Understanding variability and uncertainty in clinical chemistry metrology a case study of endogenous and exogenous molecules size and polarity

Almekdad, Dima

Awarding institution:
King's College London

Understanding variability and uncertainty in clinical chemistry metrology: a case study of endogenous and exogenous molecules size and polarity

Dima Almekdad

In partial fulfilment of the requirements for the award of Doctor of Philosophy in Analytical Science

2021

Supervisors:
Dr Leon Barron (formerly King's College London,
currently Imperial College London)
Prof. David Cowan (King's College London)
Dr Vincenzo Abbate (King's College London)
Mr Chris Hopley (National Measurement Laboratory, LGC)

Former supervisors:
Mr Chris Mussell (formerly National Measurement Laboratory, LGC)
Dr Mark Parkin (formerly King's College London)

Department of Analytical, Environmental and Forensic Sciences
Faculty of Life Sciences & Medicine
King's College London
London, UK

Table of Contents

List of Figures

List of Tables

List of Common Abbreviations

3MT	3-methoxytyramine
ANOVA	analysis of variance
BIPM	Bureau International des Poids et Mesures: International Bureau of Weights and Measures
CATs	catecholamines
CB	calibration blend
CCLM	clinical chemistry and laboratory medicine
cCRM	candidate certified reference material
CCQM	Consultative Committee on the Quantity of Material
CI	confidence interval
CID	collision induced dissociation
CITAC	Eurachem and the Co-operation of International Traceability in Analytical Chemistry
CLSI	Clinical and Laboratory Standards Institute
CMS	capillary micro-sampling
CRM	certified reference material
DBS	dried blood spots
DEM-IDMS	double exact-matched isotope dilution mass spectrometry
df	degrees of freedom
ECD	electrochemical detectors
EIA	enzymatic immunoassays
EMEA	European Medicines Agency
EQAS	external quality assurance scheme
ESI	electrospray ionisation
FDA	The Food and Drug Agency
FT-ICR	Fourier-transform ion cyclotron resonance mass spectrometry
FWHM	full width at half of the maximum peak height
GC-MS	gas chromatography with mass spectrometry
GUM	the Guide to the Expression of Uncertainty in Measurement
h	hour
HILIC	hydrophilic interaction liquid chromatography
HPLC	high performance liquid chromatography
i.d.	internal diameter
IDMS	isotope dilution mass spectrometry
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IRMM	Institute of Reference Materials and Measurements
ISD	immunosuppressant drug
ISO	International Organization for Standardization
JCTLM	Joint Committee for Traceability in Laboratory Medicine
LC	liquid chromatography

LC-MS	liquid chromatography with mass spectrometry
LC-MS/MS	liquid chromatography with tandem mass spectrometry
LogP	partition coefficient
m/z	mass to charge ratio
Met	metanephrine
METs	metanephrines
MS	mass spectrometry
ms	millisecond
MS/MS	tandem mass spectrometry
MU	measurement uncertainty
MW	molecular weight
NICE	National Institute for Health and Care Excellence
NIST	National Institute of Standards and Technology
NMI	National Measurement Institute
NML	National Measurement Laboratory
Normet	normetanephrine
PEEKsil™	polyether ether ketone with fused silica lined
PFP	pentafluoropentyl
PPGL	pheochromocytoma and paraganglioma
PRESSOR	Pheochromocytoma Research Support Organisation
PTS	proficiency testing scheme
QC	quality control
qNMR	quantitative Nuclear Magnetic Resonance
RMP	reference measurement procedure
RSD	relative standard deviation
$S:N_{p/p}$	signal-to-noise ratio peak to peak
SB	sample blend
SI	Système International: the international system of units
SILIS	stable isotopically-labelled internal standard
SPE	solid phase extraction
SRM	selected reaction monitoring
TDM	therapeutic drug monitoring
TIPTS	Tacrolimus International Proficiency Testing Scheme
TOF	time-of-flight
U	expanded uncertainty
u	standard uncertainty
UHSM	University Hospital of South Manchester
VAMS	volumetric absorptive micro-sampling
WCX	weak cation exchange

List of Poster Presentations

1. Improving transplant patient's welfare; Standardising and advancing the therapeutic drug monitoring of immunosuppressant drugs. The 3rd Annual Conference of The Association for Mass Spectrometry & Advances in the Clinical Laboratory (MSACL), Salzburg, Austria, 2016.
2. Improving transplant patient's welfare; Standardising and advancing the therapeutic drug monitoring of immunosuppressant drugs. (simplified version poster for public), STEM for Britain, Houses of Parliament, London, UK 2016.
3. Improving transplant patient's welfare; Standardising and advancing the therapeutic drug monitoring of immunosuppressant drugs. The 64th Annual American Society for Mass Spectrometry, San Antonio, Texas, 2016. (*The poster was presented by Mr Chris Mussell due to visa rejection*)
4. Development of micro-flow liquid chromatography mass spectrometry reference method for metanephrines in plasma. The 14th International Symposium on Hyphenated Techniques in Chromatography and Separation Technology, Cardiff, UK, 2018.

List of Grant Awards

1. This work is part of the Chemical and Biological Metrology programme, funded by the Department for Business, Energy and Industrial Strategy. This program supports the UK's current chemical and bio-measurement infrastructure and addresses emerging measurement needs to support innovation.
2. MSACL 2016 EU Young Investigator Travel Grant.

Acknowledgements

In the science arena

My deepest appreciation goes to Dr Leon Barron. Thank you for being my constant pillar of support, guidance and care through what was an everchanging seemingly endless PhD journey. And thanks for not leaving me behind! I am forever grateful.

Prof David Cowan, thank you for not only accepting to join my PhD journey despite your retirement, but also for your support and your patience in mentoring me, I know it was not easy, especially in the last months.

Mr Chris Hopley, thank you for taking on a project you would not have necessarily picked for yourself. Thanks for your help in the lab for the past three years.

Prof John Langley, I will always smile from my heart when I remember you saying I am your “surrogate PhD student”. Your exceptional support and compassion have been among the few constants in this PhD from day one. I will always be grateful.

Mr Chris Mussell, thank you for creating this opportunity for me and for always believing I could do it even when I doubted myself the most. I will always be grateful for the friendship we have.

Dr Simon Cowen, thank you for all your help with the statistics. I could not have managed to get it all right on my own. I really appreciate your kind help.

Dr Mark Parkin, thank you for embarking on this journey with me and for your support for the first two years of it.

Dr Vincenzo Abbate, thank you for accepting to liaise and manage this PhD at KCL for the past year. Thanks for making it a smooth and straight forward process.

Over the years of this PhD journey, I was blessed to have the scientific, technical and socio-emotional support of some great scientists as external experts, mentors, colleagues and friends, among them are:

Dr Ian Mudway, Prof Frank Kelly, Dr Andrew Kicman, Dr Milena Quaglia, Dr Bryan McCollough, Dr Danielle Moncrieffe, Dr David Douglas, Dr Caroline Pritchard, Dr Camilla Liscio, Ms Sabine Biesenbruch, Dr Cailean Clarkson, Ms Emily Whyte, Dr Kharmen Billimore, Dr Marjaana Viljanto and Ms Tabatha Hambidge.

To all my colleagues (past and present) at the National Measurement Laboratory and at King’s College London, thank you for sharing my journey and for contributing to my work.

In the personal arena

Amal and Saif, from you not only life began, but also dreams, hope, unimaginable generosity and endless support. There will never be enough words to express my appreciation, love and admiration. I love you *baba w mama*.

Aya, the most loving, truthful and caring ‘perfect little sister’. You made the weeks before submission a peaceful and grounded experience despite the emotional storm. Thanks for all the hugs, the food and care. Thanks for being my anchor, *Tati*.

Yamen, my partner in crime, my vertebra, *akhi Yabes*, thank you for being a constant source of inspiration and support every day. Thank you for being the only one to say it was ok to quit when I needed to hear it the most.

Riad, the dearest sweetest most generous *khal* anyone could wish for. I will always be grateful to you for being by my side free of judgement and full of love in the darkest times through the years.

Daniela, my sister and growth partner and witness, you have been the kindest, warmest source of support. Your eloquent sweet and kind words have given me strength and kindled warmth in my scared heart when I needed it the most.

Dr Loubaba Mamluk, “adulting” with you is a bliss! Forever grateful to you for holding my hand **every step** of the final miles of this marathon. You have anchored me in the stormiest seas of this PhD, *Louliti elghalieh shukran*.

Kholoud Mansour, thank you for always believing in me and inspiring me to grow further than my own expectations since I was 16 and until today. *Shukran Jojaity*!

Jutta ten Herkel, you made this PhD a spiritual challenge to overcome, something I would have never been able to do without your wisdom and support. Thank you.

The past years of my life through this PhD were eventful to say the least, along the way I had amazing support of many people among them are:

The Bitars, Dr Camilla Liscio, Danielle Brooker, Dr George Al Ain, Haya Fakoush, Holly Strain, Joon Lynn Goh, Julia Rampen, Karishma Woocheet, Maya Alkateb, Nadia Qtet, Raya & Hussam Zane, Rayan Azhari & Valeria Cabrera, Sabah Hamed, Dr Samer Khoury, Dr Wael BoukleyHassan & Marwa Sabbagh, Yazan Abbas and Cristina Podavitte-Soldo. Thank you for all your support and all the joy you bring me.

The biggest blessing of my life is the people in it. I am blessed with a big quantity of top-quality people whose names would take up pages. My family and friends in Damascus, my family in Bosra Al Sham, my family and friends in the diaspora, thank you all for all the love, support and inspiration you give me.

To all Syrians on your PhD journeys - past and present - thank you for inspiring strength and hope.

“O Double Lion who came from the sky, I have not reduced measures.”

Plate 31. The Papyrus of Ani. 1250 B.C.E.

The Egyptian Book of the Dead, the declaration of innocence before the Gods of the Tribunal, edited by J. Wasserman, Intrinsic Books Inc. 1998

Declaration:

I confirm that the work presented in the thesis is my own and all references are cited accordingly.

17-Dec-2020

Abstract

Standardisation of laboratory medicine and clinical chemistry applications enables the comparison of measurements that are performed at two different laboratories, in different locations and at different times. Therefore, it allows hospital laboratories to improve the accuracy of results and consequently provide better patient care. The first step to measurement standardisation of small molecules in biological samples is having a higher-order reference measurement procedure (RMP) and higher-order certified reference materials (CRMs) to achieve traceability to the SI unit. This research investigates the sources of variability and uncertainty in the development and application of higher-order RMPs and CRMs for two clinical applications. These two applications provide an extensive evaluation of the standardisation approach because 1) of the molecular weight up to about 1000 Da, these molecules cover the two ends of the molecular weight range of small molecules,2) and the nature of the compounds is rather different; being endogenous polar vs. exogenous non-polar 3) additionally they are measured in two different matrices. These case study applications are metanephrines (*m/z* 167-197) in plasma and tacrolimus (*m/z* 804) in whole blood. Both applications in this research need standardisation because they suffer from high variability among laboratories as demonstrated by external quality assurance schemes (EQAS).

The key characteristics of reference measurements are traceability to the SI unit and low measurement uncertainty estimates. In this work, novel candidate higher-order RMPs and CRMs were developed to provide the cornerstone for the standardisation of these two applications. The work described developed candidate higher-order RMPs using liquid chromatography tandem mass spectrometry (LC-MS/MS) with the double exact-match isotope dilution (DEM-IDMS) calibration approach. This thesis investigates the sources of variability and uncertainty along with the challenges to developing RMPs and CRMs to the level of accuracy, precision and traceability required to achieve higher-order reference measurements. These RMPs were applied to assign

reference values to a candidate higher-order CRM, underpinning the comparison of hospital laboratories' measurements of tacrolimus. The tacrolimus in pooled patient blood CRM that is characterised in this work is the first higher-order CRM to be produced where tacrolimus is incurred rather than spiked onto blank blood as is the case of other commercially available materials. Similarly, the RMP of plasma metanephrines is the first candidate higher-order RMP that will be submitted for listing on the Joint Committee for Traceability in Laboratory Medicine and Clinical Chemistry (JCTLM) database. Furthermore, the developed RMPs and CRMs were used for the assessment of a novel micro-sampling technique to demonstrate other areas of application of RMPs and CRMs.

The assignment of reference values by RMP and the use higher-order CRMs support: a) EQAS laboratories to perform better; b) global initiatives towards standardisation of best clinical practice guidelines; c) establishing diagnostic reference ranges; d) production of secondary CRMs; e) assigning measurement uncertainty estimates to secondary methods and f) providing a traceable standard to attain measurement traceability. Ultimately, standardisation of laboratory medicine and clinical chemistry applications aims to improve patient health care and quality of life.

1 Introduction

1.1 Standardisation: metrology & traceability

The Egyptian pyramids standing for over 4600 years are a spectacular example of how early on human civilisation identified the need for standardisation, by standardising the range of stones' dimensions used in the building. Similarly, it was only after establishing the standardisation of the nut and bolt dimensions by the British engineer Joseph Whitworth in 1841 that the glamorous Eiffel Tower in Paris was built in 1889 (1). Less spectacularly, yet crucial examples that enable our everyday life to run smoothly is the standardisation of weights, measuring tapes/rulers, vehicle tyres, railway tracks, watches/clocks, printing paper sizes, keyboards and many more that.

The need for the comparison of measurements made in different locations and at different times identified the need for a common reference standard, such as the kilogram and the meter. The earliest standard of measurement known is the Egyptian cubit, which is defined as the length of the forearm from the elbow to the tip of the middle finger (2). As the global economy was redefined post the industrial revolution, a consistent measurement system was required. The technological and scientific advances enabled this on the 20th May 1875 when seventeen nations signed the diplomatic accord of International System of Units (SI), Convention du Mètre/the Metre Convention and established the intergovernmental organisation, Bureau International des Poids et Mesures (BIPM). BIPM grew over the years and today has 62 member states and 41 associate states and economies that work on measurement science and standards (3). When the Metre Convention took place in France, the meter

was defined as 1/10 000 000 of the distance between the Equator and the North Pole as measured along the quadrant that passes through Paris. One cubic decimetre of water was defined as the kilogram. Both these units were then represented by platinum artefact standards and were preserved in Paris (4). An important advantage that the SI introduced were decimals to the units of measurement unlike the imperial measurement system. As an example, is distance and length measurement where the imperial mile denominations vary from the thou, inch, foot, yard, chain to the furlong and the longer distance unit than the mile is the league. The SI introduced a unit which is the meter with its decimals and multiplications, from the nanometre to the kilometre (5,6). The SI has been evolving with the scientific advances where the SI standards were and are being redefined into scientific constants instead of artefacts due to the risk of change, loss or damage of the artefacts. The first SI unit to be redefined to a conceptual scientific definition was the meter. The meter was redefined several times from the length of an artefact through to '*an atomic reference transition to the fixed numerical value of the speed of light*' (5–7). The advantages seen with the redefinition of the meter were among the reasons that led to defining all units by defining constants instead of artefacts. The SI unit of mass, the kilogram (kg), was redefined from the artefact metal kilogram in Paris to a new definition based on constants in 2018. The kg is '*defined by taking the fixed numerical value of the Planck constant h to be* $6.626\ 070\ 15 \times 10^{-34}$ *when expressed in the unit J s, which is equal to* $kg\ m^2\ s^{-1}$ *, where the metre and the second are defined in terms of c and* $\Delta\ v_{Cs}$' (7).

To understand standardisation, it is crucial to understand two intertwined concepts: metrology and traceability. While the above described the background of common standards and standardisation, today metrology is defined as *"the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology"* (8). Over the years and with advances in science, the concepts of metrology have extended to the fields of biology and chemistry to compare quantities of materials. Expressions and units were defined to express quantities numerically; mass fraction is expressed by weight per weight e.g., ng/g; molar fraction is defined by mole per mole e.g., nmole/mole.

The second core concept of standardisation is traceability, which is the route to relate any measurement to the internationally defined SI. Traceability is defined as "*property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty*" (8). The traceability chain is a chain consisting of calibrators and reference measurement procedures (RMPs) that link the SI to a final measurement result, see Figure 1.1. In metrological terms, the highest link in the chain is the primary standard and/or the primary method (9). In clinical chemistry and laboratory medicine, this is referred to as 'higher-order' certified reference material (CRM) which is assigned a value via a higher-order RMP and is usually of the smallest measurement uncertainty attainable.

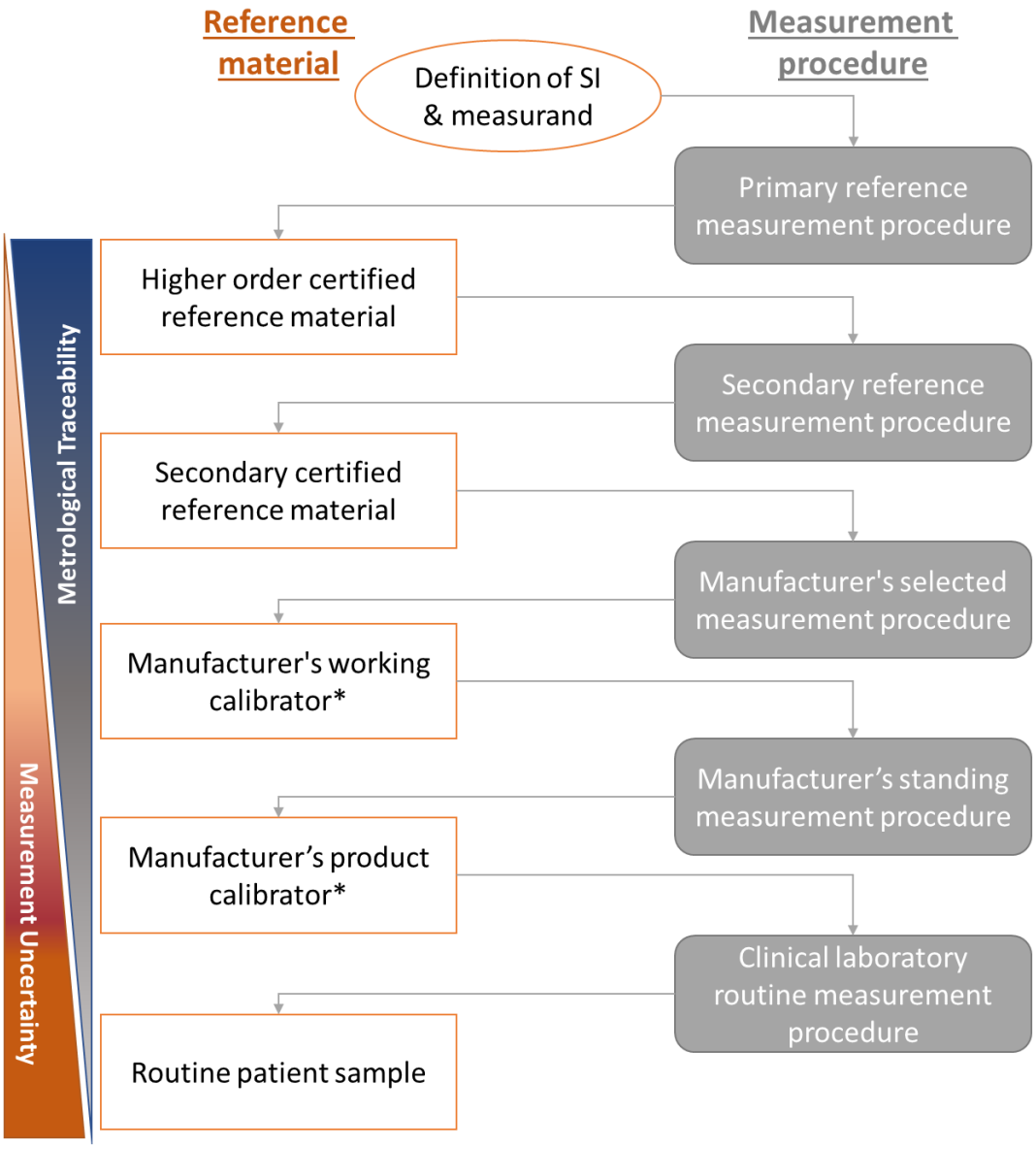


Figure 1.1 Metrological traceability chain diagram demonstrating the link between the top SI units to the reported patient result at the end of the chain by a series of reference value assignments using calibrated measurement procedures. Adapted from (10)

The definitions of primary methods for the measurement of amount of substance evolved over the years. The metrological definitions are assigned by the BIPM consultative committee for amount of substance (CCQM) (11). In 1995, the term primary method of measurement was introduced and defined as "*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, and whose results are,*

therefore, accepted without reference to a standard of the quantity being measured" (12). This definition was modified in 1998; "*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*". This also included categorising primary method as "direct" and "ratio" methods, where; *"A primary direct method measures the value of an unknown without reference to a standard of the same quantity"* and *"A primary ratio method measures the value of a ratio of an unknown to a standard of the same quantity; its operation must be completely described by a measurement equation"* (13). Hence, a primary direct method measures directly in relation to the SI or a fundamental constant (e.g., Faraday constant), an example of a primary method is gravimetry and colourimetry. The primary ratio method, on the other hand, measures the quantity in relation to a traceable reference standard. These methods are used to assign values of a substance in a medium such as measurement of mass fraction of a substance within a matrix (9). Primary ratio methods are used in clinical chemical metrology to provide a high level of traceability to higher-order CRMs (Figure 1.1). When double exact-matched isotope dilution mass spectrometry (IDMS) is performed correctly and with gravimetric preparation, it is a primary ratio method. This primary ratio method is the basis of the work described in this thesis (14,15).

1.2 Standardisation of laboratory medicine applications

The clinical community has identified the need for standardisation and harmonisation for better clinical practice (e.g., diagnosis, treatment) and improved patient care. The standardisation process is achieved in laboratory

medicine through the utilisation of CRMs; both solvent and matrix calibrants and RMPs. Each higher-order CRM is a standard that contains an analyte of interest that is quantified and assigned a reference value with corresponding measurement uncertainty using a higher-order RMP. The International Federation for Clinical Chemistry and Laboratory Medicine (IFCC) identified the necessity for traceability to achieve standardisation in the field. The use of higher-order CRMs when used correctly can provide traceability to the SI unit against which subsequent CRMs and RMPs are calibrated. Hence, using the appropriate RMPs and/or CRMs enables producing accurate and precise measurements that are comparable across different laboratories in different locations and at different times (16).

1.2.1 Traceability in laboratory medicine

Significant clinical and public health decisions are based on laboratory measurement. These measurements from individual patients or populations are usually compared across measurement systems, time and locations to inform clinical research whose outcome is translated into many forms of patient care from disease diagnosis, prognosis, prevention, treatment and control. To enable the comparison of measurement, these are compared against a reference value determined using reference measurement procedures. This can be a reference clinical decision point, a set of results obtained from a population or the same patient at an earlier time (e.g., before certain symptoms occurred). Additionally, clinicians compare laboratory measurement across different studies in different locations to formulate common clinical decisions and translate research outcome into patient care. Hence, a fundamental goal of laboratory medicine is that measurements are

comparable regardless of the measurement system at the clinical laboratory. When a number of routine measurement procedures (e.g. analytical methods), which have analytical specificity and selectivity, are traceable to the same higher-order CRM or RMP they should produce equivalent values for comparable samples irrespective of the laboratory producing the result, the location or the time of the procedure (10,17–19).

Traceability is the cornerstone that enables the comparison of measurement results required in laboratory medicine research and patient care. Traceable measurements enable:

a) the definition of common clinical reference ranges, intervals and cut-off points instead of method-specific ranges;

b) a comparison of measurement from different studies universally to improve medical research and hence patient care and

c) the formulation of consistent medical care practices through best laboratory medicine practice guidelines, clinical practice guidelines, numeric and evidence-based medicine (17,20–23). Armbruster *et al.* reported in their review of the work of the joint committee of traceability in laboratory medicine and clinical chemistry (JCTLM), that the traceability of measurements must be established for all clinical analytes. They stated that the lack of traceable results could jeopardise patient safety if results were misinterpreted against the common clinical reference ranges. For example patients undergoing unnecessary biopsies or missing necessary biopsies as a result of inaccurate non-traceable values of prostate-specific antigen (16). Measurement procedures should be validated to prove performance, and ideally all measurements performed would be traceable to the SI unit however in reality this is a very difficult goal to achieve. This would be improved by achieving

traceability in the analytical instrumentation being used in hospital laboratories. Since traceability of all measurements is very difficult to achieve, the performance and quality of hospital laboratory measurements are assessed through proficiency testing schemes (PTS) or external quality assurance schemes (EQAS). These schemes are usually independent from hospital laboratories and samples are periodically sent to subscribing laboratories to evaluate the performance of the methods for specific clinical tests. The data are then pooled and compared and reported to the subscribing laboratories to provide them with an evaluation of their methods. These schemes in many cases lack a reference value assignment to the samples which makes the interpretation of the comparison data not necessarily accurate. The United Kingdom Accreditation Service (UKAS) medical laboratory accreditation (ISO 15189:2012) requires medical laboratories to subscribe to PTS and EQAS where possible (24). This is an important step moving to standardisation and harmonisation of the analytical work of routine hospital laboratories because it allows comparison of results and would supply information on whether there is a need for standardisation or not for a specific application. This will then determine the initiation of a standardisation initiative by setting up a reference measurement system.

A reference measurement system is required to establish traceability. This system consists of four components: (a) defining the measurand (b) stated RMP and/or CRM (c) knowledge of measurement uncertainties and (d) unbroken chain of calibration and value assignments, see traceability chain in Figure 1.1 (8,17).

For a traceability-based standardisation initiative to take place several essential components are required. Firstly, and most importantly is defining the measurand of the analysis. IUPAC defines the measurand as a "*particular quantity subject to measurement*" (8,25), the definition of the measurand encompasses not only the compound but also a specific definition of the matrix in which the compound is present and the representation of the sample (i.e. whether the measurement relates to the sample being analysed only or to a larger population from which a sample comes). The less specific term 'analyte' is widely used in the clinical area but does not replace the measurand (26). The necessity of defining the measurand stems from the clinical importance of the compound, the form it exists in and the matrix it is measured in. If a compound exists naturally as a free compound or bound to proteins or conjugates, it is key to know which of the two forms or if both forms are of clinical relevance. As an example, cholesterol is an endogenous compound analysed in serum where it is naturally present in two forms; 'free' and bound to proteins, depending on the clinical diagnosis, the clinician requests either cholesterol (free) or total cholesterol (both free and bound). Hence, for the standardisation initiative of cholesterol in serum both measurands were identified (27). Once the measurand is defined, the first step is to develop a higher-order RMP, whose development is usually performed by national measurement institutes (NMIs). When the RMP is used to assign a reference value measurement, a reference material is usually used to verify the traceability of the measurement (5,6,13). While hospital laboratories evaluate performance and quality of measurement through subscribing to EQAS and PTS, NMI laboratories compare their reference measurements through international comparison studies under the umbrella of the consultative

committee for amount of substance: metrology in chemistry and biology (CCQM), these are referred to as CCQM studies. In these studies samples are sent to the NMIs in the different countries and the measurements are compared against a reference value (11).

1.2.2 Commutability of CRMs

Standardisation initiatives developed RMPs and produced CRMs for over 50 measurands. However, variability in results are still observed when analytes are measured by different laboratory measurement procedures (19). This has been observed in EQAS data where samples sent to the different subscribing laboratories showed variability in the results produced by different laboratories. One of the reasons for this variability was the commutability of some CRMs. Commutability of CRMs is defined by the clinical and laboratory standards institute (CLSI) as *"the property of a reference material, demonstrated by the equivalence of the mathematical relationships among the results of different measurement procedures for a reference material and for representative samples of healthy and diseased individuals of the type intended to be measured"* (19,26,28–30). To simplify, the study of commutability of CRMs compares the performance of the different analytical methods for a specific measurand and the similarity of the CRM behaviour in a method to that of patient samples (19). In routine clinical laboratory measurements, the analytical accuracy of a measurement is mostly based on two steps, transferring traceability and trueness from a higher-order CRM and through monitoring the stability of the performance on the method's trueness over time. Hence, these CRMs must demonstrate behaviour similar to that of

patient samples and give the same result independent of the analytical method used (31).

Lack of commutability of some cholesterol materials is a reported example of the impact of commutability on the variability observed in interlaboratory comparison studies. This was when laboratories and EQAS used material that were not of high metrological order and higher variability was observed when using lyophilized materials instead of fresh or fresh-frozen serum. The interlaboratory variability of results was improved and comparable to that of fresh serum only by using materials that were lyophilised in the presence of sucrose (31,32). In another study where 480 laboratories participated, a non-commutable material of lyophilised serum was in part the reason of the difference observed between the IDMS based method measurement and the mean of the interlaboratory measurements (31,33). Another example of the importance of commutability of CRMs and materials used for interlaboratory assessment studies is creatinine. In a Canadian assessment of clinical secondary reference methods performed in different laboratories in Ontario, a variety of control materials were used to assess interlaboratory variability. The methods were only found consistent with the use of the higher-order CRM produced by National Institute of Standards and Technology (NIST) (31,34). NIST is the national measurement laboratory of the United States.

The classification of CRMs in matrix as higher-order CRMs requires that commutability of the CRM be evaluated. Historically, the importance of commutability validation has been underappreciated. However, reviewing the list of the database of the joint committee for traceability in laboratory medicine (JCTLM) shows that the number of CRMs validated for commutability has

increased (19,35). This is mainly due to the importance of excluding the non-commutability as a source of variability in data obtained from different laboratory results. Hence if a CRM is to be analysed by a routine clinical laboratory, commutability must be validated among the methods and techniques that are used for the analysis, including the reference measurement procedure when appropriate (19).

The JCTLM database is a key reference for standardisation of CCLM, which is managed by the BIPM and IFCC. The database includes three categories of materials: biological matrices, calibration solutions and high purity materials. The high purity solid/powder materials are intended for use by higher-order RMPs and the production of secondary calibration materials and methods is more for the use by routine clinical laboratories. Biological matrices, on the other hand, could be used by both routine and reference measurement procedures (35). Some examples of higher-order materials of measurands in biological matrices include; cadmium in human blood, albumin in human serum, cholesterol in human serum, and creatinine in human serum (35). The decision on which CRMs are to be produced is usually based on the needs indicated by the clinical community to the national measurement laboratories and the final decision would usually be based on the review panels of government funding in each country. This was the case for the work reported in this thesis where the need was identified by the clinical community. In response the National Measurement Laboratory (NML) hosted at LGC submitted the proposals to The Chemical and Biological Metrology programme (CBM), funded by the UK government Department for Business, Energy & Industrial Strategy (BEIS). The CBM supports the UK's current

chemical and bio-measurement infrastructure and addresses emerging measurement needs to support innovation.

Given that higher-order materials are generally more expensive and more time and effort consuming to be produced, they are mainly targeted for the use by manufacturers of analytical instrumentation and calibrators. Essentially, a CRM is used to confirm that results of clinical patient samples analysed by routine clinical laboratories, regardless of their measurement procedure, and provide numerical values that are equivalent and are traceable to the SI (19).

1.3 Clinical chemistry and laboratory medicine (CCLM) applications & techniques

Laboratory medicine is a massive field that covers hundreds of measurements of thousands of samples across many areas from anatomic pathology to clinical pathology. The latter includes a wide range of laboratory testing fields including microbiology, haematology, clinical chemistry and many more. The work reported in this research relates to clinical chemistry unit of the hospital laboratory where many analytical measurements are performed on patient samples, each with its own unique clinical significance. Of the wide range of areas in clinical chemistry testing are clinical diagnostic tests and therapeutic drug monitoring. Mass spectrometry started being used in the clinical chemistry laboratory due its ability to perform qualitative and quantitative measurements, initially it was used with gas chromatography mass spectrometry (GC-MS). With the evolution of coupling liquid chromatography with mass spectrometry (LC-MS), LC-MS started being used in the clinical chemistry laboratory too (36,37). This introduction of GC-MS and LC-MS into

the clinical laboratory improved the potential for standardisation of different clinical tests as the technique enables development of traceable reference methods using isotope dilution mass spectrometry (IDMS) and the low measurement uncertainty calibration mode of double exact matching IDMS.

GC-MS is a gold standard in terms of cost-performance and superb chromatographic separation. However, it is limited to analysing volatile compounds with separation in the gas phase. LC-MS on the other hand works in the same phase as the biological matrix that is fluid which could overcome possible stability issues depending on the compound. The technique that was selected for this research was LC-MS for many reasons including: a) to avoid the need to add a derivatisation step for metanephrines to the sample preparation procedure as it is already long and complex due to gravimetric blend preparation; b) instrument availability and c) LC-MS has been successfully used for several candidate higher-order RMPs.

Many candidate higher-order RMPs for clinical chemistry tests and applications reported in the literature use ID LC-MS methods. Some examples include; folate in serum (38), cortisol in serum (39), glucose in plasma (40), 25-hydroxyvitamin D_3 and 25-hydroxyvitamin D_2 in serum (41,42) and paracetamol in plasma and serum (43). Moreover, upon reviewing the JCTLM list of reference measurement methods/procedures for drugs in biological matrices, only 3 out of 15 listed methods use GC-MS. These GC-MS methods were listed in 1984, 1994 and 2002, all before the current technological advances in LC-MS (35).

1.3.1 Clinical diagnosis and the selection of plasma metanephrines for this research

The classic definition of medical diagnosis is the detection or exclusion of disease, this definition however expanded slowly (44). The generally used definition in the field agrees with Knotterus *et al.*'s definition, which expanded from the detection and exclusion of disease to include the assessment of disease prognosis, evaluation of disease risk, patient monitoring, measurement of general fitness/health, etc. (44,45). The development of technology and the different techniques that are used in the clinical diagnosis has necessitated that clinical diagnostic tests are assessed for their "diagnostic accuracy". Diagnostic accuracy is defined as *"a medical test ability to provide accurate information about diagnosis, prognosis, risk of disease and other clinical issues"* (46). This is not to be confused with an accuracy assessment of a technical measurement method. Diagnostic accuracy of a medical test is a measure of the test's reliability. It is assessed through applying the test on a large population of patients and determining the false positive and false negatives, where the former is that the test detects a disease that is not there or the latter when the test fails to detect the disease and excludes it while it is there.

In this research, the standardisation of the analysis of metanephrines in plasma is investigated. This is the medical test to diagnose paraganglioma and pheochromocytoma and has a diagnostic accuracy of 99 % at a confidence interval of 95 % (47). Further information about this test is provided in Chapters 2 and 3.

The choice here to study plasma metanephrines analysis was to use this application as an example case study of small, polar compounds in plasma. This is to fulfil the aim of the research in understanding variability in measurement and sources of measurement uncertainty in clinical laboratory tests of small molecules of different polarities that range in their molecular weight from 100-1000 g/mol. The choice of metanephrines as a case study covers many aspects of the analysis as they are; (a) at the smaller end of the molecular weight range with molecular weights of around 200 g/mol; (b) highly polar compounds (typically not suited for the commonly used reversed-phase liquid chromatography); (c) endogenous compounds as they are metabolites of adrenaline, noradrenaline and dopamine; and (d) they are measured in plasma, one of the commonly used matrices in clinical testing.

1.3.2 Therapeutic drug monitoring diagnosis and the selection of tacrolimus in whole blood for this research

The introduction of new analytical techniques into clinical chemistry laboratories in the 1960s enabled the measurement of the low concentrations of drugs in biological fluids during drug treatment. This was a turning point in the individualisation of patient drug therapy as it gave the opportunity to implement therapeutic drug monitoring (TDM). TDM is the measurement of the concentration of a drug in a biological sample, e.g., plasma, or whole blood, to optimise drug therapy regimen. TDM reduces the variability coming from pharmacokinetics by controlling a patient's drug therapy based on the concentrations of the drug in their body instead of using the drug dose only. Not all drugs are suited for TDM. The leading experts in the TDM of immunosuppressant drugs Prof. Johnston and Prof. Holt shortlisted four key

characteristics of a drug to be suitable for TDM (48,49) and these are: (a) the drug should have a clear relationship between the pharmacological effect and concentration of drug in blood; (b) the drug should have a narrow therapeutic index. The therapeutic index is an indicator of drug safety as it is the range of quantitative measurement of the drug in the blood at which the drug is effective and relatively safe and not toxic (50,51). A narrow therapeutic index means that even a small reduction in drug concentration in blood results in a lack of efficacy and any excess of drug concentration would result in adverse toxicities; (c) the drug should have significant inter-patient pharmacokinetic variability resulting in poor relationship between the drug's dose and the drug's pharmacological efficacy; and (d) the adverse effects and toxicities of the drug should be difficult to distinguish from the drug efficacy.

TDM of the immunosuppressant drug tacrolimus is undertaken by the analysis of the drug in whole blood samples because these drugs are lipophilic and are bound to red blood cells. In this research the standardisation of this application is investigated. Tacrolimus in whole blood analysis in this research was the example application of the analysis at the mid-large end of the small molecule range. This was to fulfil the aim of the research in understanding variability in measurement and sources of measurement uncertainty in clinical laboratory tests of small molecules that range in their molecular weight from 100-1000 g/mol. The choice of tacrolimus as a case study satisfies many aspects of the analysis including: (a) covering the other end of the molecular range as tacrolimus is 804.02 g/mol; (b) a highly lipophilic non-polar compound; (c) an exogenous drug not produced by the human body; and (d) is measured in whole blood, one of the more complex matrices in clinical testing. Further

background information is provided in Chapters 4 and 5. With the work performed on both metanephrines in plasma analysis and tacrolimus in whole blood, learnings and conclusions about the sources of variability in clinical measurement testing of small molecules was obtained. An understanding of the different matrices and their effect on the measurement was also obtained as well as the analysis of molecules of different chemistries. The two applications presented different sets of challenges as well as common sources of variability.

1.4 The need for standardisation of CCLM

Ideally calibrators of all commercial clinical tests would be traceable to higher-order CRMs and/or RMPs if they are not directly traceable to the SI themselves. In 1998 the European Union (EU) *in vitro* Diagnostic Medical Devices directive (IVD) requires that calibrators of all commercial clinical tests sold within the EU should demonstrate traceability to higher-order CRMs or RMPs where available (52). Achieving traceability in the field of CCLM achieves: (a) suitable evaluation of clinical cut-off points and values used in disease diagnosis and prognosis evaluation; (b) use of appropriate clinical interventions and therapies; (c) avoids potential misclassification of disease; and (d) supports the production of international clinical best practice guidelines.

The international measurement evaluation program (IMEP) is a scheme by the European Union Hub that provides metrological inter-laboratory comparison schemes to enable benchmarking laboratory performance (53). The international study “trace and minor constituents in serum (IMEP-17)” is an example of how variable different analytical methods and techniques can

be (54). Figure 1.2 represents 1022 measurement results of the same creatinine in serum material analysed by over 1000 laboratories around the world using different analytical methods, where data shown in the graph are grouped based on the technique/method used. The wide spread of data in comparison to the reference value, that was assigned by a dozen NMIs around the world, demonstrate the need for a reference value for the laboratories to evaluate their results against. It also demonstrates the difference between a result obtained by a RMP where a measurement uncertainty is reported (shown as a grey rectangle across the figure) and a routine method where usually the uncertainty is not reported (data points plotted in the figure as values with no error bars).

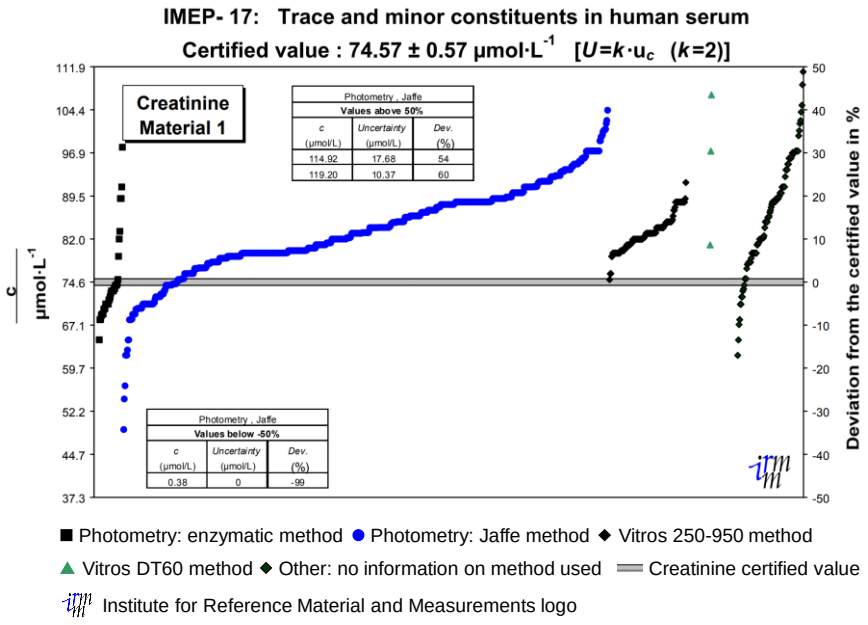


Figure 1.2 Inter-laboratories comparison study of creatinine measurements in a human serum material performed by the Institute of Reference Materials and Measurements (IRMM). The results in the graph are 1022 measurements that were reported by the different laboratories reporting different methods. The measurements were grouped based on the analytical method used. The certified value was the reference measurement with its expanded measurement uncertainty reported by the NMIs that analysed the material. Figure used with permission from (54).

The traceability and standardisation approach proved effective as it has been extensively studied for some clinical tests such as creatinine in serum (55) and cholesterol in serum (56). In the case of cholesterol in serum, adopting metrological traceability demonstrated a four-fold reduction in measurement uncertainty associated with the measurement and savings of about 100 million USD/year in misdiagnosed patients' treatment costs (57). It is worth noting at this stage that developing a higher-order RMP requires extensive research that needs resources and metrological expertise. Hence, the development of such RMPs and CRMs usually falls under the national measurement institutions (NMI) that are typically funded by governments.

Standardisation of laboratory medicine enables the creation of universal reference ranges and clinical cut-off points for the diagnosis, prognosis and treatment of different conditions. For years, clinical practice guidelines included reference ranges and clinical cut-off values for laboratory tests despite the lack of standardisation of most of these tests. This has been due to the fact that these guidelines used to be developed almost exclusively by clinicians who have very little knowledge of the actual laboratory procedures and often within a single hospital laboratory environment (58). There is significant variability in laboratory test results produced by different measurement procedures (Figure 1.2). This has been increasingly considered over the years and numerous reference systems and standardisation initiatives have taken place for several measurands.

1.5 LC-MS in the clinical laboratory

The mass spectrometer was introduced into the clinical laboratory from the forensic toxicology laboratory, when the US federal workplace drug testing programmes mandated a confirmation of a positive drugs of abuse test by GC-MS in 1988 (59,60). The first testing area for MS to be used in the clinical laboratory was in new-born screening, as the laboratory grew more familiar with GC-MS, it was used to prove the limitation of immunoassay measurement of low concentration steroids especially in females and children (61,62). With the introduction of atmospheric pressure ionisation sources and the inherent limitation of GC-MS to analyse only thermally stable and volatile compounds, caused a gradual shift from GC-MS to LC-MS (Figure 1.3). The first applications transferred to LC-MS were new-born screening, renal, cardiovascular, endocrine and drugs and toxicology in the late 1990s-early 2000s (36,59). A clear comparison between the routine techniques in the clinical laboratory with LC-MS has been summarised by Prof. Michael Vogeser in 2016, from which Table 1.1 below was adapted (63).

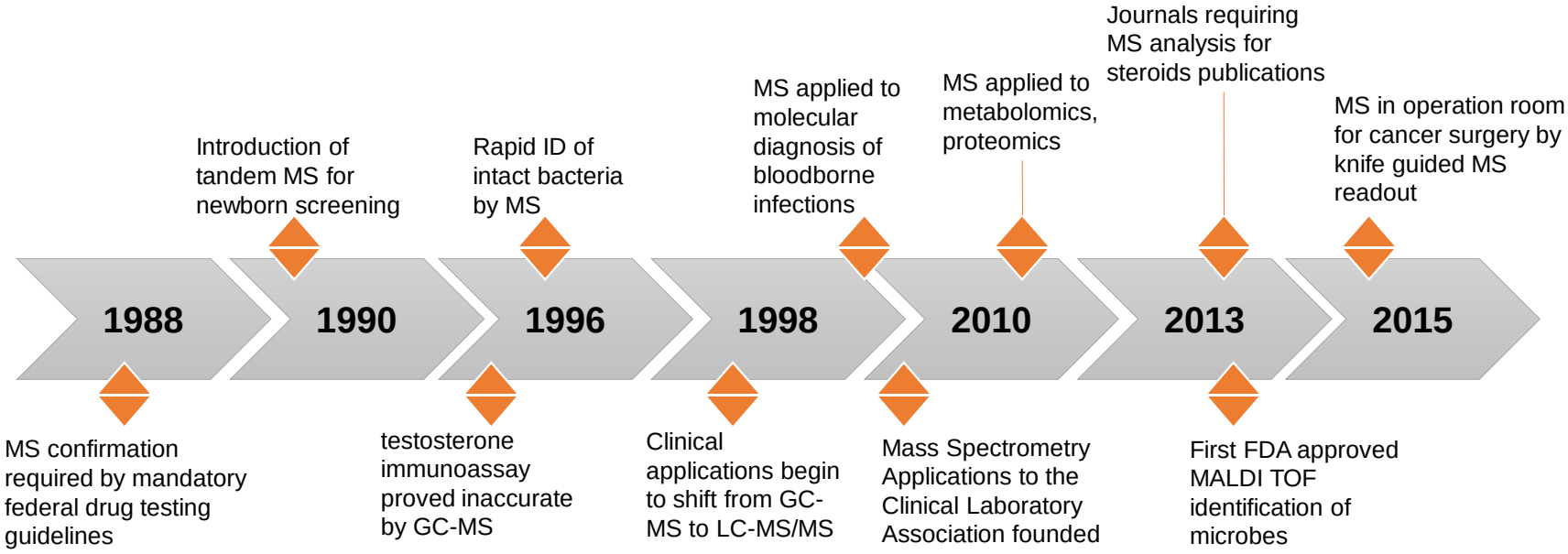


Figure 1.3 The evolution of the use of mass spectrometry in the clinical laboratory. This figure was adapted from Jannetto and Fitzgerald (59).

Table 1.1 Comparison of current standard methods used in clinical pathology and mass spectrometry. Adapted from (63).

Variable	Photometry & immunoassay	Mass spectrometry
Detection specificity	Low; indirect based on shape or enzymatic reactions	High; *m/z* and ion fragmentation
Cross-detection risk	High	Minimised
Matrix effect	No compensation	Can be compensated for by internal standard(s)
Analyte(s)	No simultaneous analysis of multiple analytes in the same method, possibly combining results of single analyte tests.	Simultaneous analysis of many analytes in one method (e.g., many SRMs for different analytes in the same method)
Sample preparation	Crude sample analysis	Extensive sample preparation required for most applications
Instrumental set up	Single technique is used	High versatility of techniques including: LC, GC, ion mobility, precursor ion selection, MS/MS fragmentation, high accuracy MS
Data acquired	Only a read-out of the quantitative result with predefined error flagging	Technical validation of results required manually or semi-automated, e.g., assessment of retention time repeatability, peak area and peak shape asymmetry

The limitations of immunoassay for the analysis of small molecules became more evident in the mid-1990s. This initiated the move towards adopting HPLC and GC-MS. About 15 to 20 years ago LC-MS pushed itself as the newly adopted technology in the clinical laboratory and it has been advancing ever since. This was due to the advantages of LC-MS that include; higher throughput, easier workflows, wide-ranging versatility and lower running costs after initial instrument installation capital (36). More importantly, LC-MS provided higher analyte selectivity and higher detection sensitivity combined. Hence many applications of the different areas within the clinical laboratory moved over to LC-MS in many laboratories around the world, with endocrine applications and TDM being two of the main areas that LC-MS proved advantageous to other techniques (36,64). Oestrogen and testosterone analysis used to be performed using radio-immunoassay (RIA) and were shifted to LC-MS methods in the late 1990s. LC-MS methods provided improved sensitivity which resulted in lower limits of detections (LOD) and improved lower limit of quantification (LLOQ) (37). Despite GC-MS methods generally being more sensitive, LC-MS was advantageous as it was compatible with a wider range of compounds and as GC-MS required more laborious sample preparation.

Oestrogens, testosterone and estradiols are among the sex steroid tests that are generally used for female reproductive function assessment, menopausal status, puberty disorders, TDM and other conditions. These tests have moved through significant advances from colourimetry (1950-1960s) to radio-immunoassays (1970-1980s) to automated immunoassays (1990-2000s) then to LC-MS in 2010-2011 onwards (65–70). Cortisol measurements in plasma

for the diagnosis of Cushing's syndrome used to be performed using different techniques including LC-UV, immunoassay and GC-MS. However, LC-MS showed better selectivity, less interference, and improved LOD, in addition to the advantage of simultaneous quantification such as for cortisone metabolite (71–75). Among the other endocrinology tests that also moved over to LC-MS is aldosterone measurement in plasma, serum or urine, which used to be analysed using RIA and GC-MS to diagnose primary aldosteronism, hyperaldosteronism and tumour confirmation. LC-MS methods are reported to improve precision and have lower LLOQs (76,77). Most recently "*the test of the decade*", vitamin D analysis, which was analysed by RIA, LC-UV and HPLC coupled to an electron capture detector moved to LC-MS. Although LC-MS has equal robustness and similar throughput to immunoassay, LC-MS methods enabled standardisation of the test reducing the large variability reported with other tests (41,78–82). LC-MS provided major advantages to large molecules and proteins analysis as well but is not covered in this introduction because this thesis is investigating measurements of small molecules in clinical testing. Moreover, as mentioned earlier in section 1.3, LC-MS has become a leading technology in higher-order reference measurement procedures. Through the work in this research a better understanding of the variability and measurement uncertainty in LC-MS based higher-order measurement procedures was obtained. Such knowledge is transferable to the clinical laboratory and can provide insights on how to reduce variability in measurement or at least how better assess measurement uncertainty estimates.

Over the past 15-20 years with the growth in the use of LC-MS in clinical laboratories and the increasing knowledge of the techniques, better awareness about the limitations of LC-MS came to light. Grebe and Singh from the Mayo clinic have summarised these as the ‘interacting triangle’ of sensitivity, specificity and throughput (36). LC-MS vendors world-wide are in constant competition to bring to the market new instruments with higher sensitivity and specificity, they are also working on automating the sample preparation processes and bringing procedures online with the instrument. An important example is the introduction of automated online solid phase extraction (SPE) techniques where once samples are pre-treated they are moved to the LC-MS on which an online SPE system is set up and the SPE procedure is performed automatically with extracts being directly injected on to the LC column (83,84).

Due to the inherent complexity of LC-MS as a technique, highly skilled users are required (36,85). With the growing use of LC-MS and the growing published literature in the different applications, the knowledge of common pitfalls and wrong practices by the users are more widely known (85). Figure 1.14 summarises a few of these in addition to an overview of the sources of variability in LC-MS measurements. More discussion about the sources of variability in the use of LC-MS is covered in 1.10.

1.6 Chromatography

With the extensive mention of the use of LC-MS thus far in this chapter, this section is dedicated to give a brief overview of chromatography and more specifically LC principles and concepts, focusing on those used in this work only.

It is customary when talking about chromatography to start with the father of chromatography, Mikhail Tsvett, the Russian botanist who in 1906 marked the beginning of the chromatography era when he separated chloroplast pigments in leaf extract and was the first to publish the scientific principles of the separation (86). Additionally, he gave the field its name coupling the Greek words for colour and to write, "χρῶμα, chroma" and "γράφειν, graphein", respectively (86). Since then column chromatography evolved, when Martins and Synge introduced the concept of liquid-liquid partition chromatography using a silica gel as a stationary phase in the 1940s (87). A decade later Martin published the potential for using a gaseous mobile phase (88,89). These were the bases of what we today know as liquid chromatography and gas chromatography.

1.6.1 Theory of chromatography

Chromatography is a process that separates a mixture of compounds into individual or groups of different compounds based on differential modes that include: size exclusion chromatography and interactions based chromatography (90,91). In interaction based chromatography the mixture of compounds is introduced to the chromatographic system as a discrete plug and each of the compounds travel through or over the stationary phase

interacting to a different degree depending on their chemical properties and elute at the end at different rates, i.e., the compounds with lower affinity for the stationary phase elute from the column first as they travel through faster (see Figure 1.4). The eluting compounds are then monitored, and signal is recorded using a detector which ultimately produces a chromatogram. The time at which the compounds elute, which is the time between the injection of the sample and the detection of a compound, is known as the retention time (t_R) and is a key characteristic for the identification of a compound. The area under the peak can also be used for quantitative purposes (88,91–93).

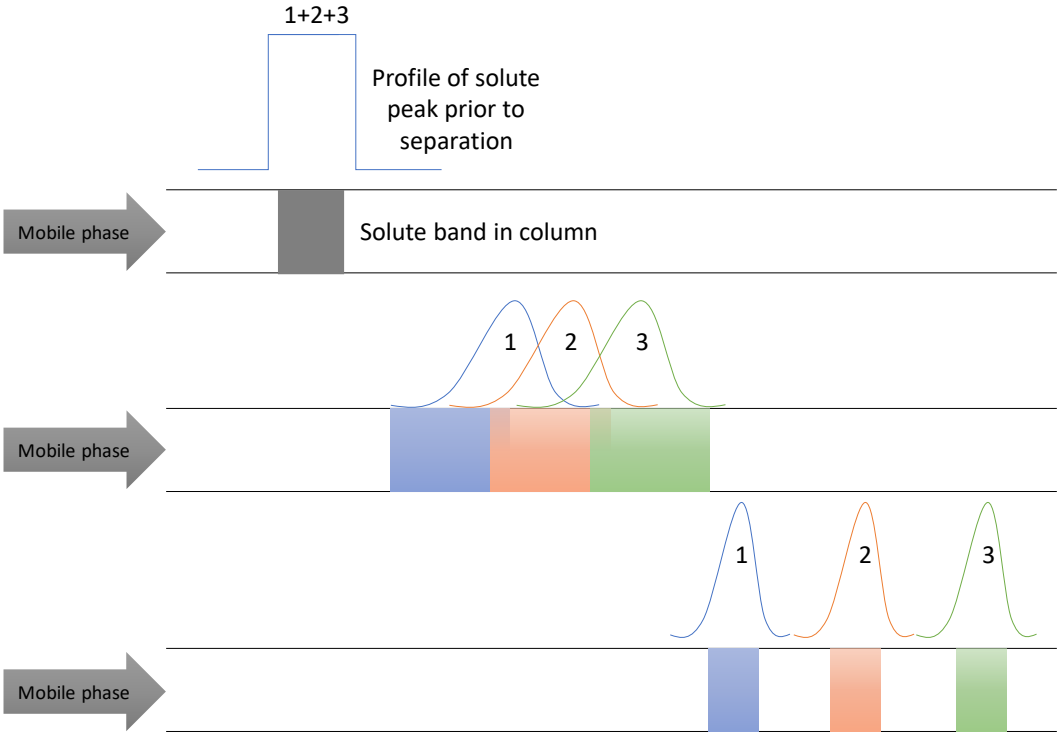


Figure 1.4 Chromatographic separation of a mixture of three compounds. Top part showing the injection of the sample as a mixture as it arrives at the beginning of the column. Middle part shows the mixture separating slowly based on the interactions with the stationery and mobile phases until the compounds are resolved on the column in the bottom part of the figure.

The main chromatography types are planar and columnar. However, many modes of chromatography were developed aiming to separate mixtures of different chemistries and sample types. Planar formats include thin layer and

paper chromatography and were not used in this thesis. Columnar chromatography is used in both gas and liquid chromatography and commonly used modes included; ion-exchange chromatography (IC), gel-permeation (molecular sieve) chromatography, affinity chromatography, hydrophobic-interaction chromatography and hydrophilic interactions liquid chromatography (HILIC) (92). Most used chromatographic techniques in clinical laboratories are GC and UHPLC. In GC, the mobile phase is gas while the stationary phase can be either solid or a liquid material grafted on to the column wall (89). In LC, on the other hand, the mobile phase is always liquid and the stationary phase is typically made of small, densely packed particles that could be solid or porous upon which different chemistries are functionalised (93).

1.6.2 Liquid chromatography

As mentioned earlier many reference methods and clinical laboratories moved to LC-MS due to its numerous advantages. The work described in this thesis used LC-MS. This section aims to give a brief background overview of the types of LC that were used in this work only, as covering the wide science of chromatography is outside the scope of this thesis.

As mentioned above the mobile phase in LC is a liquid, depending on the chemistries of the stationary and mobile phases, LC can be classified to many types including but not limited to normal phase LC (NPLC) or reversed-phase LC (RPLC). Most applications in TDM use RPLC however HILIC and IC are also used in some clinical tests. In NPLC, the mobile phase is less polar than the stationary phase whereas in RPLC the mobile phase is of higher polarity than the stationary phase. Additionally, the sample that is being loaded onto

the column would be of lower polarity to allow interaction with the stationary phase (93,94). The separation mechanisms of chromatography occur in numerous ways including via adsorption, partition, size exclusion and ion exchange (95–98).

Typically, RPLC stationary phases are constructed of silica particles on to which hydrophobic chemical functions are chemically bonded, most common stationary phases in an increasing order of polarity include C_{18}, C_8, C_4, phenyl, cyano, and amino. Mobile phases used to elute compounds off the stationary phase are more polar than the stationary phase and most notably consist of water and miscible organic solvents. Buffers are often used to assist with ionisation, solubility and chromatographic selectivity. For coupling LC to MS, these are often constituted by formate or acetate salts. The compounds are retained on the stationary phase of RPLC columns based on their varying degrees of hydrophobicity, mostly via van der Waals interactions, with the more hydrophobic compounds retained longer and require stronger mobile phase to be eluted. The hydrophobicity of the compounds is typically indicated by their Log P (i.e., Log K_{ow}) and Log D which are the measures of the partitioning of the analyte in between two immiscible solvents (octanol and water) in standard conditions. Log P is the descriptor for non-ionisable compounds whereas Log D is the descriptor for ionisable compounds, for neutral compounds Log P = Log D. The strength of the mobile phase on the other hand is determined by the polarity of the solvent, water is the ‘weakest’ of solvents as it is most polar and would not elute hydrophobic compounds. Hence, organic modifiers are required to achieve elution, these modifiers’ strength is expressed by the Snyder index ε^o which indicates the polarity of

the solvents. The most common organic solvents in the order of decreasing elution strength i.e. increasing polarity are tetrahydrofuran, acetonitrile, methanol and water (93,94,99).

This makes RPLC the first choice for the chromatography of many compounds; non-polar, neutral polar compounds, weak bases and acids, proteins, peptides and homologues. RPLC fails to retain and separate very hydrophilic polar compounds, strong bases and acids, and for these, IC or HILIC and occasionally NPLC are used. However, RPLC is the most popular mode of LC used to date. For NPLC, the polar mobile phase solvents used are not readily compatible with electrospray ionisation mass spectrometry (93,94,99). This resulted in the development of HILIC. The first publication describing HILIC separation and giving it this name was by Alpert in the early 1990s (100). HILIC uses polar stationary phases such as functionalised and unfunctionalised silica like in NPLC, but uses polar solvents similar to RPLC solvents (95,101–106). Over the years, many HILIC stationary phases were developed commercially for different applications and are constantly growing. However, the C_{18} phase is still considered the most adaptable stationary phase available. HILIC mobile phases are generally organic solvents miscible with water and usually would have a lower percentage of water in them throughout the separation in comparison to RPLC. Like RPLC, HILIC separations are established by either a gradient or an isocratic elution, for the latter the percentage organic would be high whereas for the gradient elution starts with a high percentage organic and increasing aqueous percentage as it progresses to end with high aqueous. The common assumption about the HILIC retention mechanism is a liquid/liquid extraction system where the

compound is distributed between two layers; the ‘water-rich’ layer that the mobile phase creates on the polar stationary phase against a ‘water-deficient’ mobile phase, see Figure 1.5 (95,100,107).

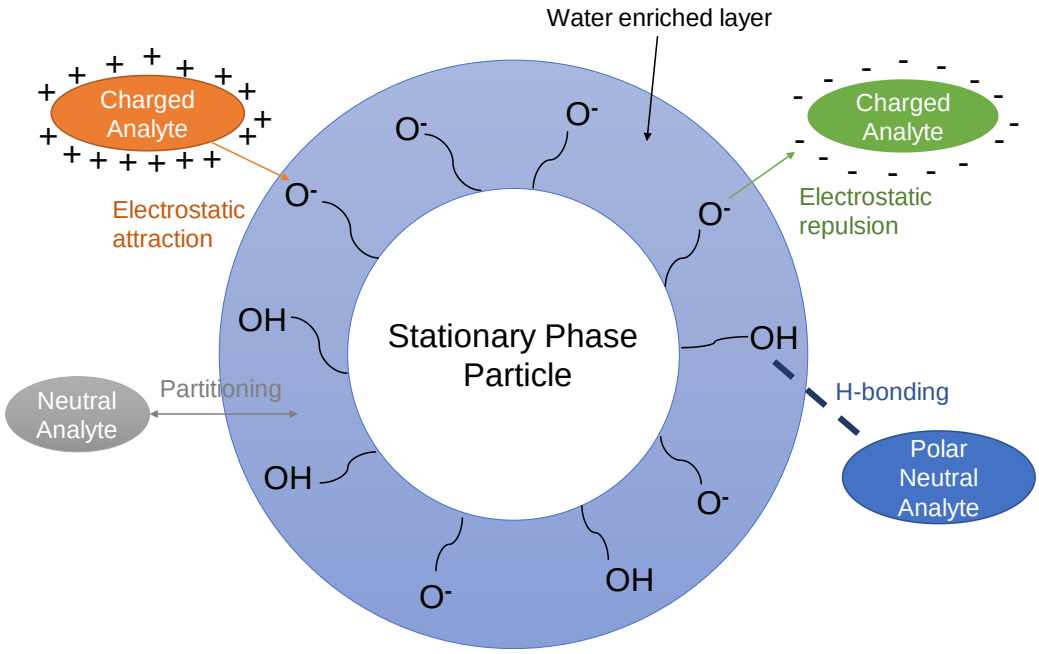


Figure 1.5 Schematic of interactions between HILIC stationary phase with the water enriched layer and different types of polar analytes. Adapted from (95,108).

Nevertheless, the retention and separation mechanisms of LC can often be a combination of different types of interaction; chemical (e.g., hydrogen bonding, electrostatic), physical (e.g., dipole-dipole), hydrophobic interactions (e.g., van der Waals forces). The separation mechanism has been proposed to in three models: (a) partitioning of the compounds between both the stationary and mobile phases (95,109,110); (b) reversible electrostatic/polarity-based interactions with the stationary phase (111,112) and; (c) a combination of preferential adsorption of the organic mobile phase onto the adsorbent stationary phase followed by compound partitioning into the adsorbed mobile/stationary phase layer (113).

Elution of compounds from the stationary phases is typically achieved by three different approaches depending on the resolution and separation of the compounds. These three approaches are: a) isocratic elution, mobile phase composition is constant throughout the method; b) gradient elution, the composition of one component of the mobile phase is increased gradually and c) step elution, when the composition of the mobile phase is changed periodically (88,93).

1.7 Mass spectrometry

Mass spectrometry has been a fundamental analytical technique since the early 1900s. In 1913, J.J. Thompson studied the effects of electric and magnetic fields on ionised neon particles (114). His work was the basis on which Francis Aston built the first mass spectrometer in 1919 that enabled him to identify the naturally occurring isotopes of compounds which are not chemically separable but only mass resolved using mass spectrometry (MS) (115). Ever since, the mass spectrometer has been used in most of the sciences to separate ionised species based on their mass-to-charge ratio (*m/z*). Figure 1.6 represents a schematic of the basic components of a mass spectrometer including an ionisation source, a mass analyser and a detector. The sample is introduced to the mass spectrometer through the sample inlet. The sample inlets have several forms, among them are direct infusion of the sample using a syringe pump or the eluate of an additional separation technique such as GC or LC. Using the latter combinations is referred to as hyphenated MS techniques.

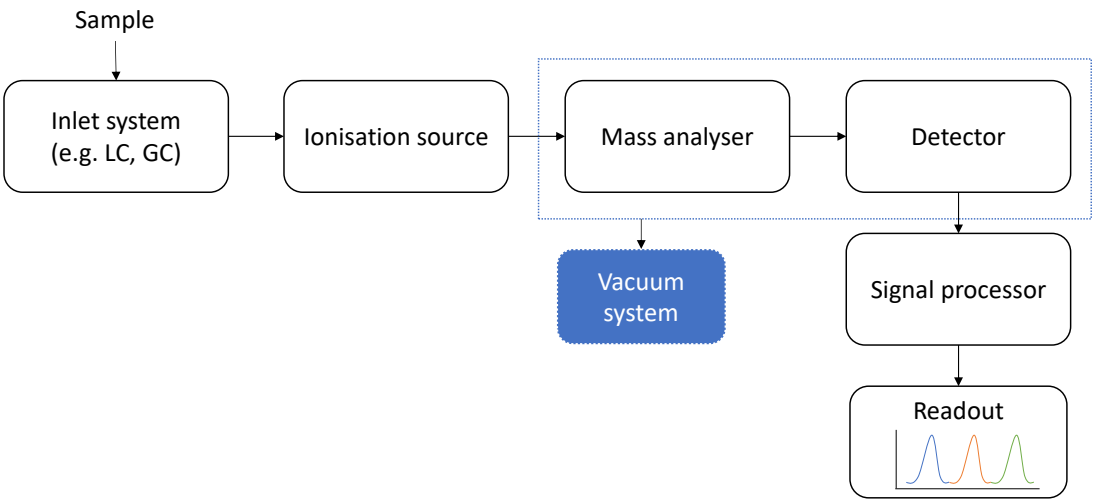


Figure 1.6 Schematic of the MS components showing the positioning of the inlet system followed by the basic components of the mass spectrometer and an example read out chromatogram. The ionisation source is typically at ambient temperature whereas other components are under vacuum.

1.7.1 Ionisation: electrospray ionisation

To enable the ions to be separated by an electric or magnetic field, an optimum path free of other molecules is needed. Therefore, MS instruments are designed to operate under vacuum. Hence, for a sample to be analysed by MS, it needs to be ionised and must be transformed from its physical state, i.e., liquid or solid to a gas phase. The ionisation of the sample takes place at the ionisation source. There are numerous types of ionisation sources that are used in mass spectrometry, but only the electrospray ionisation (ESI) source is discussed here as it was the source used to perform the work for this research. ESI was selected as the ionisation source of choice as it is easily coupled to LC, and because the compounds that were analysed in this work were readily ionisable in positive mode ESI.

In 1985, 5 years after Dole *et al.* introduced electrospray ionisation, Fenn *et al.* adapted their work to couple LC to MS (116,117). This was the most impactful advancement in the MS field in the past century and a Noble Prize

was awarded in 2002 to Fenn and his team. The ESI source can be operated in either negative or positive ionisation modes. Figure 1.7 illustrates a positive ionisation process.

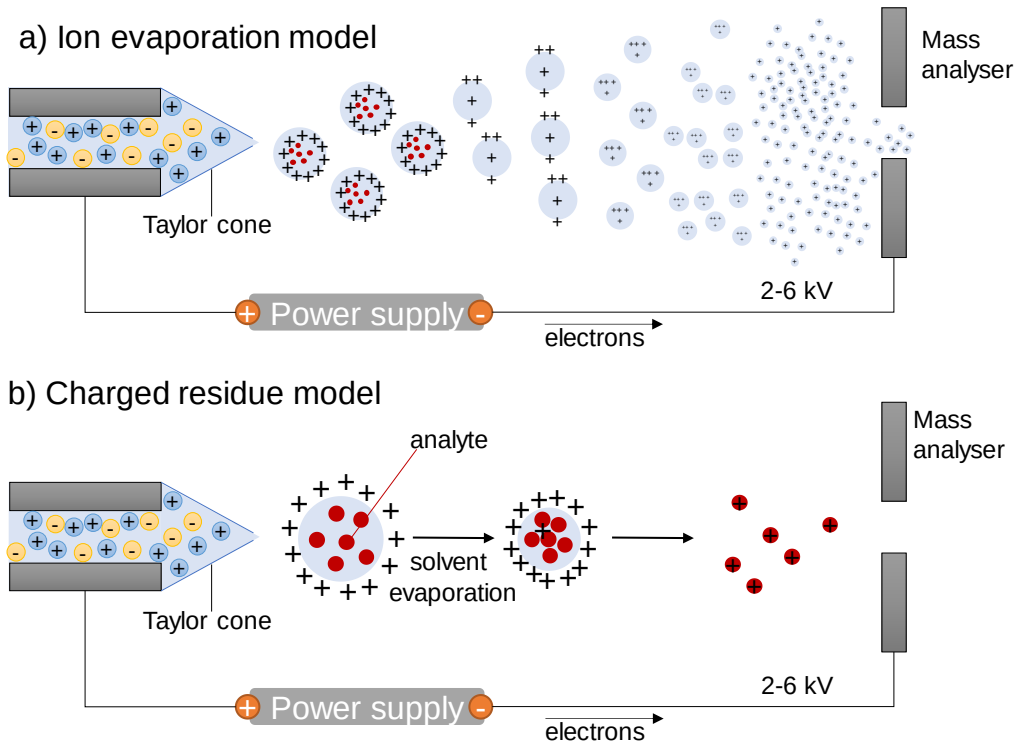


Figure 1.7 Schematic of positive ion mode ESI spray solvent evaporation in two ionisation models; (a) ion evaporation model and (b) charged residue model. In both models Taylor cone and formation of gas phase ions is demonstrated. Adapted from (118,119)

A sample in liquid form, usually an eluate of a LC system or through direct infusion by syringe pump, is introduced into the ionisation source at atmospheric pressure. The molecules in the eluate pick up the charge by several potential mechanisms among them could be a pH effect, a buffering effect in the liquid phase or by an electrochemical effect at the tip of the capillary when the sample passes through the very narrow conductive capillary, on which a potential difference of 2 to 6 kV is applied. The ions in the solution then get to the gas phase through electrospray ionisation which starts at the ESI capillary. At the capillary tip, a thin of the eluate Taylor cone

of liquid is formed (Figure 1.7). When the density of the charge in the liquid is high enough, the Taylor cone is shattered into droplets, forming the ‘spray’. The spray droplets move in the ionisation chamber influenced by the electric field between the capillary needle and the sampling cone at atmospheric pressure. The source chamber has an inert and heated desolvation gas applied causing the evaporation of the solvent in the droplets. As the radius of the droplet decreases, the Coulombic repulsion becomes greater than the surface tension causing Coulombic fission once the Rayleigh limit is surpassed. This incidence occurs several times to produce very small charged droplets. There are two models that are thought of as to how gas-phase ions are produced: a) the ion evaporation model which bases the formation of gas-phase ions on direct ion emission that would take place once a droplet radius is ≤10 nm; b) the charged residue model which is based on continued Coulombic fission of the droplet to the point that only a single macromolecule remains in the gas ion form. The first model is believed to be applicable for small molecules forming ions whereas the latter is presumed to be the basis for the ionisation of proteins and macromolecules forming gas-phase ions (116,118,120–123).

ESI is classified as a ‘soft’ ionisation technique where molecules are desorbed from the liquid with limited fragmentation due to lower the energy being applied in the ionisation process. Although the response of ESI could be saturated at high concentration of analyte, it is still considered ideal for quantitative analysis as the response is linearly related to the mole content of the analytes present in the source rather than the flow rate (116). With the advances in the manufacturing of ESI, signal stability has improved significantly. What remains an important issue however is ionisation

enhancement and suppression hence there is a crucial need for the use of internal standardisation in mass spectrometry (85,123,124). Moreover, the ionisation process is a competitive one where all ionisable analytes in the ionisation source are prone to pick up the electric charge. Hence, the background noise, coming from other matrix components, influences the sensitivity thus the limit of detection of a compound in a complex matrix is influenced by the background noise due to the competitiveness in ionisation between noise ions and target analyte ions. When analysing analytes in matrices, this phenomenon in the ionisation is called matrix effects which could be ionisation suppression or enhancement. Ionisation suppression is the reduction in the signal of target analyte due to the other matrix components and ionisation enhancement is the increase in the signal of the target analyte due to charging effects of other components co-eluting (118,124,125). Improving ionisation could be achieved by using a better chromatographic separation to minimise matrix components with the target analyte. Additionally, a mass analyser of higher mass resolution could minimise chemical noise caused by co-eluting compounds and hence improved overall signal to noise ratio (see 1.7.2).

1.7.2 Mass analysers: triple quadrupoles

There is a wide range of mass analysers that is used in mass spectrometers including the quadrupole ion trap, time-of-flight, magnetic sector, ion cyclotron and others. The choice of the suitable mass analyser is crucial and is based on the precise aim of the analytical method. The several factors to consider when selecting a mass analyser include: sensitivity, resolution, selectivity and scanning speed. For this research, the aim is high accuracy quantification

hence the most important factors are sensitivity and selectivity. The speed of acquisition is often sacrificed when developing higher-order RMP hence it was not considered a primary criterion in the selection of the mass analyser. Therefore, the mass analysers used in this work were triple quadrupoles, the theory of which is now described.

The triple quadrupole mass analyser consists of three parts: a first quadrupole, a collision cell and a third quadrupole (see Figure 1.8).

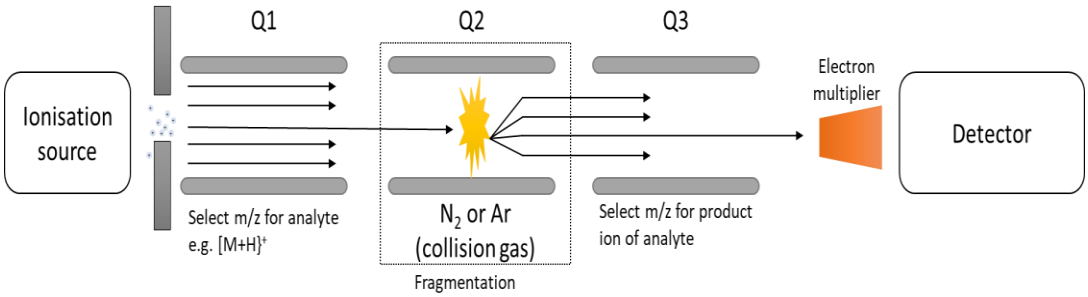


Figure 1.8 Schematic of triple quadrupole mass analyser with a demonstration of the SRM transition acquisition mode where a single ion is selected by Q1 to undergo fragmentation in Q2 and in Q3 only the selected product ion is captured.

1.7.2.1 Quadrupole

A quadrupole is a mass filter that consists of four hyperbolic parallel metal rods that are electrically coupled in pairs. The two pairs are then situated to form a cylinder like shape with a passage through the centre (see Figure 1.9). An alternating radio frequency (RF) potential is applied to the rods with a static direct current (DC) potential offset, resulting in one pair being positively charged and the other, negatively charged. As ions enter the quadrupole they are drawn to the pole of opposing charge, which is quickly switched in polarity, repelling the ion, enabling the ion to traverse the quadrupole to the detector. Therefore, ions are transmitted through the quadrupole passage and are

influenced by the potential to either go through to the detector or are neutralised by hitting the rods and hence not detected (see Figure 1.9). On the other hand, if the potential is set to scan, many ions over a range of *m/z* can pass through sequentially.

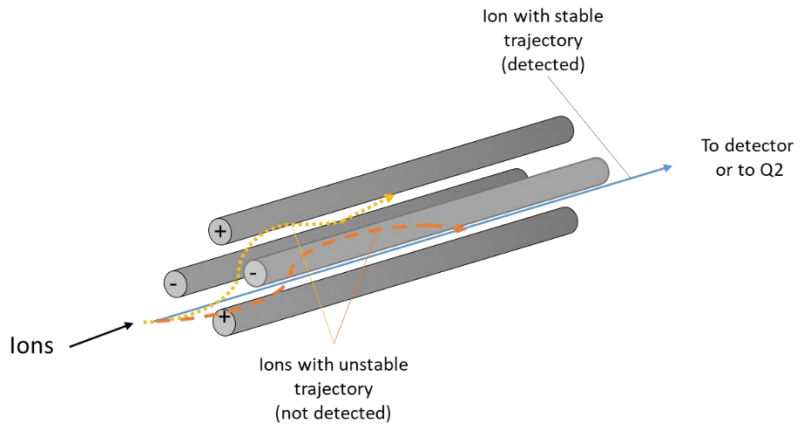
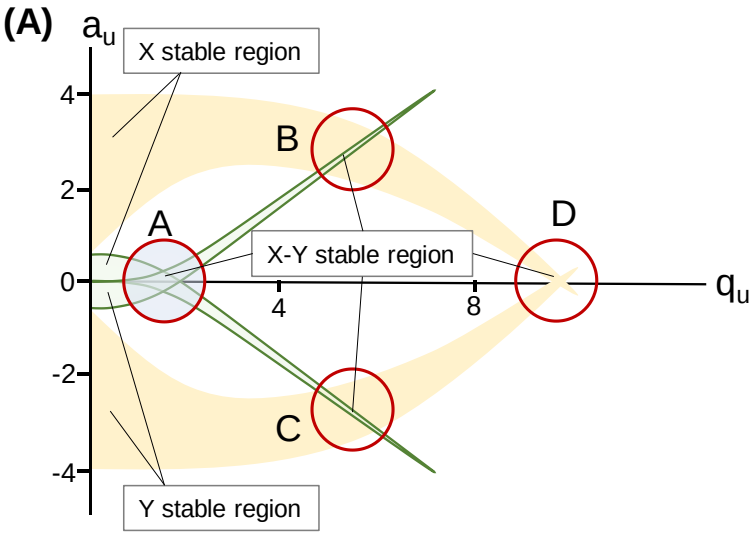


Figure 1.9 Schematic of quadrupole analyser selecting sing *m/z* stable ions (blue) to pass through while non-stable ions (yellow and orange) are neutralised by colliding with the quadrupole rods. Adapted from (119).

By manipulating the DC potential and the RF frequency, control over the ion trajectory is obtained. As such a single ion could be selected to pass through to the detector by fixing the DC and RF allowing the quadrupole to perform as a mass filter. The ratios of the RF and DC voltages through which the ion motion is stable can be represented by the Mathieu diagram and equations (see Figure 1.10.A and Figure 1.10.B). For an ion to make it through the quadrupole, it needs to be stable on the X and the Y axes. There are several regions where the stability zones of both X and Y overlap. The first stability region (A) is where the voltage requirement is the lowest where Figure 1.10.C shows a closer view of the region. The stability region is dependent on the ion's *m/z*. Typically, when scanning with a quadrupole over a range, the RF and the DC are ramped linearly with time. As the voltages are increased the

scan line cuts through the stability regions and the ions of that particular masses are transmitted through the quadrupole to the detector, see Figure 1.10.C. A plot is then curated as signal vs time which is the mass spectrum. The masses are resolved from each other when the scan line does not cross two stable regions at the same time. The DC and RF ratios can be increased so that the scan line passes through the tips of the stability regions which results in the reduced peak width hence the resolution is increased (119,126).

The operator selects the *m/z* range required; the lower limit is typically set to *m/z* 50 while the upper limit of most quadrupoles is *m/z* 4000. Based on the filtering capabilities of the quadrupoles, sequencing more than one quadrupole provides enhanced selectivity and specificity. When operating a triple quadrupole mass analyser in a selected reaction monitoring (SRM) mode a single *m/z* is allowed through the first quadrupole. This ion is then transmitted into the collision cell to undergo fragmentation (fragmentation or collision induced dissociation mechanism is described later in 1.7.2.2). The fragments produced in the cell are then passed through to the following 'third' quadrupole where a single product *m/z* is selected to pass through the detector which provides a high selectivity and specificity when quantifying compounds (119).



(B)

$$a = \frac{8zU}{m\omega^2 r_0^2} \qquad q = \frac{4zV}{m\omega^2 r_0^2}$$

X and **Y** are the distances along the axises
U *a* resolving DC voltage
V *is the applied voltage*
z the charge of the electron
ω *is the frequency*
m mass of the ion
r_0^2 is the effective radius of the quadrupole from the centre to the rod surfaces

(C)

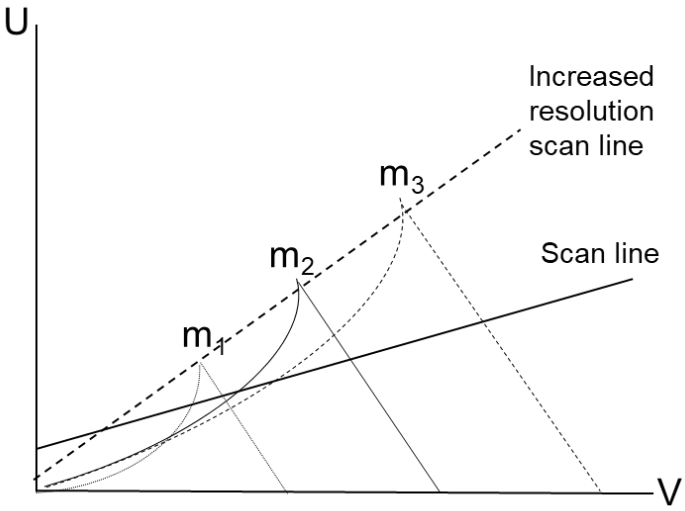


Figure 1.10 Mathieu equations and diagrams on the stability of ion trajectory through the quadrupole. (A) Mathieu diagram noting the stable regions A, B, C and D; (B) Mathieu equations (C) Scan line trajectory in the first stability region A. Adapted from (119).

1.7.2.2 Collision cell

The collision cell is located between the first and third quadrupole in the triple quadrupole mass analyser. The cell contains an inert gas as pressure of 10^{-4} to 10^{-2} torr (127). The gases used usually are helium, argon or nitrogen. The collision cell typically has a RF current that helps keep the ions focused during travel through the collision cell. The collisional activation of the ions is achieved by accelerating the ions to collide with the gas using a potential difference leading to increase of internal energy of the compound and fragmentation by chemical bond breakage. This process is referred to as collision induced dissociation (CID) (128,129). In SRM acquisition mode, the first quadrupole would be set to allow a specific *m/z* (precursor ion) only to pass through to the collision cell and the third quadruple would be set to allow a specific *m/z* (product ion) only to reach the detector.

The first CID spectra were reported in 1968 by Jennings *et al.* and by Haddon and McLafferty. These spectra were a result of an accidental observation of a faulty system where air leaked into the vacuum chamber of the mass spectrometer and caused collisional dissociation of some ions. These incidences and subsequent publication founded the basis for tandem mass spectrometry (130,131). MS/MS is what enabled structural elucidations and quantitative analysis to advance drastically (128).

1.8 Isotope dilution mass spectrometry (IDMS)

MS is not an inherently quantitative technique hence it requires the use of an internal standard with a calibration system to compensate for any losses in the sample preparation process and instrumentational effects such as matrix effects (i.e. suppression and/or enhancement) in the ESI ionisation process (119). There are several types of internal standard and calibration approaches used. More information about these and the available choices for each are covered in 1.10.2 and 1.10.3.

IDMS is the basis of a primary ratio method (see 1.1) that can achieve the highest attainable metrological order when performed correctly. Accurate and precise SI-traceable results that have small measurement uncertainty are obtained by accurately performed double exact-matched IDMS and using traceable materials (15). IDMS is based on measuring the ratio of the compound to its stable isotopically-labelled internal standard. Stable isotopically-labelled internal standards (SILIS) are identical compounds to the target analyte and have virtually identical physicochemical properties apart from in mass. This made SILIS the most accurate internal standards as they mimic the physiochemical properties of the target analyte. The mass difference is due to enrichment with heavier atoms that include ^{13}C, ^{15}N or ^{2}H. These atoms are usually used in the production of SILIS as their natural abundance is low. Therefore, the compound and its 'labelled' analogue compound should theoretically behave identically throughout the sample preparation and chromatographic separation but are separated by MS but this is not always the case. During the LC separation, SILIS should ideally elute at the same time as the natural compound, however; SILIS that are labelled with

deuterium only in particular were found to elute slightly earlier with a small difference in t_R (132). This was explained by Wieling by the difference in physiochemical properties between hydrogen and deuterium atoms where the later form stronger binding with carbon atoms (133). SILIS are added to the sample at known quantity and as early as possible in the sample preparation procedure. It should also be 'equilibrated' in the sample to ensure identical behaviour within the matrix as does the target compound. Therefore they do not only account for any losses in the sample preparation process, but also for any instrumental variation assuming they have the same extraction efficiency (85,134,135).

For a labelled compound to be used as a SILIS, it needs to be enriched with isotopes having at least three mass units difference from the natural compound to avoid overlapping with the natural isotopologue of the analyte and causing interference in the monitored SRM channels. Ideally, there should be no interferences between the compound and its SILIS compound. ^{13}C and/or ^{15}N are generally preferred to deuterated internal standards because they do not suffer from kinetic isotopic effects to the same extent. However, deuterated internal standards are more readily available due to their ease of production and lower prices compared with the ^{13}C and/or ^{15}N standards. The difference in LC retention of deuterated internal standards is of high importance when analysing an analyte in a complex matrix (e.g., plasma) as the matrix effects (e.g., ionisation suppression) would differ depending on the type of co-eluting components in that region of the chromatogram (134).

Equilibration of SILIS in the sample is crucial to obtain accurate quantitative results. It is of high importance to understand how the labelled compound

behaves in the matrix. For example, if the target compound is highly bound to red blood cells (e.g., tacrolimus), it is important that the internal standard is equilibrated in the sample and allowed to bind/interact with the matrix components. This ensures identical behaviour of the compound and its labelled equivalent. There are several approaches to applying IDMS. Each approach has its own use and manifests a different level of complexity. A reference method usually applies double exact-matching IDMS where high accuracy, high precision and low measurement uncertainty can be achieved. However, such a method requires longer periods of time and greater expertise. Equally it would not be used for larger numbers of samples, due to long times required for gravimetric sample preparation, long instrument run times which ultimately results in increased cost. Routine clinical testing laboratories usually use single IDMS (see 1.8.1) when isotopically-labelled internal standards of the compounds are available. Figure 1.11 summarises the some approaches of IDMS that are further explained in this section (135).

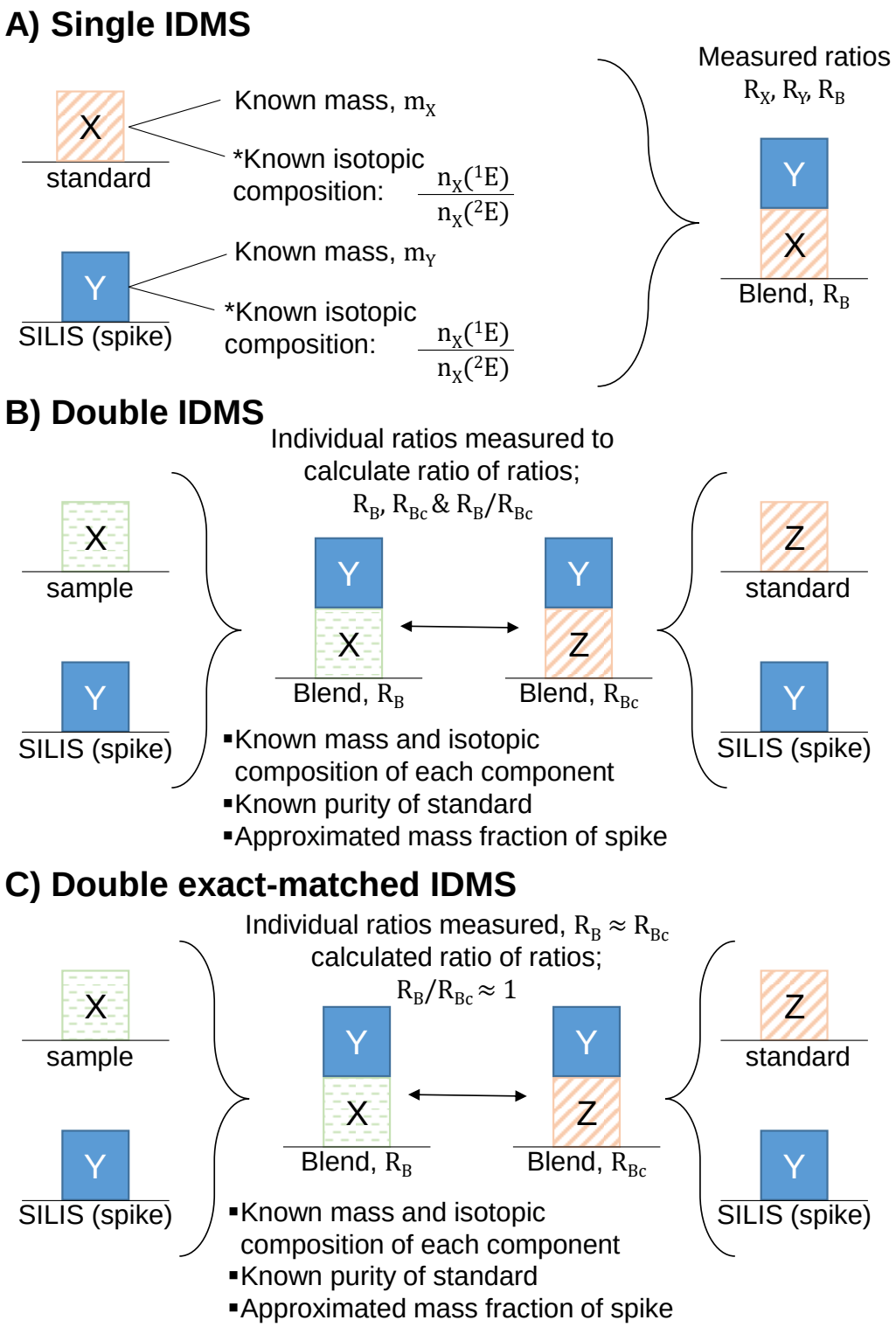


Figure 1.11 Summary of different blends prepared in different IDMS calibration approaches, showing the importance of knowing the SILIS mass fraction in single IDMS whereas with the use of double IDMS approaches measuring by an independent calibration blend the SILIS mass fraction is approximated and does not affect the measurement (given there is no isotopic interferences in the SILIS).

1.8.1 IDMS Equations

This section is a prompt to the subsequent sections where the different approaches to IDMS are described. In the introduction to the different IDMS equations that will follow, below is the convention of notation that is used to describe the different solutions and blends in these equations:

- Sample: index X
- Spike/SILIS: index Y
- Primary analyte standard/calibration standard: index Z
- Sample blend (i.e., sample + SILIS): index $B=X+Y$
- Calibration blend (i.e., primary standard + SILIS): index $Bc=Y+Z$

These equations were derived from the field of inorganic analysis. When applying IDMS to inorganic analysis. There is an obstacle that organic analysis does not suffer from. That is the isotopic purity of isotopically enriched/labelled compounds. As an example, when an inorganic element such as silver (Ag) is to be measured, the natural abundance of this element to the isotopically enriched element is 48 %:52 % of ^{107}Ag:^{109}Ag. On the other hand, carbon isotopes for example, occur in nature as ^{12}C:^{13}C:^{14}C in 98.9 %:1.1 %:<0.0001 %, similarly hydrogen isotopes ^{1}H:^{2}H occur in nature as 99.985 %:0.015 %. This allows the simplification of the IDMS equations when used for in the organic analysis field (136,137).

The concentrations (mol/L) are interchangeable with mass fraction (g/Kg) in these equations (136,137). The below notation was followed for the rest of this section and for the different equations:

- n_X: amount of substance in X (mol)

- n_Y: amount of substance in Y (mol)
- c_x: concentration of analyte in sample X (mol/L)
- c_Y: concentration of analyte in spike/SILIS Y (mol/L)
- c_Z: concentration of analyte in primary standard Z (mol/L)
- m_Y: mass of spike Y added to the sample X to prepare the blend B (g)
- m_X: mass of sample X added to the spike Y to prepare the blend B (g)
- m_Z: mass of primary standard Z added to the spike Y to prepare the blend B (g)
- m_{Zc}: mass of primary standard Z added to the spike Y to prepare the calibration blend Bc (g)
- m_{Yc}: mass of the spike Y added to primary standard Z to prepare the calibration blend Bc (g)
- R_B: isotope amount ratio of sample blend B
- R_{Bc}: isotope amount ratio of calibration blend Bc
- R_X: isotope amount ratio of sample X
- R_Y: isotope amount ratio of sample Y
- R_Z: isotope amount ratio of sample Z

1.8.2 Single IDMS

Single IDMS is the simplest IDMS approach and is the one that does not apply a calibration. It is based on ratio measurement. The IDMS ratio is a critical term to IDMS and is the basis of the measurement procedure. It refers to the ratio of the response of the analyte measured by MS to the response of the isotopically-labelled internal standard. In single IDMS, a known amount of SILIS is added to the sample that contains the analyte of interest at an unknown mass fraction, i.e., a quantity expressed by weight per weight, e.g., ng/g. The SILIS addition is referred to as the ‘spike’. It is added to the sample early in the workflow and should be equilibrated in the sample. The ratio of analyte:spike remains constant through the workflow (assuming the SILIS is stable and of high purity); hence sample handling would effects are considered compensated and hence not influencing the accuracy of the measurement. The ratio is used to deduce the mass fraction of the unknown analyte. Nevertheless, the ratio is susceptible to mass bias occurring in the

mass spectrometer. IDMS has a fully defined measurement equation, Equation 1.1. For single IDMS to be applied, the isotopic purity of both the analyte and its corresponding SILIS needs to be known. Moreover, the mass fraction of the spike needs to be known. As mentioned earlier, SILIS are expensive to produce and for purity analysis fairly large amounts are needed e.g. 10-50 mg for quantitative nuclear magnetic resonance (qNMR) purity analysis (135).

Equation 1.1 The single IDMS equation used to measure the mass fraction of sample X.

$$c_X = c_Y \cdot \frac{m_Y}{m_X} \cdot \frac{R_Y - R_B}{R_B - R_X} \cdot \frac{R_X + 1}{R_Y + 1}$$

1.8.3 Double IDMS

Double IDMS is more complex than single IDMS and requires more effort and time. However, it not only produces lower measurement uncertainties but additionally helps in overcoming the need to know the pure mass fraction and isotopic distribution of the SILIS (138). This is due to the use of a calibration blend to measure against. In double IDMS two ‘blends’ are prepared. The ‘sample blend’ (SB) and the ‘calibration blend’ (CB); both blends would be prepared to contain the exact amount of SILIS. A sample blend, like in single IDMS, would contain the sample with the SILIS addition. The calibration blend, on the other hand, would be prepared by the addition of a well characterised reference standard of the analyte to the SILIS (135). The mass fraction of the analyte in the SB is then determined by the ratio of the ratios of both blends. Hence, the accurate isotopic purity and mass fraction of the SILIS becomes irrelevant as the addition would cancel out in the measurement equation (139,140). Additionally, this simplifies the measurement uncertainty equation

to use mainly the values obtained from the preparation of the two blends and the accuracy of the ratio measurements performed by MS. The double IDMS measurement equation (Equation 1.2) would be applicable when: a) there is no unlabelled compound in the stable isotopically-labelled internal standard; b) there is no contribution of unlabelled standard isotopes to the signal of the heavy SILIS; and c) the isotopic ratios are equal $R_X = R_Z$.

Equation 1.2 Simplified double IDMS equation used to measure the mass fraction of sample X. This equation is used when the SILIS is of high isotopic purity.

$$c_X = c_Z \cdot \frac{m_Y}{m_X} \cdot \frac{m_{Zc}}{m_{Yc}} \cdot \frac{R_Z - R_{Bc}}{R_{Bc} - R_Y} \cdot \frac{R_Y - R_B}{R_B - R_Z}$$

1.8.4 Double exact matched IDMS (DEM-IDMS)

DEM-IDMS is based on the double IDMS approach, a SB and a CB are prepared using the same amount of SILIS in both. However, in DEM-IDMS the preparation of the blends is performed repetitively in an iterative manner until the measured ratio of analyte to SILIS by MS is equal to 1 (i.e., R_B = 1 and R_{Bc} = 1). Hence, the 'ratio of ratios' R_B:R_{Bc} would also equal 1. It would take several attempts to achieve that ratio of ratio to equal 1. This approach considers that both the signal intensities of the compound and its SILIS in both the SB and the CBs are equal so that any measurement effect occurs on both the analyte and the internal standard.

DEM-IDMS is considered a highly costly, lengthy and labour-intensive measurement procedure. DEM-IDMS was developed from double IDMS about 23 years ago to refine the measurement procedure and produce lower measurement uncertainty (141). DEM-IDMS is not used for routine measurement methods but to assign values of reference materials/standards

using higher-order RMP. Typically, DEM-IDMS used tighter bracketing of SBs by CBs where the CB is analysed before and after the SB, the mean ratio of the two measured ratios of the two CBs would then be used as R_{Bc}.

For organic compounds and their SILIS, the differences in isotopic enrichment can be controlled unlike the case for the analysis of inorganic elements where the isotopic enrichment is naturally occurring. This allows the use of the simplified version of the equation, Equation 1.3.

Equation 1.3 Simplified DEM-IDMS equation

$$c_X = c_Z \cdot \frac{m_Y}{m_X} \cdot \frac{m_{Zc}}{m_{Yc}} \cdot \frac{R_B}{R_{Bc}}$$

1.9 Measurement uncertainty (MU)

1.9.1 Measurement uncertainty and error

It is crucial to differentiate between measurement uncertainty (MU) and error. ISO 3534:2006 on Statistics' definition of Error is '*the difference between an individual result and the true value of the measurand*'. The same ISO standard defined the true value as: '*value which characterises a quantity or quantitative characteristic perfectly defined in the conditions which exist when that quantity or quantitative characteristic is considered'* (142). However, the concept of absolute true values is quite abstract, although it is the commonly used way of reporting a measurement. The ISO 3534:2006 and the Eurachem and the Co-operation of on International Traceability in Analytical Chemistry (CITAC) guide noted that '*the true value of a quantity or quantitative characteristic is a theoretical concept and, in general, cannot be known exactly*' (143). Additionally, the term "conventional true value" which is defined by IUPAC

Gold Book and the Guide to the expression of uncertainty in measurement (GUM) as: *'value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose'* (144,145).

Therefore, an observed error in a measurement indicates the difference between the measurement and the reference value. While error is also noted as an idealised concept that cannot be known exactly, if an error, theoretical or observed is known can be used as a correction factor to the measurement. On the other hand, MU, is an interval/range that is estimated for a measurement procedure or material/sample. MU cannot be used as a correction factor of a measurement (143).

1.9.2 Measurement uncertainty definition, sources and components

The MU is defined today as a *"non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used"* (146). MU expresses the confidence in a measurement by a MU estimate that is associated with a probability which is referred to as the confidence interval (CI). The GUM which was published in the mid-1990s by ISO is considered the master document on measurement uncertainty. Subsequently, several supplementary documents and an updated document of the GUM were published to further explain and improve the ease of use of the GUM (147,148).

The uncertainty on a measurement result could result from many sources, these include; the measurand's incomplete definition, reference values,

interferences, matrix effects, uncertainties of masses, environmental conditions, any assumptions made in the measurement procedure and random variation (146). The MU estimate should encompass each potential source of uncertainty in the measurement procedure. Among these sources could be the variability in instrumentation performance, accuracy and linearity of balances used in gravimetric preparation, the purity of standards used and many others. For every measurement procedure it is crucial to identify the sources of uncertainty of each step of the process. These would then be combined to provide a final MU estimate (149).

Since the final MU estimate is constituted of the different uncertainty components, to calculate a MU estimate each of the components needs to be identified and must be quantified. The quantitative estimation of the individual uncertainty components is expressed as standard uncertainty. This can be obtained in several ways depending on the MU component. For example, from the certificates of reference standards or can be calculated statistically using the standard deviation of repeat measurements. The final MU is then calculated using the measurement equation in addition to combining the standard uncertainties. This is performed as a square root sum of the squares of the values and their respective standard uncertainties (see Equation 1.4).

1.9.3 DEM-IDMS measurement uncertainty components

In the case of DEM-IDMS, the components to MU are those used in the measurement equation (Equation 1.4). These are mainly:

(a) the purity of the reference standard (u_{cZ});

(b) the gravimetric preparation of blends (u_{mX}, u_{mY}, u_{mYc}, u_{mZ},) which is typically the smallest component of the uncertainty budget due to the accuracy of advanced modern analytical balances used; and

(c) the precision of the ratio measurements obtained from the MS (R_{Bc}, R_B). The latter is usually the largest component of the uncertainty budget and could be reduced by repeat measurements. However, in cases where the precision of the ratio of measurements is minimal and the purity of the standard could become the largest component.

Finally, the total relative uncertainty of the mass fraction of an analyte (c_x) may be calculated using the equation below (see Equation 1.4) (149).

Equation 1.4 Total relative measurement uncertainty of the amount of substance in a sample measured by DEM-IDMS (135,143,148)

Uncertainty of amount of substance content of the sample	Uncertainty of mass fraction of standard	Uncertainty of mass of sample in sample blend	Uncertainty of mass of spike in sample blend	Uncertainty of mass of standard in calibration blend	Uncertainty of mass of spike in calibration blend	Uncertainty of ratio of isotope ratios of blends

$$u(w_x) = \sqrt{\left(\frac{u(w_z)}{w_z}\right)^2 + \left(\frac{u(m_x)}{m_x}\right)^2 + \left(\frac{u(m_y)}{m_y}\right)^2 + \left(\frac{u(m_z)}{m_z}\right)^2 + \left(\frac{u(m_{yc})}{m_{yc}}\right)^2 + \left(\frac{u\left({}^{R_B}/_{R_{Bc}}\right)}{{}^{R_B}/_{R_{Bc}}}\right)^2}$$

mass fraction of sample	mass fraction of standard	mass of sample	mass of spike in sample blend	mass of standard in calibration blend	mass of spike in calibration blend	Ratio of isotope ratios of blends

As mentioned above the MU is reported with its associated probability/confidence. The above equation gives a total MU representing one standard deviation that is 68 % confidence that the reported measurement will be in this range. When results are assumed to be of normal distribution, a coverage factor k (typically k = 2, depending on the number of degrees of freedom) is used to ensure that the MU covers the largest percentage of the potential values, resulting in an expanded measurement uncertainty (U) at a confidence interval (CI) of 95 %. The coverage factor definition by International Vocabulary of Metrology (VIM) is '*a number larger than one by which a combined standard measurement uncertainty is multiplied to obtain an expanded measurement uncertainty'* (8,150). When reporting a measurement, it is reported with its U; *X ± U units at 95 % CI* e.g., 5.3 ± 0.21 ng/g at 95 % CI.

It is common to discuss what would be the ideal MU estimate, however, the best estimate is the one that is 'fit for purpose'. This is dependent on how the method will be used and what the requirement of a measurement procedure. As an example, for an analytical method performed in a routine hospital laboratory the MU estimate could be as high as 30-40 % due to variability in biological samples. Such high MU estimates could potentially be fit for purpose depending on the clinical relevance of the test. On the other hand, to produce a 'reference value' for a higher-order CRM that would be at the top of the traceability chain (Figure 1.1), the MU estimate must be as low as analytically and technically possible. Reducing the MU estimate of any method

even by small amounts requires tremendous time, effort and expertise (143,149).

To assign a MU estimate in the case of a routine hospital laboratory, the decision would be based on the balance between available resources and what is clinically 'fit for purpose'. Over the years, hospital laboratories were reporting values without their MU estimates. As UK hospital laboratories have moved towards ISO 15189:2012, they are now required to report the estimated MU of their results in line with the ISO standard. The ISO statement is *"the laboratory shall determine measurement uncertainty for each measurement procedure in the examination phases used to report measured quantity values on patients' samples. The laboratory shall define the performance requirements for the measurement uncertainty of each measurement procedure and regularly review estimates of measurement uncertainty"* (ISO 15189:2012) (151,152). Figure 1.12 is an example to demonstrate the importance of reporting measurement uncertainty when interpreting comparative measurements. The data shown in Figure 1.12.A where measurement uncertainty is not reported could be misinterpreted as the data reported by laboratory C would be assessed as positively biased. However, when measurement uncertainty is reported (Figure 1.12.B), where the measurement uncertainty bars overlap with the reference value, it shows agreement of data. This would influence major decisions in the analytical procedure is the samples analysed are QC materials for example, a batch of analysis could be failed because of a failed QC measurement. This is a

practical demonstration of what the consequences of not reporting measurement uncertainty could be.

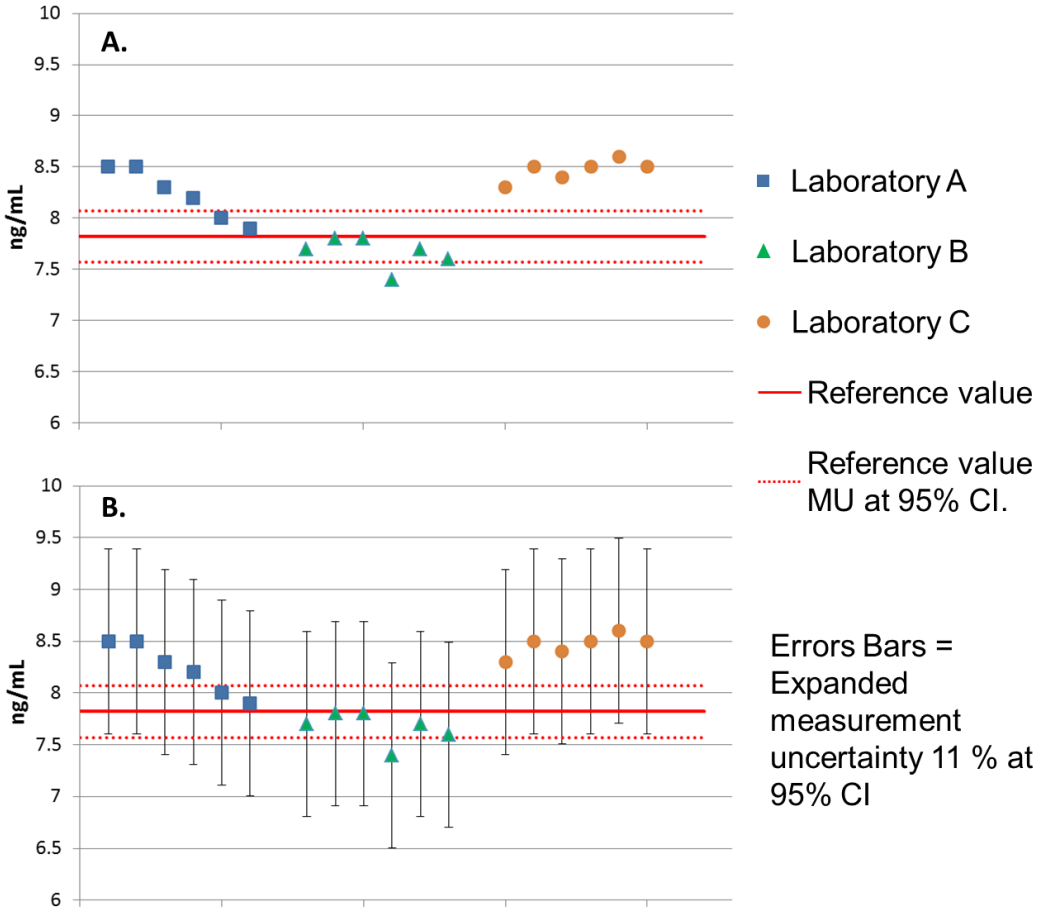


Figure 1.12 Representation of reporting measurement data with (A) and without (B) reporting measurement uncertainty. The numbers in the figure are only a visual representation of the concept (not real data).

1.10 Sources of variability in LC-MS measurements

Despite the great advantages that LC-MS brings to the clinical laboratory, it carries a set of challenges that need to be addressed. Errors and inaccuracies in the analytical procedure could take place at many stages. These could be classified into two categories: quantitative analytical errors and manual handling errors (due to human error) (85). The latter class of error is mainly

due to lack of automation of many LC-MS applications. Figure 1.13 summarises some examples of the potential manual errors. Other examples would include dispensing volumes for SPE, SPE recoveries, reproducibility in preparation of buffers and standards, retention time variance due to eluent preparation. All of these are manual errors and not only by the user but potentially the manufacturers of SPE, standards and other consumables.

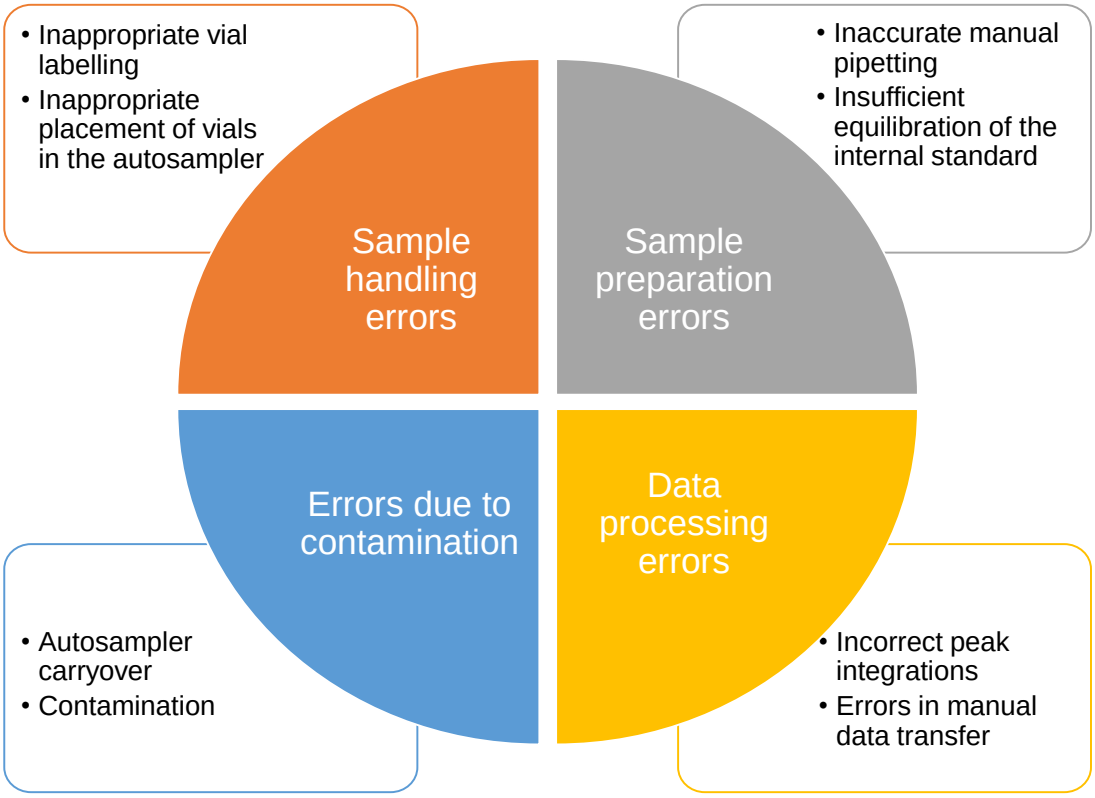


Figure 1.13 Examples of potential sources of error due to manual handling (human error)

The quantitative analytical sources (Figure 1.14) of inaccuracy are mainly divided between the ionisation source and the mass analyser of the mass spectrometer. The mass detection sources of inaccuracy all stem from inadequate ion selections for monitoring whereas the possible sources of variability during ionisation are numerous. As mentioned in the ionisation section (1.7.1) earlier the inherent poor signal stability of ESI mandates the

use of internal standardisation for compensation (85,124). Below is an overview of the most common sources of variability and error in LC-MS based methods.

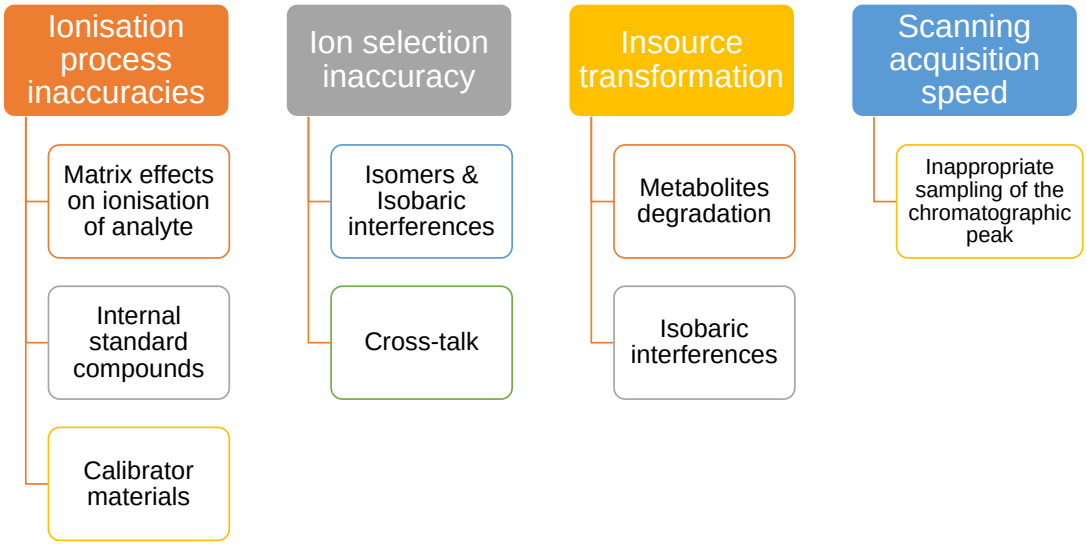


Figure 1.14 Chart summarising examples of potential sources of error due to MS effects

1.10.1 Ionisation suppression and enhancement

Matrix effect is the impact of the ionisation of other molecules on the ionisation of the target of molecule. The molecules present in the sample that are ionised, could be in the biological matrix and were extracted alongside the target analyte during the sample clean up (e.g., SPE or protein crash). These compounds are sometimes referred to as ‘co-extractives’ and they can cause either “ion enhancement” or “ion suppression”. During the initial stages of LC-MS method development the analyte is typically analysed as an analytical standard in solvent form where the solid powder standard is dissolved in a solvent that the molecule is soluble in. Impurities in water and the organic solvents (e.g. acetonitrile, methanol), that are used in mobile phases and sample preparation, could interact with the analyte in the ionisation

chamber (153–156). The analyst would therefore establish a baseline for the standard in solvent by comparing against an injection of the solvent without the analyte. The expression “pure standard in solvent” is used below based on this explanation.

Therefore, ion suppression is the scenario that is observed when the signal of the analyte is lower when it is ionised in a matrix extract than when it is ionised in a pure solvent of equal concentration. When the situation is the opposite i.e. the signal of the pure standard in solvent is lower than that of the analyte in matrix it is referred to as ion enhancement (85,157,158). This competitive ionisation is a result of the relatively small electric current available in the ionisation source where the charges are taken up by all the ionisable molecules. Other reasons for ion suppression or enhancement could be hydrophilic molecules and salts that would compete in the ionisation process as well. Ion suppression is a common issue and is the reason for the need for extensive sample clean-up of the complex biological samples to isolate the target analyte. Among the many sample preparation and clean-up techniques are SPE, liquid-liquid extraction (LLE) and protein precipitation (85).

The evaluation of matrix effect is performed usually by one of two approaches: post-extraction spiking and post-column infusion. Post-extraction spiking is performed by spiking an extract of a biological sample with a small amount of a relatively high concentration standard in solvent after the sample preparation is completed and comparing the MS signal of the post-extraction spiked material to a solvent standard of equal concentration (85,123,159,160). This gives a semi-quantitative estimation of the matrix effect. Another approach to

assess matrix effects is using post-column infusion which provides a qualitative assessment of the matrix effects. In this procedure, a blank matrix extract (not containing the analyte) is injected by LC using the LC method of the analysis with a post-column infusion of the analyte in a solvent standard using a T-piece. The solvent standard will provide a constant signal throughout the LC run whereas the matrix co-extractives would disturb the signal (suppression or enhancement) for periods of times as they elute from the LC. This will help know where in the LC gradient program is the highest suppression and would assist the analyst in manipulating the elution of the target analyte to avoid eluting in the suppression region (85,123,159,160). Therefore, a more thorough evaluation would include both post-extraction spiking and post-column infusion. However, the selection of post-extraction spiking and post-column infusion is dependent on the aim of the analysis and whether the assessment of matrix effects needs to be relatively quantified.

1.10.2 The choice of internal standardisation

It was discussed in the isotope dilution mass spectrometry section (see 1.8) why SILIS are considered the best to be used with LC-MS and why ^{13}C-labelled internal standards are better compounds to use than deuterium-labelled compounds. However, in cases where the choice of the SILIS is incorrect, inaccuracies in the measurement occur. As an example in carvedilol analysis, the deuterated internal standard suffers from hydrogen-deuterium exchange while in the ionisation chamber (132). Moreover, some deuterated internal standards do not co-elute with the target analyte (161) and hence, matrix effects on each is different, or in other cases (e.g. piperaquine in

plasma) the deuterated SILIS was suppressed significantly more than the target analyte (162,163). Both cases are further evidence that ^{13}C internal standards are superior to deuterated internal standards (85). Such cases may lead to major errors in the quantitative analysis because the basis on which the measurement is performed is the ratio of the peak area of the analyte to the peak area of the SILIS assuming that they are behaving in the same way through the analytical procedure. Hence when using deuterated SILIS it is crucial to assess the internal standard compounds and their behaviour in the method. Moreover, in cases where SILIS for the compound is not available or is too expensive, hospital laboratories use analogues and structurally related compounds as internal standards. These internal standards are more susceptible to being possible sources of error in the measurement as the internal standard compound could behave differently to the target compound in terms of retention times, ionisation and stability (85,164).

1.10.3 Calibration and quality control materials

As with the choice of SILIS, it is crucial to select the appropriate quality control (QC) and calibration materials used in the clinical laboratory. In National Measurement Institutes (NMI) laboratories, calibration blends and QCs that are used for quantification and validation are usually either prepared in-house from higher-order traceable certified solid standards with known purities, or from higher-order CRMs available and produced by other NMIs. However, given how expensive these types of CRMs are and how laborious it is to prepare highly accurate calibration blends, hospital and clinical testing laboratories tend to rely on commercially available calibration and QC

materials, referred to in metrology as secondary reference materials. Therefore, it is fundamental for the clinical laboratory to carefully select calibration materials with similar matrix effects as those of the samples. There are many reasons why the commercially available materials may not exert the same matrix effects as samples and some of these are: (a) freeze-thaw cycles of the materials while patient samples are usually analysed fresh or fresh-frozen; (b) production processes of the material including virus inactivation or lyophilisation (85) and (c) whether the calibration material is spiked or is an endogenous compound in matrix.

Vogeser *et al.* reported instrument specific inaccurate results of tacrolimus measurements with different commercial materials which resulted in a systematic negative bias of clinical sample measurements where the matrix effects on the analyte were different from the analogue and structurally related internal standard (165). Furthermore, in individual patient samples specific matrix effects could take place. To identify these sample cases, it is recommended to compare the internal standard peak areas over a sequence of samples to spot any lower or larger peak areas in a specific sample where enhancement or suppression could occur. Such cases are usually in patient samples who are prescribed medications that elute at the same retention time as that of the internal standard and/or the analyte (85).

1.10.4 In-source transformation

Although ESI is considered a soft ionisation technique it could still cause fragmentation/dissociation of ions in the ionisation source as the result of breakage of weak bonds in the molecule. When a peak is observed in the

same SRM trace as the analyte, it suggests an isobaric interference or an interference due to in-source transformation. The term “in-source transformation” was put forward by Vogeser and Seger (85) to describe the situation where a compound metabolite is *transformed* to the target analyte in the ionisation source. This is sometimes observed with glucuronide and/or sulfate-conjugated metabolites, for example. It is hence important that the chromatography can resolve the peaks of the analyte and the metabolite. Identifying in-source transformation products is possible when running long chromatographic separations or UHPLC separations of the samples/QC materials that contain the metabolites. Suggested ways to control in-source transformation is to optimise the tuning of the ESI source to be specific to the target analyte and to retune and revalidate periodically (85).

1.10.5 MS-based assays challenges and compound identification criteria

To ensure that an analytical method is highly specific, it is vital to identify the compound and the ions used in the measurement methods with high confidence (166). In bioanalysis, the level of interferences that could be detected by the mass spectrometer is high due to the richness of the matrix. This makes the specificity of a method a crucial part of the validation process to trust the quantitative results. The interferences that could be detected are those with the same nominal mass to the target analyte or internal standard and are referred to as isobaric interferences. The compounds could be coming from the rich biological matrix, they could be structurally different from the target analyte or structural isomers of the compound with the same elemental

formula (e.g., testosterone and its epimer epitestosterone). Moreover, the interferences could be larger molecules that are multiply charged in the ESI source producing the same *m/z* of the analyte at the detector. These interferences could be problematic with low resolution MS instruments if they are not operated to resolve at one mass unit. The use of high-resolution mass spectrometer reduces the level of interferences as these instruments are highly specific and can differentiate between compounds of the same nominal masses but with different chemical structures unless they are isobaric or have the same elemental composition. The use of chromatography would be important to separate the latter two types before they reach the mass spectrometer. These instruments include mass spectrometers with different mass analysers such as time-of-flight (TOF), orbitrap, Fourier transform ion cyclotron resonance (FT-ICR) however these instruments are not currently commonly used in hospitals and other clinical testing laboratories because of their complexity and cost (85,119). With the cases of true isobars such as 11-hydroxycortisol and 21-hydroxycortisol, where the two compounds have the same elemental formula but a different structure, the mass spectrometer fails to separate them even at high resolution and, for these cases, chromatographically resolving the two compounds is required (85,167).

Tandem MS is commonly used due to its higher analytical specificity. Many of the MS instruments in clinical laboratories are triple quadrupole instruments with SRM scan modes as these are highly specific due to identifying the target analyte by two ions the precursor and the product ions from the fragmentation. Yet, SRM scanning mode is prone to isobaric interferences which have the

same SRM transition as the target analyte due to the complexity of biological matrices, many cases have been reported in the literature (83,84). Sporadically, in clinical samples, patients could be taking different medications that could represent isobaric interferences but are normally evaluated routinely during method development and validation stages (85).

The different approaches to identify isobaric interferences in SRM transitions include acquiring a minimum of two SRM transitions and comparing the measurements of the two transitions with an allowable degree of acceptance to the possible difference between the two measurements and this is usually set arbitrarily (85). In the cases where only two transitions are acquired and there is a major discrepancy between the two transitions there is no way to know which of the two measurement is accurate i.e., which SRM transition suffers from the isobaric interference. Hence, ideally three or more SRM transitions are monitored where possible (85). This could be challenging in screening applications as there would be a high number of compounds being screened. Running multiple compound assays with several SRM transition per compound could be dependent on the slower scan speed and could as a result limit number of SRM transition. This would be limited by the chromatographic peak width and the need to sample the peak accurately with a sufficient number of scan points. In quantitative analysis however running at least two or more SRM transitions is key to improve the method selectivity (85).

Another approach to identify isobaric interferences that has been extensively studied by Kushnir *et al.* is the branching ratio approach, where the ratios are defined by the quantitative fragmentation patterns of the target analyte and

the internal standard. The ratios are then calculated as a ratio of the peak areas or heights of multiple SRM transitions of the target analyte and the internal standard (166). There would be predefined acceptance ranges for the branching ratios which determine the confidence in the measurement results. This approach is commonly used in GC-MS and in the legally regulated areas such as pesticide control and forensic toxicology, yet it is rarely used in hospital and clinical testing laboratories (85,166,168,169). One of the several reasons for this is the fact that fragmentation patterns differ due to the specific instrument characteristics (85,171–173). The fragmentation patterns also differ due to the diverse matrix effects (such as potential matrix related isobaric interferences with the precursor ion hence requiring constant revalidation of the branching ratios accepted ranges (170). Additionally, branching ratios are not applicable to the many compounds that produce only a single product ion with sufficient signal intensity for quantitative measurement under CID.

In higher-order RMPs, monitoring at least one additional SRM transition is a requirement to ensure method's specificity and selectivity and the SRM transitions are referred to as the quantification and confirmation transitions and results of both are compared for confirmation of measurements. The quantitative SRM is usually the one with higher signal intensity and/or better signal-to-noise ratio to improve precision of measurement and reduce instrument variability in repeated measurements. Furthermore, in RMP at method development stage, significant effort is invested in achieving optimum chromatographic separation prior to the MS, usually requiring longer running times to reduce the chance of co-eluting the target analyte with an isobaric

interference. However, in hospital and clinical testing laboratories long run times are not usually attainable. This is due to many reasons such as the increased cost of the analysis and potential delays in reporting results.

A final issue to mention when it comes to inaccuracies arising from ion selections is cross-talk, although it has become less of a problem in recent years with the new generations of MS but cross-talk could still occur in methods where there is a large number transitions acquired simultaneously. Cross-talk is the effect that takes place when two SRM transitions that are acquired in the same method have the same mass for the product ion e.g., *m/z* 353 → 165 and *m/z* 400 → 165. This occurs due to poor optimisation of the inter-scan delay which could result in incomplete emptying of the collision cell before acquiring the next SRM transition, which results in error/inaccuracy in the measurement because a higher signal of the product ion would be recorded that is coming from the previous SRM transition rather than the target analyte's SRM which would result in a bias (85).

1.10.6 Data quality

Due to limited resources and lack of traceable higher-order CRMs and methods, hospital laboratories are relying on the data acquired through EQAS to assess their results (See 1.2.1 and 1.2.2). However, the EQAS samples do not always have a reference value but rather a consensus mean value to which tolerance windows are assigned merely based on the data collected from the laboratories. Hence, if there is an inherent bias in the measurements, the hospital laboratories will not know. Similarly, since most data is reported without measurement uncertainty the comparison is not an entirely reliable

representation of the actual method performance nor does it provide an estimate of the methods accuracy and traceability (26).

At NMI laboratories, when developing higher-order reference measurement procedures and using these methods to produce certified reference materials, two cornerstones to build upon are accuracy and precision to achieve traceability and low measurement uncertainty, respectively. The accuracy is based on the use of higher-order certified solid standards that are assigned a reference purity value using a primary method e.g., qNMR and are traceable to the SI unit. Additionally, all sample preparation is performed gravimetrically to ensure an unbroken traceability chain of the SI unit. This is achieved by using high accuracy balances that are calibrated by ISO 17025 accredited calibration laboratories against check weights that are traceable to the SI unit. Gravimetric preparation is also more accurate than volumetric preparation as it does not suffer from the environmental effects on the volumes of liquid such as density. However gravimetric preparation is laborious, time consuming and requires skilled analysts. Such highly accurate work ensures traceability to the ‘true value’. While the term ‘true value’ is often used it is not accurate description because there is no such value as a measurement alone without its measurement uncertainty, the ‘true value’ is the true range of values at an assigned confidence interval (143).

To achieve the true range of values that is fit for purpose, the measurement uncertainty of the method needs to be as small as possible by improving the reference method’s performance. For a primary ratio reference method as double DEM-IDMS, the most important factor to drive down the measurement

uncertainty is by the improvement of the precision of the analyte to labelled internal standards ratios as this is typically the largest component of the uncertainty budget (143). This could be achieved in several ways including: (a) increasing number of replicate injections to five or more injections as the tighter the variability between injections the smaller the uncertainty would be; (b) ensuring the number of scans across the peaks is an accurate representation of the peak by optimising the dwell time to have a minimum ten scans across the peak (174–176); (c) consistent integration of the peaks and analysing samples in replicates. A thorough literature search highlighted that very few works investigate the impact of these factors on the measurement uncertainty.

While traceability is established by using higher-order certified reference solid (powder) materials and gravimetric preparation. Higher-order certified reference materials in biological matrix are then used for method validation. Higher-order reference measurement methods can provide crucial support to help hospital laboratories assess and improve their methods. This is by certifying solid CRMs, biological matrix CRMs, assigning reference traceable values for EQAS samples and supporting the laboratories that are running the EQAS.

Since a key aspect of data validity is the comparability of results across time and space (be it between laboratories in the same country or in different countries), the need for traceability and MU is evident (see section 1.2.1 and 1.9). Although it has not been reported by clinical laboratories in the past, the move towards it is taking place. It is crucial to have traceability and a MU

estimate as part of the method validation process. Validation is defined by the ISO as '*confirmation by examination and provision of objective evidence that the particular requirements of specified intended use are fulfilled'* (177). The goal of the validation process is to achieve reliable results, and this is by: (a) establishing what are all the possible elements that influence the measurement; (b) ensuring traceability to available higher-order references (methods and materials) and/or the SI unit; and (c) determining the sources of uncertainties associated with each of the influencing factors and with the references. Hence, validation is a tool to establish traceability. Therefore, the distinction between accuracy and traceability; a method could be accurate in terms of assessing the result against what is considered the 'true value' but is not necessarily traceable. However, whenever a method is traceable, it is accurate (178–180).

1.11 The need for standardisation of plasma metanephrines analysis

As mentioned earlier (see section 1.3.1) plasma metanephrines was the selected case study as a small polar compound measured for clinical diagnostic test for this research. This is to investigate the sources of variability in higher-order reference measurement procedure development of small endogenous polar molecules in biological samples. Here, the need for the standardisation of this analysis is explained. Plasma metanephrines analysis is a test that is used for the diagnosis PPGLs (181). In addition to the analysis of metanephrine and normetanephrine, the metabolite of dopamine, 3-methoxytyramine (3MT) is sometimes added to the test however it is not part

of the universal diagnostic clinical test. Recently, the clinical value of measuring 3MT has been investigated (181). Currently, there are no higher-order RMPs to measure METs and hence no higher-order CRMs. Moreover, there is only one available EQAS for plasma METs worldwide. This EQAS started as a pilot scheme in 2007 by the Australian Association of Clinical Biochemists. The pilot project then became an EQAS provided by through the Royal College of Pathologists of Australia (RCPA) (182). The RCPA EQAS monthly ships out of Australia lyophilised plasma samples that are both patient plasma and plasma spiked with metanephrines standards. The EQAS team at RCPA summarises the measurements provided by the subscribing laboratories in a report to enable the comparison of measurements (see Figure 1.15). In the absence of higher-order RMP to assign reference values to the samples, a consensus mean value is used to evaluate the data (47,183). Figure 1.15 shows an example of reported EQAS data of the same sample. The use of a consensus mean value instead of a reference value can provide an error in the interpretation of the comparative data. This is consensus mean value is assigned based on measurements that are not traceable reference values but rather a mean of measurements made using different methods. As the example data in Figure 1.15 show that for the same sample, metanephrine is analysed at 875 pmol/L by two laboratories whereas another three other laboratories measured it at 1668 pmol/L. This is nearly double the amount of metanephrines measured for the same sample. It is a similar case for normetanephrine in the same example data. With the absence of a reference measurement that is assigned using a traceable higher-order reference measurement procedure none of these measurements can be accurately

assessed. It could be that the laboratories measuring at the lower concentration are the ones providing an accurate measurement or it could be the other way around. Hence, it is crucial to provide a reference value whenever comparing measurements performed in different time and space. This is the basis of any standardisation approach (see section 1.1).

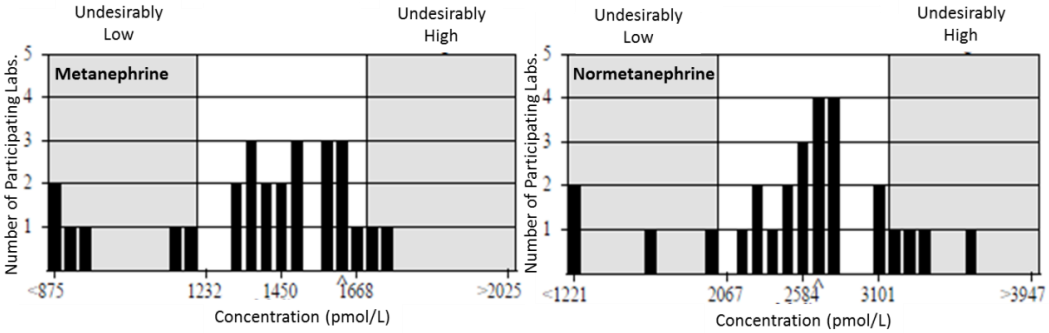


Figure 1.15 Example of RCPA-EQA plasma metanephrines analysis data. Data shown is of one sample analysed by 27 participating laboratories, 23 of which are using LC-MS. Charts in the figure used with permission from The Consultant Clinical Scientist of the Liverpool University Hospital, Mr Andrew Davison, from their laboratory's participation in the EQAS report.

Therefore, it is critical to assign reference values to the EQAS samples to enable hospital laboratories to evaluate their analytical methods to a reference traceable value rather than a consensus mean value which is currently used. The pheochromocytoma and paraganglioma research support organization (PRESSOR) *"is a non-profit consortium of health science professionals at scientific, medical and academic institutions around the world dedicated to research into improved diagnosis, localization, management and treatment of pheochromocytoma and paragangliomas, particularly malignant pheochromocytoma and paragangliomas".* PRESSOR has identified the need for reference value assignment through a higher-order RMP. They have also

identified the need for standardisation and harmonisation as a critical step to improving diagnosis and patient care (184).

1.12 The need for standardisation of tacrolimus analysis

As mentioned earlier (see 1.3.2) tacrolimus in whole blood analysis was the selected case study, as a larger exogenous compound measured for clinical testing, for this research. This is to investigate the sources of variability in higher-order reference measurement procedure development of exogenous non-polar small molecules in biological samples. Here, the need for the standardisation of this analysis is explained.

Adverse toxicity of calcineurin inhibitor based immunosuppressant drugs, namely tacrolimus, are evident. Tacrolimus minimisation strategies were suggested in 2009 in the European Consensus Conference on tacrolimus and in the findings of the ELiTE-Symphony study (Efficacy Limiting Toxicity Elimination) (185–188). Ekberg *et al.* reported in this large-scale study that 3-5 % of allografts in renal transplant patients are lost annually, resulting in death with a functioning allograft or an immunosuppressant induced nephropathy of the allograft in the long term. Hence, they performed a study on 1645 renal transplant patients and introduced combination therapies to reduce the dose needed from the immunosuppressant drugs. This has affirmed the need for TDM of immunosuppressant drugs. Clinical testing laboratories worldwide perform tacrolimus monitoring to ensure patients are on the minimum effective correct dose to protect them from adverse toxicities while ensuring efficacy and no allograft rejection episodes. In 2015/2016 tacrolimus was one of the top twenty highest cost prescribed/issued

medications in the community and hospitals in England at over 71 million GBP for renal transplant patients (189). It had been on the top 20 costing medicines list since 2013 only to move lower in the list but remaining among highly costing medicines in 2017 (189–192). The patients prescribed tacrolimus undergo TDM routinely for the rest of their lives.

Currently, tacrolimus is analysed using immunoassay-based methods and LC-MS methods. Immunoassay kits including all necessary calibrants and QC materials are usually commercially available by several suppliers. Immunoassay kits have been reported to cross-react with the metabolites of tacrolimus and some suffer from interference with heterologous antibodies in some samples (193–197). LC-MS based methods on the other hand are mostly in-house developed methods (198). The need for standardisation of tacrolimus in whole blood analysis has been investigated in several studies (151,198).

Levine *et al.* reported their assessment of the need for standardisation in a comparative study across 22 laboratories in 14 countries. Upon comparing the measurements, high variability of the results in comparison to the reference values of the samples was observed. The reference values were assigned using a candidate higher-order reference method. Figure 1.16 demonstrate the spread of results across laboratories and across analytical techniques. The study showed the variability in data among laboratories that used the Abbott ARCHITECT kit was the least, followed by laboratories using LC-MS methods while laboratories using Dade Dimensions showed the highest variability. Although harmonisation of the results of the LC-MS based methods

was attempted by using the same calibration across the different laboratories, this failed to bring the data together. The fact that calibration materials are different among methods is a major concern that highlight the need for traceable materials that would facilitate traceability of methods and pave the way towards standardisation (198). An example of the data variability and its clinical importance is the results of sample P-4 (Figure 1.16-B) where if the clinical decision point is 8 µg/L, patients who are tested by the lowest biased laboratories would require an increase in the dose and subsequently a rise in adverse reactions. On the other hand, patients tested at the highest biased laboratories would have their dose reduced potentially resulting in allograft rejection (198). Hence, harmonisation and standardisation of tacrolimus analysis is not only a step towards providing consistent patient care but is also an important step to improve patient care by ensuring better quality of test.

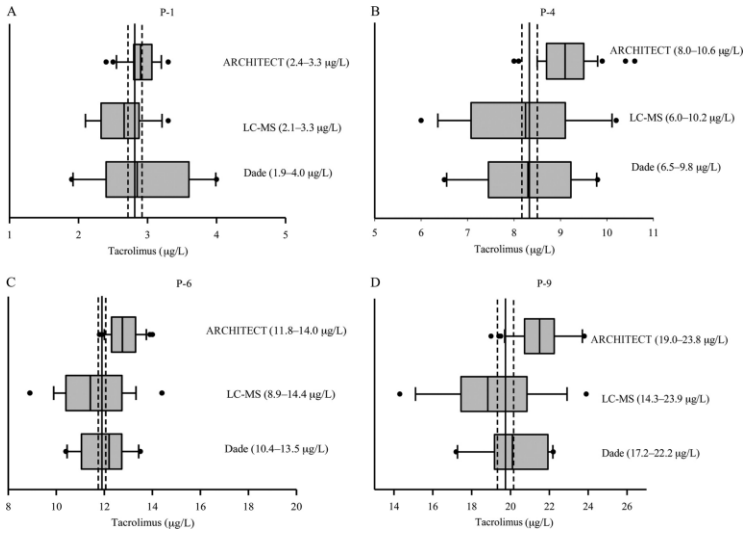


Figure 1.16 Data of 4 samples analysed by different laboratories using three analytical techniques. The results are shown in Box-and-whiskers plots with the range of attained concentrations in brackets. Reference values and their expanded measurement uncertainties of each of the samples, indicated by line and dashed lines respectively, were determined using DEM-IDMS method. Figure used with permission from Clinical Chemistry (198).

Data from the Tacrolimus International Proficiency Testing Scheme (TIPTS), which was founded by Prof. Holt and Prof. Johnston and recently acquired by LGC Proficiency Testing (UK) (199), show that majority of the laboratories are now using LC-MS based methods. These methods are in-house developed methods mostly which lack standardisation of measurement procedures such as the internal standards, calibration, sample preparation steps including solid phase extraction (SPE) and chromatographic methods, etc. Such discrepancy in procedures leads to high inter-laboratory variability, imprecision and inaccuracy; the imprecision is increased by the lack of standardisation of the assignment of values to calibration materials (151).

Annesely *et al.* attempted standardisation of LC-MS based methods in 2013, in collaboration with Waters, Analytical Services International at St. George's Hospital (ASI) and the National Measurement Laboratory (NML) hosted at LGC. This study provided seven different laboratories with the commercial analytical kit MassTrack™ immunosuppressants kit by Waters™ (Massachusetts, USA). The selection of the laboratories was based on the use of the same instrumental platform, the Acquity® TQD UPLC-MS/MS. The NML's role was to assign reference values to four patient blood pool samples. The data obtained using the kits by the different laboratories showed excellent agreement. In this study the higher-order CRM of tacrolimus spiked in human blood (ERM-DA110a) was used to enable measurement uncertainty estimation for the MassTrack™ kit which demonstrated excellent agreement with different laboratories' LC-MS methods (151). This study is a good example of how higher-order candidate reference measurement procedure

and values used to compare measurements made by different methods at different times and locations. It also demonstrated the use of a higher-order CRM for assignment of measurement uncertainty of analytical methods. However, work is still required to improve the standardisation of this test overall.

Overall LC-MS brought forward great advantages to the TDM of tacrolimus over immunoassay as it increased the analytical specificity (151,200,201). Additionally, haematocrit and serum albumin are reported to influence the analysis of tacrolimus in immunoassay kits (204–206). Taylor *et al.* concluded that LC-MS is a better routine test for tacrolimus TDM due to the above reasons in combination with the technological advances in LC-MS and the modern fast sample preparation method above (151,207). This research builds on the efforts made so far in the standardisation of tacrolimus in whole blood analysis by using the advances in LC-MS to better understand the sources of variability in higher-order reference measurements of non-polar molecules like tacrolimus and developing a CRM in patient blood to improve the evaluation of assessment of immunoassay kits. Additionally, in this research, building on the existing efforts for the standardisation of tacrolimus in whole blood an assessment of a new micro-sampling device to support hospital laboratories was investigated.

This standardisation challenge is different from that of plasma metanephrines analysis for several reasons including: a) the first step in the process of developing a higher-order RMP for the standardisation in plasma METs is still missing; b) the knowledge gained through working with tacrolimus is different.

While METs and Tacrolimus are considered small molecules, they provide an opportunity to investigate polar opposites of the range of small molecules in many aspects, this is further discussed in 1.13.1.

1.13 Thesis outline

1.13.1 Scope of research

In this thesis the concept of standardisation and harmonisation by using DEM-IDMS based reference methods is explored to investigate and understand the sources of variability in higher-order reference measurements in clinical chemistry and laboratory medicine applications. This was performed by both developing a candidate higher-order RMP and the application of one to analyse samples and assign reference values. This was investigated for tacrolimus in whole blood and metanephrines in plasma. These were selected as case studies of small molecule analysis in biological samples for clinical testing (see sections 1.3.1 and 1.3.2).

As mentioned earlier, although these compounds classified as small molecules, they could not be any more disparate in chemical and biological properties. Firstly, tacrolimus is a drug, an exogenous compound, that is administered to patients whereas METs are endogenous metabolites that are produced within the human body. This from an analytical point of view when performing DEM-IDMS where calibration blends are made of blank matrix, analyte standard and SILIS, provide a different challenge. In the case of tacrolimus, a blank biological matrix that is free of tacrolimus could be readily available by sampling humans who were not administered tacrolimus.

However, for METs, this becomes more challenging as such blank biological matrix free of METs does not exist naturally.

Additionally, these two applications not only cover a wide range of aspects of clinical chemistry and laboratory medicine applications as a diagnostic test and a monitoring test, but they are also measured in two different biological matrices. This creates the opportunity for this research to investigate any potential sources of variability in the different matrices; whole blood and plasma.

Chemically, the compounds selected for this work have very different physiochemical properties, summarised in Table 1.2 while the structures are in Figure 2.3 and Figure 4.2. While tacrolimus is a relatively larger molecule with an average molar mass of 804.02 Da, metanephrines are significantly smaller ranging between 167-197 Da. Tacrolimus is a non-polar highly lipophilic compound which makes it suitable for the use of RP-LC whereas METs are highly polar compounds that are not readily suited for the commonly used technique RP-LC. Moreover, the levels at which the compounds are measured is different. While tacrolimus is measured at mass fractions of ng/g, METs are measured at ultra-low mass fractions range in the lower pg/g. This created an opportunity to address different technical challenges in higher-order reference measurements at different levels. Given, the relatively higher mass fraction at which tacrolimus is measured and its clinical application where patients undergo the TDM regularly, it gave the opportunity to explore how higher-order reference measurement could support in investigating novel micro-sampling tool for the test. The use of higher-order RMP to assess new

sampling technologies was investigated for tacrolimus whereas such a possibility is not available for METs due to their ultra-low levels.

Table 1.2 Aspects of tacrolimus and METs. (Metanephrines Log P, pKa were obtained the Human Metabolome Database (208) and tacrolimus LogP, pKa were predicted by ChemAxon and obtained from The Toxin and Toxin Database (209)).

COMPOUND(S)	METANEPHRINES	TACROLIMUS
MASS $[M+H]^+$	*m/z* 167-197	*m/z* 804
NATURE	Endogenous	Exogenous
MASS FRACTION RANGE	1-100 pg/g	1-30 ng/g
MATRIX	Plasma	Whole blood
LogP	Normet -0.71 Met -0.27 3MT 0.41	3.19
pK_a *(basic)*	Normet 9.06 Met 9.25 3MT 9.64	-2.9
pK_a *(acidic)*	Normet 9.99 Met 10.05 3MT 10.39	9.96

By investigating these two applications, a comprehensive understanding of the process of standardisation by development and application of higher-order RMPs is gained. This research aims to investigate and better understand the sources of variability in the process of higher-order reference measurement. Another aim of the work was to develop novel RMP for the measurement of METs and new applications for tacrolimus higher-order reference measurements. This will ultimately lead to supporting hospital laboratories improving their measurement and providing better patient care.

1.13.2 Aims and objectives

1.13.2.1 Aims

The research aimed to investigate and better understand the sources of variability in the process of higher-order reference measurement of small molecules in biological samples for clinical diagnosis and monitoring. This would advance the standardisation and harmonisation of the two clinical measurement applications that were selected as case studies for this research. This was achieved by building the cornerstone for the standardisation and harmonisation of plasma metanephrines analysis and by providing tools to advance the standardisation of tacrolimus in whole blood.

This work aimed to develop a candidate higher-order reference measurement procedure for metanephrines in plasma. Additionally, this aims to characterise the candidate higher-order CRM of tacrolimus in pooled patient blood which is the first material of its kind. Because it was the first material that was assigned higher-order reference value which was produced of pooled patient blood. Tacrolimus in the CRM was incurred in the blood through patients taking the medication rather than added to it as a standard to healthy individuals blood pool. This meant the material had tacrolimus metabolites which would aid in better evaluation of immunoassay methods in hospital laboratories. Finally, in the work, it was the first time to use standardisation tools of higher-order RMPs and CRMs to evaluate a novel micro-sampling device (Chapter 5).

1.13.2.2 Objectives

The above aims were to be achieved by delivering the following objectives:

- Develop higher-order RMP for small polar molecule in biological samples by; developing a candidate higher-order RMP for plasma metanephrines analysis using LC with DEM-IDMS as a preferred technique.
- Evaluate the performance of the candidate RMP against a recognised ISO standard and international guidelines.
- Advance standardisation of larger small molecules measurements in whole blood samples by characterising a candidate higher-order CRM of tacrolimus in patient whole blood.
- Investigating the use of higher-order measurement tools (RMPs & CRMs) for the evaluation of novel alternative sampling devices.

1.13.3 Novelty

This work is first to investigate the sources of variability in higher-order reference measurement of small molecules in biological samples for clinical applications. It is the first to suggest detailed technical solutions for mitigating sources of variability in higher-order reference measurement of small molecules. As mentioned earlier, this research presents the first candidate higher-order RMP for plasma metanephrines analysis and subsequently the first to produce traceable reference measurements of samples. Additionally, the incurred higher-order candidate reference material for tacrolimus in whole blood characterised in this research is the first of its kind. Moreover, at the time of the work, there was no evaluation for the use of the commercially

available micro-sampling device (Mitra®) for the analysis of tacrolimus which was achieved using reference methods and reference values and reference. These micro-sampling devices will support hospital laboratories move toward a more robust and accurate micro-sampling technique instead of dried blood spots that suffer from high variability due to haematocrit effect. This work outlines the challenges of higher-order reference measurements, method development and application, and will benefit those working in clinical analysis of small molecules up to 804 Da in molar mass.

1.13.4 Value

This research will impact the quality of data achieved for those working on clinical measurements of small and large molecules, such as metanephrines and tacrolimus, for which standardisation of measurement is currently lacking. Every standardisation initiative for any laboratory medicine test starts with developing a higher-order reference measurement procedure that enables assigning traceable reference values to enable data comparison among laboratories. The need for standardisation of plasma metanephrines has been identified by the pheochromocytoma and paraganglioma research support organisation (PRESSOR) as a crucial step to improve diagnosis of patients with these tumours (184). As for tacrolimus in whole blood measurements, there have been efforts for standardisation that did not only pave the way for the work in this research to take place but more importantly showed the need for an incurred higher-order certified reference material to improve the evaluation of hospital laboratories methods especially immunoassay-based methods. Finally, micro-sampling is an ongoing area of investigation for any

suitable laboratory medicine application to ultimately provide better quality patient experience. Given the frequency, an organ-transplant immunosuppressed patient needs to visit the hospital to undergo tacrolimus blood analysis, micro-sampling is evidently an area that is worth investigating to improve the patient's quality of life.

1.14 References

1. Singer CJ, Holmyard EricJ, Hall AR, Williams TA. A history of technology. Reprint 1979. Vol. 2. United Kingdom: Clarendon Press; 1954. 804 p.
2. Clagett M. Ancient Egyptian Science: Ancient Egyptian mathematics. American Philosophical Society; 1989. 465 p.
3. BIPM - About the BIPM [Internet]. [cited 2018 Aug 13]. Available from: https://www.bipm.org/en/about-us/
4. McDonald DM, Scarre C. The origins of metrology: collected papers of Dr. Daniel McLean McDonald. McDonald Institute for Archaeological Research; 1992. 156 p.
5. Giacomo P. The new definition of the metre. J Opt. 1984;15(4):167.
6. Quinn TJ. Practical realization of the definition of the metre (1997). Metrologia. 1999;36(3):211.
7. SI-Brochure-9-EN.pdf [Internet]. [cited 2019 Dec 12]. Available from: https://www.bipm.org/utils/common/pdf/si-brochure/SI-Brochure-9-EN.pdf
8. Barwick V, Prichard E. International vocabulary of metrology -- Basic and general concepts and associated terms (VIM) 3rd ed [Internet]. 2012 [cited 2017 Nov 14]. Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf
9. Milton MJT, Quinn TJ. Primary methods for the measurement of amount of substance. Metrologia. 2001;38(4):289.
10. Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardization and harmonization of clinical laboratory tests123. Am J Clin Nutr. 2016 Sep;104(Suppl 3):907S-912S.
11. Kaarls R, Quinn TJ. The Comité Consultatif pour la Quantité de Matière: a brief review of its origin and present activities. Metrologia. 1997;34(1):1.
12. BIPM. Proceedings of the 1st meeting of the CCQM, Bureau International des Poids et Mesures (BIPM). 1995. pages 17-20
13. BIPM. Proceedings of the 4th meeting of the CCQM, Bureau International des Poids et Mesures (BIPM). 1998. pages 24-33
14. Thienpont LM, Van Uytfanghe K, De Leenheer AP. Reference measurement systems in clinical chemistry. Clin Chim Acta. 2002 Sep 1;323(1):73–87.
15. Siekmann L. Establishing measurement traceability in clinical chemistry. In: Traceability in Chemical Measurement [Internet]. Springer, Berlin, Heidelberg;

2003 [cited 2017 Nov 19]. p. 147–59. Available from: https://link.springer.com/chapter/10.1007/3-540-27093-0_25

16. Armbruster D, Miller RR. The Joint Committee for Traceability in Laboratory Medicine (JCTLM): A Global Approach to Promote the Standardisation of Clinical Laboratory Test Results. Clin Biochem Rev. 2007 Aug;28(3):105–14.
17. Vesper HW, Thienpont LM. Traceability in Laboratory Medicine. Clin Chem. 2009 Jun 1;55(6):1067–75.
18. White GH. Metrological traceability in clinical biochemistry. Ann Clin Biochem. 2011 Sep 1;48(5):393–409.
19. Miller WG, Myers GL, Rej R. Why Commutability Matters. Clin Chem. 2006 Apr 1;52(4):553–4.
20. Cali JP. Standardization of diagnostic materials. 2. Problems of standardization in clinical chemistry. Bull World Health Organ. 1973 Jun;48(6):721–6.
21. Cali JP. Problems of standardization in clinical chemistry. Bull World Health Organ. 1973;48(6):721–6.
22. Büttner J. Reference methods as a basis for accurate measuring systems. Eur J Clin Chem Clin Biochem J Forum Eur Clin Chem Soc. 1991 Apr;29(4):223–35.
23. Tietz NW. A model for a comprehensive measurement system in clinical chemistry. Clin Chem. 1979 Jun 1;25(6):833–9.
24. ISO 15189:2012 - Medical laboratories -- Requirements for quality and competence [Internet]. [cited 2018 Aug 13]. Available from: https://www.iso.org/standard/56115.html
25. Lehmann HP, Fuentes-Arderiu X, Bertello LF. Glossary of terms in quantities and units in Clinical Chemistry (IUPAC-IFCC Recommendations 1996). Pure Appl Chem. 1996;68(4):957–1000.
26. Greaves RF. A Guide to Harmonisation and Standardisation of Measurands Determined by Liquid Chromatography – Tandem Mass Spectrometry in Routine Clinical Biochemistry. Clin Biochem Rev. 2012 Nov;33(4):123–32.
27. Cohen A, Hertz HS, Mandel J, Paule RC, Schaffer R, Sniegoski LT, et al. Total serum cholesterol by isotope dilution/mass spectrometry: a candidate definitive method. Clin Chem. 1980 Jun 1;26(7):854–60.
28. EN ISO17511. In vitro diagnostic medical devices - Measurement of quantities in biological samples - Metrological traceability of values assigned to calibrators and control materials [Internet]. 2003. Available from: https://www.iso.org/standard/30716.html
29. ISO 15194:2009 - In vitro diagnostic medical devices -- Measurement of quantities in samples of biological origin -- Requirements for certified reference materials and the content of supporting documentation [Internet]. [cited 2018 May 31]. Available from: https://www.iso.org/standard/42022.html
30. Characterization and Qualification of Commutable Reference Materials for Laboratory Medicine; Approved Guideline. [Internet]. Clinical and Laboratory Standards Institute.; 2010 [cited 2018 May 31]. Available from: https://clsi.org/media/1419/ep30a_sample.pdf
31. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem. 1998 Aug 1;31(6):449–57.
32. Baadenhuijsen H, Demacker PN, Hessels M, Boerma GJ, Penders TJ, Weykamp C, et al. Testing the accuracy of total cholesterol assays in an external quality-control program. Effect of adding sucrose to lyophilized control

sera compared with use of fresh or frozen sera. Clin Chem. 1995 May;41(5):724–30.

33. Malavasi B, Catapano A, Galli G, Franzini C. A collaborative trial for the evaluation of blood cholesterol measurement in clinical laboratories in Italy. Eur J Clin Chem Clin Biochem J Forum Eur Clin Chem Soc. 1992 Mar;30(3):157–61.
34. Hill SA, Heathcote JC, McQueen MJ. The significance of matrix effects on the measurement of lactate dehydrogenase (LD) activity using kodak dry slide technology in the ontario laboratory proficiency testing program. Clin Biochem. 1990 Apr;23(2):179–81.
35. Database of higher-order reference materials, measurement methods/procedures and services [Internet]. [cited 2018 Feb 6]. Available from: https://www.bipm.org/jctlm/
36. Grebe SK, Singh RJ. LC-MS/MS in the Clinical Laboratory – Where to From Here? Clin Biochem Rev. 2011 Feb;32(1):5–31.
37. Vogeser M, Seger C. LC-MS/MS in clinical chemistry. J Chromatogr B Analyt Technol Biomed Life Sci. 2012 Feb 1;883–884:1–2.
38. Blackmore S, Pfeiffer CM, Lee A, Fazili Z, Hamilton MS. Isotope Dilution–LC-MS/MS Reference Method Assessment of Serum Folate Assay Accuracy and Proficiency Testing Consensus Mean. Clin Chem. 2011 Jul 1;57(7):986–94.
39. Hawley JM, Owen LJ, MacKenzie F, Mussell C, Cowen S, Keevil BG. Candidate Reference Measurement Procedure for the Quantification of Total Serum Cortisol with LC-MS/MS. Clin Chem. 2016 Jan 1;62(1):262–9.
40. Wang Y, Hu X, Tang J, Wang W, Dong L, Gu C, et al. Traceability to a primary reference measurement procedure (ID-LCMS); A key step in validating the clinical accuracy and safety of hospital blood glucose monitoring systems. Clin Chim Acta. 2018 Nov 1;486:275–81.
41. Tai SS-C, Bedner M, Phinney KW. Development of a Candidate Reference Measurement Procedure for the Determination of 25-Hydroxyvitamin D3 and 25-Hydroxyvitamin D2 in Human Serum Using Isotope-Dilution Liquid Chromatography−Tandem Mass Spectrometry. Anal Chem. 2010 Mar 1;82(5):1942–8.
42. Stepman HCM, Vanderroost A, Uytfanghe KV, Thienpont LM. Candidate Reference Measurement Procedures for Serum 25-Hydroxyvitamin D3 and 25-Hydroxyvitamin D2 by Using Isotope-Dilution Liquid Chromatography–Tandem Mass Spectrometry. Clin Chem. 2011 Mar 1;57(3):441–8.
43. Bylda C, Thiele R, Kobold U, Volmer DA. Simultaneous quantification of acetaminophen and structurally related compounds in human serum and plasma. Drug Test Anal. 2014;6(5):451–60.
44. Group BMJP. Evaluation of diagnostic procedures. BMJ. 2002 Jun 8;324(7350):1391.
45. Knottnerus JA, Weel C van, Muris JWM. Evaluation of diagnostic procedures. BMJ. 2002 Feb 23;324(7335):477–80.
46. Bruns DE, Huth EJ, Magid E, Young DS. Toward a Checklist for Reporting of Studies of Diagnostic Accuracy of Medical Tests. Clin Chem. 2000 Jul 1;46(7):893–5.
47. Lenders JM, Pacak K, Walther MM, et al. Biochemical diagnosis of pheochromocytoma: Which test is best? JAMA. 2002 Mar 20;287(11):1427–34.

48. Johnston A, Holt DW. Therapeutic drug monitoring of immunosuppressant drugs. Br J Clin Pharmacol. 1999 Apr;47(4):339–50.
49. Johnston A, Marsden JT, Holt DW. The continuing need for quality assessment of cyclosporine measurement. Clin Chem. 1989 Jan 7;35(7):1309–12.
50. Benet LZ, Goyan JE. Bioequivalence and Narrow Therapeutic Index Drugs. Pharmacother J Hum Pharmacol Drug Ther. 1995;15(4):433–40.
51. Levy G. What are narrow therapeutic index drugs? Clin Pharmacol Ther. 1998;63(5):501–5.
52. Directive 2002/98/ec of the European parliament and of the council of 27 January 2003. L33 pp 30-40 [Internet]. 2003 [cited 2017 Nov 14]. Available from: https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-1/dir_2002_98/dir_2002_98_en.pdf
53. EU Science Hub - European Commission [Internet]. EU Science Hub. [cited 2018 Jun 28]. Available from: https://ec.europa.eu/jrc/en/interlaboratory-comparisons/IMEP
54. Van Nevel L, Ornemark U, Smeyers P, Harper C, Taylor PDP. IMEP-17 trace and minor constituents in human serum. Rep Particip Part. 2007;1.
55. Stokes P, O'Connor G. Development of a liquid chromatography–mass spectrometry method for the high-accuracy determination of creatinine in serum. J Chromatogr B. 2003 Aug 25;794(1):125–36.
56. Briche CSJW, Carter D, Webb KS. Comparison of gas chromatography and liquid chromatography mass spectrometric measurements for high accuracy analysis of cholesterol in human serum by isotope dilution mass spectrometry. Rapid Commun Mass Spectrom RCM. 2002;16(9):848–53.
57. Semerjian HG, Watters RL. Impact of measurement and standards infrastucture on the national economy and international trade. Measurement. 2000 Apr 1;27(3):179–96.
58. Miller WG, Eckfeldt JH, Passarelli J, Rosner W, Young IS. Harmonization of Test Results: What Are the Challenges; How Can We Make It Better? Clin Chem. 2014 Jan 7;60(7):923–7.
59. Jannetto PJ, Fitzgerald RL. Effective Use of Mass Spectrometry in the Clinical Laboratory. Clin Chem. 2016 Jan 1;62(1):92–8.
60. Mandatory Guidelines for Federal Workplace Drug Testing Programs [Internet]. Federal Register. 2017 [cited 2018 Mar 5]. Available from: https://www.federalregister.gov/documents/2017/01/23/2017-00979/mandatory-guidelines-for-federal-workplace-drug-testing-programs
61. Fitzgerald RL, Herold DA. Serum total testosterone: immunoassay compared with negative chemical ionization gas chromatography-mass spectrometry. Clin Chem. 1996 May;42(5):749–55.
62. Taieb J, Mathian B, Millot F, Patricot M-C, Mathieu E, Queyrel N, et al. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. Clin Chem. 2003 Aug;49(8):1381–95.
63. Vogeser M, Seger C. Mass spectrometry methods in clinical diagnostics – state of the art and perspectives. TrAC Trends Anal Chem. 2016 Nov;84, Part B:1–4.
64. Saint-Marcoux F, Sauvage F-L, Marquet P. Current role of LC-MS in therapeutic drug monitoring. Anal Bioanal Chem. 2007 Aug 1;388(7):1327–49.
65. Anari MR, Bakhtiar R, Zhu B, Huskey S, Franklin RB, Evans DC. Derivatization of ethinylestradiol with dansyl chloride to enhance electrospray ionization:

application in trace analysis of ethinylestradiol in rhesus monkey plasma. Anal Chem. 2002 Aug 1;74(16):4136–44.

66. Santen RJ, Demers L, Ohorodnik S, Settlage J, Langecker P, Blanchett D, et al. Superiority of gas chromatography/tandem mass spectrometry assay (GC/MS/MS) for estradiol for monitoring of aromatase inhibitor therapy. Steroids. 2007 Jul 1;72(8):666–71.
67. Santen RJ, Lee JS, Wang S, Demers LM, Mauras N, Wang H, et al. Potential role of ultra-sensitive estradiol assays in estimating the risk of breast cancer and fractures. Steroids. 2008 Dec 12;73(13):1318–21.
68. Kushnir MM, Blamires T, Rockwood AL, Roberts WL, Yue B, Erdogan E, et al. Liquid chromatography tandem mass spectrometry assay for androstenedione, dehydroepiandrosterone, and testosterone with pediatric and adult reference intervals. Clin Chem. 2010 Jul 1;56(7):1138–47.
69. Vesper HW, Botelho JC. Standardization of testosterone measurements in humans. J Steroid Biochem Mol Biol. 2010 Aug 1;121(3):513–9.
70. Ketha H, Kaur S, Grebe SK, Singh RJ. Clinical applications of LC-MS sex steroid assays: evolution of methodologies in the 21st century. Curr Opin Endocrinol Diabetes Obes. 2014 Jun;21(3):217–26.
71. Hirota N, Furuta T, Kasuya Y. Determination of cortisol in human plasma by capillary gas chromatography—mass spectrometry using [2H5] cortisol as an internal standard. J Chromatogr B Biomed Sci App. 1988 Jan 1;425:237–43.
72. Kasuya Y, Furuta T, Hirota N. Determination of cortisol in human plasma by stable isotope dilution mass spectrometry. Biomed Environ Mass Spectrom. 16(1–12):309–11.
73. Murphy BEP. Commercial radioimmunoassays do not measure urinary free cortisol accurately and should not be used for physiological studies. J Clin Endocrinol Metab [Internet]. 2003 Aug 1 [cited 2018 Jun 13];88(8). Available from: https://insights.ovid.com/clinical-endocrinology-metabolism/jceme/2003/08/000/commercial-radioimmunoassays-not-measure-urinary/83/00004678
74. Urine free cortisol analysis by automated immunoassay and high-performance liquid chromatography for the investigation of Cushing's syndrome. Ann Clin Biochem. 2006 Sep 1;43(5):402–7.
75. Wood L, Ducroq DH, Fraser HL, Gillingwater S, Evans C, Pickett AJ, et al. Measurement of urinary free cortisol by tandem mass spectrometry and comparison with results obtained by gas chromatography-mass spectrometry and two commercial immunoassays. Ann Clin Biochem. 2008 Jul 1;45(4):380–8.
76. Taylor PJ, Cooper DP, Gordon RD, Stowasser M. Measurement of aldosterone in human plasma by semiautomated HPLC tandem mass spectrometry. Clin Chem. 2009 Jun 1;55(6):1155–62.
77. Taylor PJ, van Rosendal SP, Coombes JS, Gordon RD, Stowasser M. Simultaneous measurement of aldosterone and cortisol by high-performance liquid chromatography–tandem mass spectrometry: Application to dehydration–rehydration studies. J Chromatogr B. 2010 May 1;878(15):1195–8.
78. Jones G. Assay of vitamins D2 and D3, and 25-hydroxyvitamins D2 and D3 in human plasma by high-performance liquid chromatography. Clin Chem. 1978 Feb 1;24(2):287–98.

79. Binkley N, Krueger D, Cowgill CS, Plum L, Lake E, Hansen KE, et al. Assay variation confounds the diagnosis of hypovitaminosis D: a call for standardization. J Clin Endocrinol Metab. 2004 Jul 1;89(7):3152–7.
80. Singh RJ. Are clinical laboratories prepared for accurate testing of 25-hydroxy vitamin D Clin Chem. 2008 Jan 1;54(1):221–3.
81. Phinney KW. Development of a standard reference material for vitamin D in serum. Am J Clin Nutr. 2008 Aug 1;88(2):511S-512S.
82. Singh RJ, Taylor RL, Reddy GS, Grebe SKG. C-3 Epimers can account for a significant proportion of total circulating 25-hydroxyvitamin D in infants, complicating accurate measurement and interpretation of vitamin D status. J Clin Endocrinol Metab. 2006 Aug 1;91(8):3055–61.
83. Adaway JE, Peitzsch M, Keevil BG. A novel method for the measurement of plasma metanephrines using online solid phase extraction-liquid chromatography tandem mass spectrometry. Ann Clin Biochem Int J Biochem Lab Med. 2014 Jul 21;0004563214546560.
84. Wright MJ, Thomas RL, Stanford PE, Horvath AR. Multiple reaction monitoring with multistage fragmentation (MRM3) detection enhances selectivity for LC-MS/MS analysis of plasma free metanephrines. Clin Chem. 2015 Jan 3;61(3):505–13.
85. Vogeser M, Seger C. Pitfalls associated with the use of liquid chromatography-tandem mass spectrometry in the clinical laboratory. Clin Chem. 2010 Aug;56(8):1234–44.
86. Kirk-Othmer (Editor). Kirk-Othmer Encyclopedia of Chemical Technology. 5th ed. Vol. 2. John Wiley and Sons Inc.; 2007 - pages 408-520
87. Martin AJP, Synge RLM. A new form of chromatogram employing two liquid phases. Biochem J. 1941 Dec;35(12):1358–68.
88. Polykarpou E. Optimisation of chromatography for downstream protein processing [Internet]. [London, UK]: University College London; 2011 [cited 2018 Jun 4]. Available from: http://discovery.ucl.ac.uk/1334599/1/1334599.pdf
89. McNair HM, Miller JM. Basic Gas Chromatography. John Wiley and Sons Inc.; 1998 - pages 15-35 and 211-224
90. McKetta JJ. Encyclopedia of Chemical Processing and Design. Vol. 8. Marcel Dekker Inc.; 1976 - pages 245-260 and pages 315-330.
91. Anderson J, Berthod A, Pino V, Stalcup AM. Analytical Separation Science, 5 Volume Set. John Wiley & Sons; 2016. 2148 p.
92. Coskun O. Separation techniques: Chromatography. North Clin Istanb. 2016 Nov 11;3(2):156–60.
93. Snyder LR, Kirkland JJ, Dolan JW. Introduction to modern liquid chromatography. 3rd ed. John Wiley and Sons Inc.; 1974 - pages 373-417
94. Dorsey JG, Foley JP, Cooper WT, Barford RA, Barth HG. Liquid chromatography: theory and methodology. Anal Chem. 1992 Jun 15;64(12):353–89.
95. Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. Anal Bioanal Chem. 2012 Jan 1;402(1):231–47.
96. Wolcott RG, Dolan JW, Snyder LR, Bakalyar SR, Arnold MA, Nichols JA. Control of column temperature in reversed-phase liquid chromatography. J Chromatogr A. 2000 Feb 11;869(1):211–30.

97. Gritti F, Félix G, Achard MF, Hardouin F. Laterally attached liquid crystalline polymers as stationary phases in reversed-phase high-performance liquid chromatography: V. Study of retention mechanism using linear solvation energy relationships. J Chromatogr A. 2001 Jul 13;922(1):51–61.
98. Jaroniec M. Partition and displacement models in reversed-phase liquid chromatography with mixed eluents. J Chromatogr A. 1993 Dec 17;656(1):37–50.
99. Snyder LR. Principles of adsorption chromatography: The separation of nonionic organic compounds. [Internet]. Vol. 3. Marcel Dekker, Inc.; 1968. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jps.2600571053
100. Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr A. 1990 Jan 19;499:177–96.
101. Hemström P, Irgum K. Hydrophilic interaction chromatography. J Sep Sci. 29(12):1784–821.
102. Guo Y, Gaiki S. Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. J Chromatogr A. 2005 May 13;1074(1):71–80.
103. Li R, Huang J. Chromatographic behavior of epirubicin and its analogues on high-purity silica in hydrophilic interaction chromatography. J Chromatogr A. 2004 Jul 2;1041(1):163–9.
104. Olsen BA. Hydrophilic interaction chromatography using amino and silica columns for the determination of polar pharmaceuticals and impurities. J Chromatogr A. 2001 Apr 13;913(1):113–22.
105. Garbis SD, Melse-Boonstra A, West CE, van Breemen RB. Determination of Folates in Human Plasma Using Hydrophilic Interaction Chromatography−Tandem Mass Spectrometry. Anal Chem. 2001 Nov 1;73(22):5358–64.
106. Oyler AR, Armstrong BL, Cha JY, Zhou MX, Yang Q, Robinson RI, et al. Hydrophilic interaction chromatography on amino-silica phases complements reversed-phase high-performance liquid chromatography and capillary electrophoresis for peptide analysis. J Chromatogr A. 1996 Feb 16;724(1):378–83.
107. Naidong W. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. J Chromatogr B. 2003 Nov 5;796(2):209–24.
108. The Basics of HILIC [Internet]. [cited 2019 Dec 23]. Available from: https://www.chromacademy.com/channels/hplc-training-courses/technique/hilic/
109. Dill KA. The mechanism of solute retention in reversed-phase liquid chromatography. J Phys Chem. 1987 Mar;91(7):1980–8.
110. Ying PT, Dorsey JG, Dill KA. Retention mechanisms of reversed-phase liquid chromatography: determination of solute-solvent interaction free energies. Anal Chem. 1989 Nov 15;61(22):2540–6.
111. Wang HL, Duda JL, Radke CJ. Solution adsorption from liquid chromatography. J Colloid Interface Sci. 1978 Aug 1;66(1):153–65.
112. Riedo F, Kováts E sz. Adsorption from liquid mixtures and liquid chromatography. J Chromatogr A. 1982 Apr 30;239:1–28.

113. Knox JH, Pryde A. Performance and selected applications of a new range of chemically bonded packing materials in high-performance liquid chromatography. J Chromatogr A. 1975 Jan 1;112:171–88.
114. Thomson PSJJ, M O, S FR. Bakerian Lecture:—Rays of positive electricity. Proc R Soc Lond A. 1913 Aug 1;89(607):1–20.
115. Aston FW. A critical search for a heavier constituent of the atmosphere by means of the mass-spectrograph. Proc R Soc Lond Ser Contain Pap Math Phys Character. 1923;103(722):462–9.
116. Whitehouse CM, Dreyer RN, Yamashita Masamichi, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. Anal Chem. 1985 Mar 1;57(3):675–9.
117. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989 Oct 6;246(4926):64–71.
118. Kebarle P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. J Mass Spectrom JMS. 2000 Jul;35(7):804–17.
119. Watson JT, Sparkman OD. Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation, Fourth Edition [Internet]. 4th ed. John Wiley and Sons; 2008. Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470516898 pages 53-228 and 639-688
120. Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD, Alice MB. Molecular Beams of Macroions. J Chem Phys. 1968 Sep 1;49(5):2240–9.
121. Iribarne JV, Thomson BA. On the evaporation of small ions from charged droplets. J Chem Phys. 1976 Mar 15;64(6):2287–94.
122. Smith RD, Loo JA, Edmonds CG, Barinaga CJ, Udseth HR. New developments in biochemical mass spectrometry: electrospray ionization. Anal Chem. 1990 May 1;62(9):882–99.
123. Sargent M, Mussell C, Godfrey R, Sage A, Wolff C, Neville D, et al. Guide to achieving reliable quantitative LC-MS measurements [Internet]. RSC Analytical Methods Committee; 2013. 68 p. Available from: https://www.rsc.org/images/AMC%20LCMS%20Guide_tcm18-240030.pdf
124. Atmospheric pressure ion sources - Covey - 2009 - Mass Spectrometry Reviews - Wiley Online Library [Internet]. [cited 2018 Mar 12]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/mas.20246/full
125. Electrospray: From ions in solution to ions in the gas phase, what we know now - Kebarle - 2009 - Mass Spectrometry Reviews - Wiley Online Library [Internet]. [cited 2018 Mar 12]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/mas.20247/full
126. Konenkov NV, Cousins LM, Baranov VI, Sudakov MYu. Quadrupole mass filter operation with auxiliary quadrupolar excitation: theory and experiment. Int J Mass Spectrom. 2001 Jul 18;208(1):17–27.
127. Douglas DJ. Applications of Collision Dynamics in Quadrupole Mass Spectrometry. J Am Soc Mass Spectrom. 1998 Feb 1;9(2):101–13.
128. Mayer PM, Poon C. The mechanisms of collisional activation of ions in mass spectrometry. Mass Spectrom Rev. 2009 Aug;28(4):608–39.
129. McLuckey SA. Principles of collisional activation in analytical mass spectrometry. J Am Soc Mass Spectrom. 1992 Sep 1;3(6):599–614.

130. Jennings KR. Collision-induced decompositions of aromatic molecular ions. Int J Mass Spectrom Ion Phys. 1968 Aug 1;1(3):227–35.
131. Haddon WF, McLafferty FW. Metastable ion characteristics. VII. Collision-induced metastables. J Am Chem Soc. 1968 Aug 1;90(17):4745–6.
132. Stokvis E, Rosing H, Beijnen JH. Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not? Rapid Commun Mass Spectrom RCM. 2005;19(3):401–7.
133. Wieling J. LC-MS-MS experiences with internal standards. Chromatographia. 2002 Jan 1;55(1):S107–13.
134. Boyd RK, Basic C, Bethem RA. Wiley: Trace Quantitative Analysis by Mass Spectrometry - Robert K. Boyd, Cecilia Basic, Robert A. Bethem [Internet]. 2008 [cited 2017 Nov 14] - pages 177-200 - Available from: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470057718.html
135. Bedson P. Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS) [Internet]. 2002 [cited 2018 Jun 25]. Available from: http://pubs.rsc.org/en/Content/eBook/978-0-85404-418-4 - pages 1-34.
136. Sargent M, Harte R, Harrington C. Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS). In: Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS) [Internet]. 2002 [cited 2018 Sep 17]. pages 3–24. Available from: https://pubs.rsc.org/en/content/chapter/bk9780854044184-00001/978-0-85404-418-4
137. Burke DG, Mackay LG. Complete equation for the measurement of organic molecules using stable isotope labeled internal standards, exact matching, and mass spectrometry. Anal Chem. 2008 Jul 1;80(13):5071–8.
138. Vogl J. Characterisation of reference materials by isotope dilution mass spectrometry. J Anal At Spectrom 2007,22, 475-492 [Internet]. 2007 Feb 27; Available from: http://pubs.rsc.org/en/content/articlelanding/2007/ja/b614612k/unauth#!divAbstract
139. Bièvre PD, Laeter JRD, Peiser HS, Reed WP. Reference materials by isotope ratio mass spectrometry. Mass Spectrom Rev. 1993 May 1;12(3):143–72.
140. Trace Element Analysis in Biological Specimens, Volume 15 - 1st Edition [Internet]. [cited 2018 Sep 17]. Available from: https://www.elsevier.com/books/trace-element-analysis-in-biological-specimens/herber/978-0-444-89867-8
141. Henrion A. Reduction of systematic errors in quantitative analysis by isotope dilution mass spectrometry (IDMS): an iterative method. Fresenius J Anal Chem. 1994 Dec 1;350(12):657–8.
142. ISO 3534-2:2006 Statistics - Vocabulary and Symbols - Part 2: Applied statistics. [Internet]. ISO, Geneva, Switzerland; 2006. Available from: https://www.iso.org/standard/40147.html
143. Ellison S, Williams A. EURACHEM/CITAC Guide CG4 - Quantifying Uncertainty in Analytical Measurement [Internet]. Eurochem & CITAC - Co-Operation on International Traceability in Analytical Chemistry; 2012 [cited 2017 Nov 14]. Available from: https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf
144. IUPAC - conventional true value (C01316) [Internet]. [cited 2020 Dec 4]. Available from: https://goldbook.iupac.org/terms/view/C01316

145. GUM - English - Annex B. General metrological terms [Internet]. [cited 2020 Dec 4]. Available from: https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e_FILES/MAIN_C045315e/AB_e.html
146. ISO/IEC Guide 99:2007 - International vocabulary of metrology -- Basic and general concepts and associated terms (VIM) [Internet]. [cited 2017 Nov 14]. Available from: https://www.iso.org/standard/45324.html
147. ISO/IEC Guide 98-3:2008 - Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) [Internet]. [cited 2017 Nov 14]. Available from: https://www.iso.org/standard/50461.html
148. Joint Committee for Guides in Metrology. JCGM 100:2008 - GUM 1995 with minor corrections. Evaluation of measurement data — Guide to the expression of uncertainty in measurement [Internet]. 1st ed. 2008. 134 p. Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
149. Pritchard C. Traceable quantification of proteins using mass spectrometry. [Leeds, UK]: University of Leeds; 2012.
150. [VIM3] 2.38 coverage factor [Internet]. [cited 2019 Nov 29]. Available from: https://jcgm.bipm.org/vim/en/2.38.html
151. Annesley TM, McKeown DA, Holt DW, Mussell C, Champarnaud E, Harter L, et al. Standardization of LC-MS for Therapeutic Drug Monitoring of Tacrolimus. Clin Chem. 2013 Jul 31;clinchem.2013.209114.
152. White GH, Farrance I. Uncertainty of Measurement in Quantitative Medical Testing. Clin Biochem Rev. 2004 Nov;25(4):S1–24.
153. Annesley TM. Methanol-Associated Matrix Effects in Electrospray Ionization Tandem Mass Spectrometry. Clin Chem. 2007 Oct 1;53(10):1827–34.
154. Napoli KL. More on Methanol-Associated Matrix Effects in Electrospray Ionization Mass Spectrometry. Clin Chem. 2009 Jun 1;55(6):1250–2.
155. Keller BO, Sui J, Young AB, Whittal RM. Interferences and contaminants encountered in modern mass spectrometry. Anal Chim Acta. 2008 Oct 3;627(1):71–81.
156. Luo Y, Korfmacher W, Ho S, Shen L, Wang J, Wu Z, et al. Evaluation of two blood micro-sampling approaches for drug discovery PK studies in rats. Bioanalysis. 2015 Sep 1;7(18):2345–59.
157. Annesley TM. Ion Suppression in Mass Spectrometry. Clin Chem. 2003 Jul 1;49(7):1041–4.
158. Taylor PJ. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. Clin Biochem. 2005 Apr 1;38(4):328–34.
159. Matuszewski BK, Constanzer ML, Chavez-Eng CM. Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC−MS/MS. Anal Chem. 2003 Jul 1;75(13):3019–30.
160. Bonfiglio R, King RC, Olah TV, Merkle K. The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. Rapid Commun Mass Spectrom. 1999 Jun 30;13(12):1175–85.
161. Wang S, Cyronak M, Yang E. Does a stable isotopically labeled internal standard always correct analyte response?: A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma. J Pharm Biomed Anal. 2007 Jan 17;43(2):701–7.
162. Lindegardh N, Annerberg A, White NJ, Day NPJ. Development and validation of a liquid chromatographic-tandem mass spectrometric method for

determination of piperaquine in plasma: Stable isotope labeled internal standard does not always compensate for matrix effects. J Chromatogr B. 2008 Feb 1;862(1):227–36.

163. Bunch DR, M. El-Khoury J, Gabler J, Wang S. Do deuterium labeled internal standards correct for matrix effects in LC-MS/MS assays? A case study using plasma free metanephrine and normetanephrine. Clin Chim Acta. 2014 Feb 15;429(Supplement C):4–5.
164. Taylor PJ, Brown SR, Cooper DP, Salm P, Morris MR, Pillans PI, et al. Evaluation of 3 Internal Standards for the Measurement of Cyclosporin by HPLC–Mass Spectrometry. Clin Chem. 2005 Jan 10;51(10):1890–3.
165. Vogeser M. Instrument-Specific Matrix Effects of Calibration Materials in the LC-MS/MS Analysis of Tacrolimus. Clin Chem. 2008 Aug 1;54(8):1406–8.
166. Kushnir MM, Rockwood AL, Nelson GJ, Yue B, Urry FM. Assessing analytical specificity in quantitative analysis using tandem mass spectrometry. Clin Biochem. 2005 Apr 1;38(4):319–27.
167. Carvalho VM, Nakamura OH, Vieira JGH. Simultaneous quantitation of seven endogenous C-21 adrenal steroids by liquid chromatography tandem mass spectrometry in human serum. J Chromatogr B Analyt Technol Biomed Life Sci. 2008 Sep 1;872(1–2):154–61.
168. Careri M, Mangia A. Validation and qualification: the fitness for purpose of mass spectrometry-based analytical methods and analytical systems. Anal Bioanal Chem. 2006 Sep;386(1):38–45.
169. Calbiani F, Careri M, Elviri L, Mangia A, Zagnoni I. Matrix effects on accurate mass measurements of low-molecular weight compounds using liquid chromatography-electrospray-quadrupole time-of-flight mass spectrometry. J Mass Spectrom JMS. 2006 Mar;41(3):289–94.
170. Kaufmann A, Butcher P, Maden K, Widmer M, Giles K, Uría D. Are liquid chromatography/electrospray tandem quadrupole fragmentation ratios unequivocal confirmation criteria? Rapid Commun Mass Spectrom RCM. 2009 Apr;23(7):985–98.
171. Jansen R, Lachatre G, Marquet P. LC-MS/MS systematic toxicological analysis: Comparison of MS/MS spectra obtained with different instruments and settings. Clin Biochem. 2005 Apr 1;38(4):362–72.
172. Hopley C, Bristow T, Lubben A, Simpson A, Bull E, Klagkou K, et al. Towards a universal product ion mass spectral library - reproducibility of product ion spectra across eleven different mass spectrometers. Rapid Commun Mass Spectrom RCM. 2008 Jun;22(12):1779–86.
173. Oberacher H, Pavlic M, Libiseller K, Schubert B, Sulyok M, Schuhmacher R, et al. On the inter-instrument and inter-laboratory transferability of a tandem mass spectral reference library: 1. Results of an Austrian multicenter study. J Mass Spectrom JMS. 2009 Apr;44(4):485–93.
174. Kruve A, Rebane R, Kipper K, Oldekop M-L, Evard H, Herodes K, et al. Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. Anal Chim Acta. 2015 Apr 22;870:29–44.
175. European U. European Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. Off J Eur Coomun. 2002 Aug 12;L221:8–36.
176. Ng KT, Rapp-Wright H, Egli M, Hartmann A, Steele JC, Sosa-Hernández JE, et al. High-throughput multi-residue quantification of contaminants of emerging

concern in wastewaters enabled using direct injection liquid chromatography-tandem mass spectrometry. J Hazard Mater. 2020 Nov 5;398:122933.

177. The selection and use of reference materials - a basic guide for laboratories and accreditation bodies. Accreditation Qual Assur. 2003 Sep 1;8(9):429–33.
178. Taverniers I, De Loose M, Van Bockstaele E. Trends in quality in the analytical laboratory. I. Traceability and measurement uncertainty of analytical results. TrAC Trends Anal Chem. 2004 Jul 1;23(7):480–90.
179. Barwick V. Eurachem/CITAC Guide Guide to Quality in Analytical Chemistry An Aid to Accreditation, Third edition [Internet]. Eurochem & CITAC - Co-Operation on International Traceability in Analytical Chemistry; 2016 [cited 2018 Sep 10]. Available from: https://www.eurachem.org/images/stories/Guides/pdf/Eurachem_CITAC_QAC_2016_EN.pdf
180. Thompson M, Ellison SLR, Wood R. Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). Pure Appl Chem. 2002;74(5):835–855.
181. Pacak K, Eisenhofer G, Lenders J. Pheochromocytoma: Diagnosis, Localization, and Treatment. John Wiley & Sons; 2008. 185 p.
182. Pillai D, Callen S. Pilot quality assurance programme for plasma metanephrines. Ann Clin Biochem. 2010 Jan 3;47(2):137–42.
183. Lagerstedt SA, O'Kane DJ, Singh RJ. Measurement of Plasma Free Metanephrine and Normetanephrine by Liquid Chromatography–Tandem Mass Spectrometry for Diagnosis of Pheochromocytoma. Clin Chem. 2004 Jan 3;50(3):603–11.
184. PRESSOR - Pheochromocytoma Research Support Organization [Internet]. [cited 2016 Jan 4]. Available from: http://www.pressor.org/
185. Nankivell BJ, Borrows RJ, Fung CL-S, O'Connell PJ, Chapman JR, Allen RDM. Calcineurin Inhibitor Nephrotoxicity: Longitudinal Assessment by Protocol Histology. Transplantation. 2004 Aug 27;78(4):557.
186. Ekberg H, Tedesco-Silva H, Demirbas A, Vítko Š, Nashan B, Gürkan A, et al. Reduced Exposure to Calcineurin Inhibitors in Renal Transplantation. N Engl J Med. 2007 Dec 20;357(25):2562–75.
187. Ekberg H., Bernasconi C., Tedesco-Silva H., Vítko S., Hugo C., Demirbas A., et al. Calcineurin Inhibitor Minimization in the Symphony Study: Observational Results 3 Years after Transplantation. Am J Transplant. 2009 Jul 16;9(8):1876–85.
188. Wallemacq P, Armstrong VW, Brunet M, Haufroid V, Holt DW, Johnston A, et al. Opportunities to optimize tacrolimus therapy in solid organ transplantation: report of the European consensus conference. Ther Drug Monit. 2009 Apr;31(2):139–52.
189. NHS England. Prescribing Costs in Hospitals and the Community England 2015/16 [Internet]. 2016. Available from: https://files.digital.nhs.uk/publicationimport/pub22xxx/pub22302/hosp-pres-eng-201516-report.pdf
190. NHS England. Prescribing Costs in Hospitals and the Community England 2013/14 [Internet]. 2014. Available from: https://files.digital.nhs.uk/publicationimport/pub15xxx/pub15883/hosp-pres-eng-201314-rep.pdf

191. NHS England. Prescribing Costs in Hospitals and the Community England 2014/15 [Internet]. 2015. Available from: https://files.digital.nhs.uk/publicationimport/pub18xxx/pub18973/hosp-pres-eng-201415-report.pdf
192. NHS England. Prescribing Costs in Hospitals and the Community England 2016/17 [Internet]. 2017. Available from: https://files.digital.nhs.uk/publication/n/1/hosp-pres-eng-201617-report.pdf
193. Martín BB, Marquet P, Ferrer JM, Puig BC, Bennasar AB, Prieto MR, et al. Rheumatoid Factor Interference in a Tacrolimus Immunoassay. Ther Drug Monit. 2009 Dec;31(6):743.
194. Hermida J, Tutor JC. Falsely Increased Blood Tacrolimus Concentrations Using the ACMIA Assay Due to Circulating Endogenous Antibodies in a Liver Transplant Recipient: A Tentative Approach to Obtaining Reliable Results. Ther Drug Monit. 2009 Apr;31(2):269.
195. Moscato D, Nonnato A, Adamo R, Vancheri M, Caropreso A. Therapeutic monitoring of Tacrolimus: Aberrant results by an immunoassay with automated pretreatment. Clin Chim Acta. 2010 Jan 4;411(1):77–80.
196. Rostaing Lionel, Cointault Olivier, Marquet Pierre, Josse Anne-Gaelle, Lavit Michel, Saint-Marcoux Franck, et al. Falsely elevated whole-blood tacrolimus concentrations in a kidney-transplant patient: potential hazards. Transpl Int. 2009 Sep 15;23(2):227–30.
197. Parikh BA, Siedlecki AM, Scott MG. Specificity of a Circulating Antibody That Interferes With a Widely Used Tacrolimus Immunoassay. Ther Drug Monit. 2010 Apr;32(2):228.
198. Levine DM, Maine GT, Armbruster DA, Mussell C, Buchholz C, O'Connor G, et al. The Need for Standardization of Tacrolimus Assays. Clin Chem. 2011 Jan 12;57(12):1739–47.
199. LGC acquires Immunosuppressive PT scheme | LGC Standards [Internet]. [cited 2020 Feb 6]. Available from: https://www.lgcstandards.com/GB/en/news/LGC-acquires-Immunosuppressive-PT-scheme
200. Moes DJAR, Press RR, de Fijter JW, Guchelaar H-J, den Hartigh J. Liquid chromatography-tandem mass spectrometry outperforms fluorescence polarization immunoassay in monitoring everolimus therapy in renal transplantation. Ther Drug Monit. 2010 Aug;32(4):413–9.
201. Barau C, Frangie C, Goujard C, Tribut O, Parant F, Taburet A-M, et al. Falsely elevated whole blood tacrolimus concentrations due to interference in an affinity column-mediated immunoassay method on Xpand dimension. Ther Drug Monit. 2009 Apr;31(2):267–8.
202. Staatz CE, Taylor PJ, Tett SE. Comparison of an ELISA and an LC/MS/MS method for measuring tacrolimus concentrations and making dosage decisions in transplant recipients. Ther Drug Monit. 2002 Oct;24(5):607–15.
203. Altinier S, Varagnolo M, Zaninotto M, Boccagni P, Plebani M. Heterophilic antibody interference in a non-endogenous molecule assay: an apparent elevation in the tacrolimus concentration. Clin Chim Acta Int J Clin Chem. 2009 Apr;402(1–2):193–5.
204. Napoli KL. Is microparticle enzyme-linked immunoassay (MEIA) reliable for use in tacrolimus TDM? Comparison of MEIA to liquid chromatography with mass spectrometric detection using longitudinal trough samples from transplant recipients. Ther Drug Monit. 2006 Aug;28(4):491–504.

205. Brown NW, Gonde CE, Adams JE, Tredger JM. Low hematocrit and serum albumin concentrations underlie the overestimation of tacrolimus concentrations by microparticle enzyme immunoassay versus liquid chromatography-tandem mass spectrometry. Clin Chem. 2005 Mar;51(3):586–92.
206. Taylor PJ, Morris RG. Tacrolimus measurement by microparticle enzyme immunoassay II. Ther Drug Monit. 2003 Jun;25(3):259–60.
207. Taylor PJ, Franklin ME, Tai C-H, Pillans PI. Therapeutic drug monitoring of tacrolimus by liquid chromatography-tandem mass spectrometry: is it truly a routine test? J Chromatogr B Analyt Technol Biomed Life Sci. 2012 Feb 1;883–884:108–12.
208. Human Metabolome Database [Internet]. [cited 2019 Nov 11]. Available from: http://www.hmdb.ca/
209. T3DB: Tacrolimus [Internet]. [cited 2019 Nov 25]. Available from: http://www.t3db.ca/toxins/T3D4732

2 Understanding measurement uncertainty and variability in higher-order reference measurement procedure development: endogenous small polar molecules in plasma samples

2.1 Introduction

2.1.1 Hypothesis and objectives

In line with the aim of this thesis expressed in Chapter 1 (section 1.13.2.1), the hypothesis for this chapter was that “all the major sources of measurement uncertainty could be reliably determined for the measurements of clinically relevant small molecules (METs) at the lower end of the molecular weight range”.

In order to test this hypothesis, the following objectives were set:

- Develop a DEM-IDMS based LC-MS method as a candidate RMP to quantify plasma METs.
- Evaluate the sources of measurement uncertainty of the different elements of the RMP.
- Optimise the elements of the RMP to mitigate the sources of variability in measurement to reduce the measurement uncertainty of the method.
- Use the candidate RMP to analyse human plasma to assess the estimated overall measurement uncertainty of the method.

2.1.2 Pheochromocytoma and Paraganglioma

The diseases known as pheochromocytoma and paraganglioma (PPGLs) are poorly diagnosed. As stated by the Pheochromocytoma Research Support Organisation (PRESSOR); '*Pheochromocytomas are rare, with an annual detection rate of two to four per million. A relatively high prevalence of the tumour in autopsy studies (1:2000) suggests that many of these tumours are missed, resulting in premature death. The actual annual incidence is therefore likely to be near 10 per million*' (1,2).

PPGLs can be fatal because if a patient was misdiagnosed and underwent surgery, the tumours secrete a sudden spike of catecholamines (CATs) during a surgical procedure. This would cause a cardiac arrest particularly in patients with cardiovascular disorders. However, these tumours are treatable by surgical removal and chemotherapy (3). Therefore, improving the biochemical diagnostic analysis could save patients' lives (3,4).

PPGLs diagnosis is informed by biochemical testing of metanephrines (METs) in urine or plasma samples after years of the diagnostic test evolved from initially measuring CATs. The METs testing proved to have a higher clinical diagnostic sensitivity and specificity (3,4). This is because METs are phase I metabolites of CATs and dopamine; where norepinephrine, epinephrine are metabolised to normetanephephrine (Normet), metanephrine (Met), respectively, and dopamine is metabolised to 3-methoxytyramine (3MT) (5).
One of the reasons the diagnosis of these tumours is poor is that the means of getting reliable measurements, that a higher-order reference measurement analysis is likely to achieve, are lacking. Therefore, because of their clinical

importance, the analysis of METs in plasma was selected as the case study application for this thesis; to understand sources of measurement uncertainty in the development of higher-order reference measurement procedures (RMPs) of small polar molecules in biological samples.

Developing a higher-order RMP for the measurement of plasma METs would enable assigning higher-order reference traceable measurements to certified reference materials in matrix and to external quality assurance scheme (EQAS) samples. This would provide the hospital laboratories with the tools to assess and improve their analytical methods hence improve diagnosis.

PPGLs are neuroendocrine tumours that arise from chromaffin cells. The tumour is called pheochromocytoma when it occurs in the chromaffin cells of the adrenal gland medulla. When the tumour occurs outside the adrenal glands in the paraganglia, near the sympathetic ganglia throughout the body, it is then called paraganglioma or extra-adrenal pheochromocytoma (Figure 2.1) (6,7). PPGLs can be either sporadic or hereditary, occurring in patients of different age groups with the highest prevalence between 40 and 50 years old. When reported in children PPGL are typically hereditary and are paragangliomas (8). All types of PPGLs secret CATs and some additionally secrete dopamine. Paraganglioma are more likely to secrete dopamine than pheochromocytoma. The secretion of CATs by PPGLs could be constant or intermittent. This increased level of CATs in the human body could manifest in several symptoms that include headache, sweating, tachycardia, palpitation and hypertension. These tumours are treatable however misdiagnosed patient could die as mentioned earlier in cases of surgical operations. PPGL

diagnosis could be a lengthy process if the biochemical test of METs in plasma or urine is inaccurate. The diagnostic process is summarised in Figure 2.2.

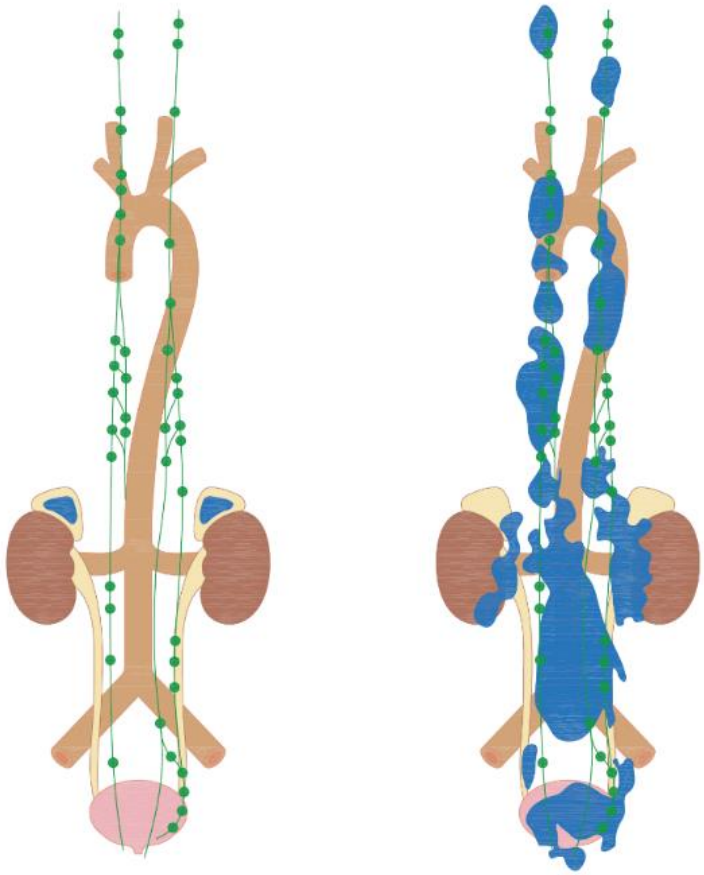


Figure 2.1 PPGLs potential sites shaded in blue. Left: Pheochromocytoma sites, Right: Paraganglioma. Figure drawn with adaptation from reference (7)

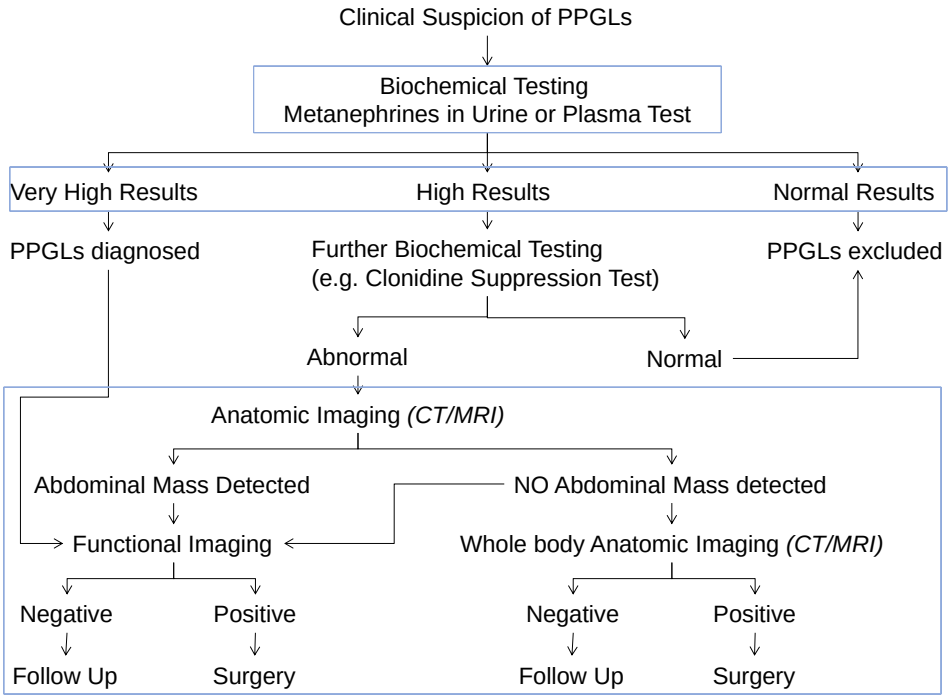


Figure 2.2, PPGL diagnostic process flowchart, adapted from (9,10)

2.1.3 Diagnostic assay evolution

Biochemical testing for PPGLs has evolved over the last two decades. Initially diagnosis was based on elevated concentrations of vanillylmandelic acid (VMA) in urine (11). This was superseded by the use of elevated CATs in plasma and urine, which was then replaced by the use of METs in plasma and urine (11). The evolution of the diagnostic assay has resulted in some confusing terms to describe the various assays (12), which are explained in Table 2.1.

Table 2.1 Terminologies commonly used for Metanephrines Analysis for PPGL Diagnosis

Names of routinely used METs assays	Analysis Description
Urinary Total (Normet & Met only)	Aliquot taken from 24-hour urine sample collection. Urine hydrolysed. Older analytical techniques unable to independently measure Met & Normet. Result reported as a total.
Urinary Fractionated (Normet & Met only)	Aliquot taken from 24-hour urine sample collection. Urine hydrolysed. Chromatographic techniques enabled Met & Normet to be measured individually – as 'fractionated' compounds.
Plasma Free (Normet & Met only)	Plasma analysed without any hydrolysis steps – any conjugated METs are excluded from diagnostic test.

2.1.4 An evaluation of assay performance using plasma and urine

Approaches to PPGL diagnosis vary from hospital to hospital depending on local decisions and equipment availability (globally and nationally). Several large-scale studies have been published which evaluated and compared plasma and urinary based assays' diagnostic selectivity and sensitivity. A large multicentre cohort study of 1003 patients (over 14 years) by Lenders *et al.*, compared diagnostic selectivity and specificity of five different assays; plasma CATs, plasma METs, urinary CATs, urinary METS and urinary VMA. The study concluded that plasma METs is preferable if only one assay is to be used for PPGL diagnosis (13).

Moreover, urine-based assays require the patient to collect all their urine over a period of 24 hours and storing the large collection vessel in their home fridge. In addition to the inconvenience of the sampling process to the patient, this also jeopardises the accuracy of the sampling process. The 24-hour urine collection could introduce a source of variability in the sampling process as it would rely on the patient collecting the entire volume of all urine at each urination for 24 hours. This could result in variability of the measurement due to the different volumes combined (14).

Although the plasma METs assay demonstrated the best diagnostic selectivity and sensitivity it still suffers from several challenges pre-analytically and analytically. At the patient sampling stage, the subsequent plasma METs assay was reported to give higher false positives when venous forearm blood samples were collected from a patient in a seated position rather than supine position (15,16). Lenders *et al*. found that the rate of false positives raised from 9 % to 25 % when patients' blood was collected while they are in a seated position without prior resting time rather than laying down (15). Similar positive biases were observed when sampling patients that were not abstaining from caffeine intake (17). Darr *et al.* have also reported the use of different cut-off limits from supine to seated position in sampling caused a drop in the diagnostic sensitivity of the plasma test from 98 % to 85 % (17). The different laboratories and clinical centres debated which cut-off limits to use and whether the supine or seated blood assay should be the first line of diagnosis of PPGL. However, the data of the different studies demonstrate clearly that the supine sampling and using supine collected cut-off limits are of the highest diagnostic specificity (11,17–19).

Several important questions arise when presented with the data about the assays in the different studies including:

- How can this data be evaluated when none of the reported measurements made are traceable?
- How can clinical cut-off ranges be assigned for clinical decisions when the measurements made suffer from high variability inter and intra laboratory?

Additionally, throughout the different publications, the measurement uncertainty of the analytical assay was not assessed, results were continuously reported with no measurement uncertainty estimate and were not traceable. Understanding and mitigating the analytical sources of uncertainty in the measurement procedure of the analytical assay is the key aim of this work, to provide measurements that ultimately are metrologically traceable and of low measurement uncertainty. This would therefore support the assignment of reference ranges and improve the diagnostic sensitivity and selectivity of the plasma METs assay.

2.1.5 Analytical approaches to METs measurement

2.1.5.1 METs, the compounds

Currently, the diagnostic assays all measure Normet and Met for the diagnosis. Although the clinical utility of 3MT measurement remains under discussion on whether to add it to the panel of the assay is ongoing, some laboratories measure it (20). The structures, molecular weights (MW), partition coefficients (LogP) and other details of the three compounds are shown in Figure 2.3. The structures and properties of the stable isotopically-labelled internal standards (SILIS) that were used in this research are also shown in Figure 2.3.

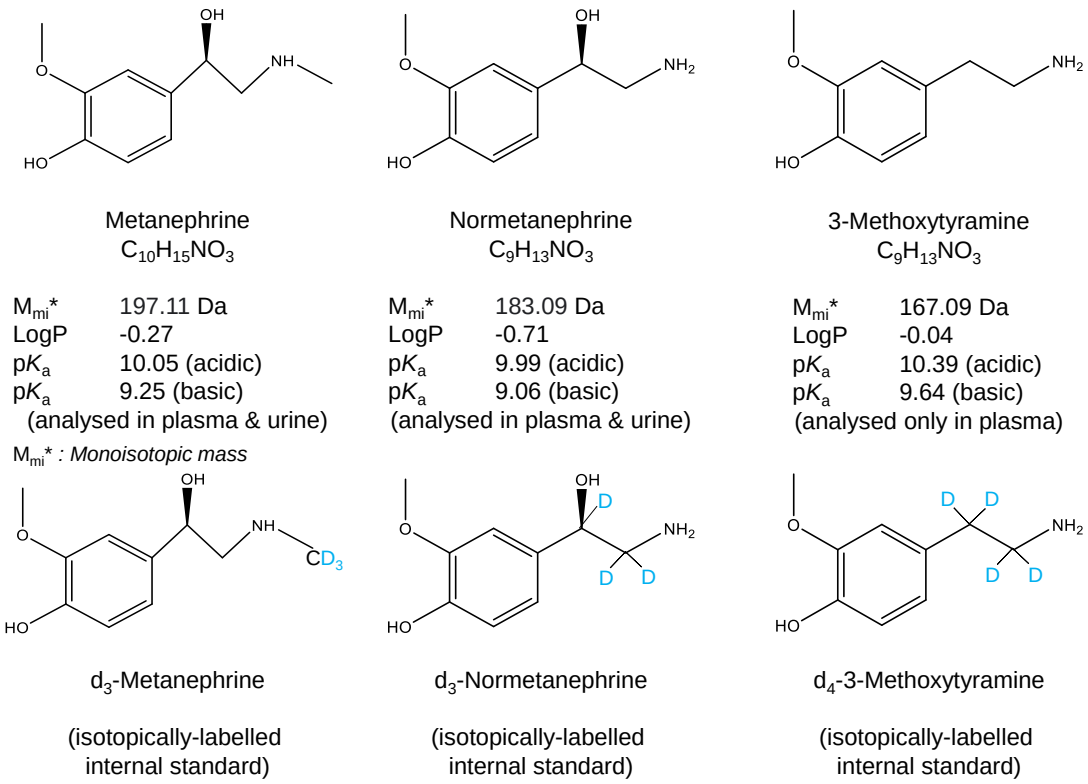


Figure 2.3 Metanephrines structures, formula, Log P, pKa are obtained the Human Metabolome Database (21). The structures of the stable isotopically-labelled metanephrines are from the certificates of analysis of the standards used in this research (see 4.2.1).

2.1.5.2 Immunoassays, LC-ECD and GC-MS based methods

In the early 1990s analytical methods for CATs mainly used competitive immunoassay approaches, these being either radio-isotopic, enzymatic or fluorometric (22). These immunoassays have since been replaced with liquid chromatography (LC) or gas chromatography (GC) coupled to electrochemical detectors (ECD). ECDs were used as detector because they were more sensitive than (UV) detectors which were typically used before. While ECD is still used as a detection method, mass spectrometry (MS) including tandem mass spectrometry (MS/MS) has become increasingly used for this assay with the rise in the use of MS in the clinical laboratory (see 1.5) (22). With the evolution of the clinical test and the advancement in the technology used in

analysis the clinical analytical community reviewed the different technologies. Most recently, the Pheochromocytoma Clinical Practice Guideline 2014 does not recommend enzymatic immunoassays (EIA) to quantify CATs and METs (11). Enzymatic immunoassay based method for the analysis of METs in plasma and in urine, are known to be prone to cross-reactivity, non-specific binding of compounds and high imprecision (23,24).

The quantification of METs in brain tissue and urine samples *via* GC-MS has been reported. When GC is used for METs analysis, a derivatisation step is required to increase the volatility of the METs which complicated the analysis by the interferences introduced in the derivatisation process, and the additional time and labour required when processing a lot of samples (25–27). METs plasma concentrations are analytically challenging being present in the pg/mL range in addition to the increased risk of the presence of isobaric interferences at low *m/z (such as* isomers of particular biomolecules and fragments of large molecules) could impact the quantification significantly (28).

LC coupled to UV, fluorescence and electrochemical detectors, have been used for the quantification of CATs and METs in plasma and urine since the 1970s (11). LC-ECD remains a frequently used technique for the analysis of METs due to the low cost of the instrumentation (12,29,30). However, LC-ECD methods have been reported to be affected by interferences from diet and medications because of their lower selectivity compared to MS/MS. Among the most commonly reported interferences in the LC-ECD methods is paracetamol as it usually co-elutes with the Normet peak (31), other studies

improved the chromatographic separation to remove this interference by using a different column and method (32). However this interference with paracetamol was also reported in urinary METs analysis by LC-ECD (33). With LC-ECD methods, unknown interferences from curry leaves through dietary intake have been reported to interfere by co-eluting with the (unspecified) Recipe® internal standard causing signal suppression and subsequently negative bias (34). The antihypertensive labetalol was identified as an interferent (35) as were the anti-inflammatory drugs sulfasalazine and mesalazine (36,37).

2.1.5.3 LC-MS based methods

The clinical laboratories moved to the use of LC-MS based methods due to several factors including; a) the shortfalls of the analytical techniques mentioned above (see 2.1.5.2), b) the advancement of MS technology and its versatility and c) its ability at reducing costs of analysis after the initial investment in the instrument purchase.

In addition to the pre-analytical challenges there are several analytical issues with plasma METs analysis by LC-MS. Measuring METs in plasma with low measurement uncertainty could be challenging due to many factors, most importantly, the very low concentration that METs are typically measured at in plasma. As an example, the University Hospital of South Manchester (UHSM) set their clinical cut-off limits to be <101 pg/mL for Met, <216 pg/mL for Normet and <30 pg/mL for 3MT (38). Measuring at these low levels with confidence is analytically challenging as it introduces technical challenges with instrument sensitivity. To obtain high confidence in the measurements, the detection

sensitivity needs to be high, this would reduce variability in measurement caused by the instrument which directly influences the measurement uncertainty.

Clinical cut-off limits are higher in urine analysis: <197 µg/24 h for Met and <550 µg/24 h for Normet whereas 3MT is not reported to be measured in urine. The unit of measurement is weight per time rather than weight per volume because the analysis is performed on an aliquot taken from a pool of all of the patient's urinations collected over 24 hours (14).

All three compounds have low relative molar mass (RMM); 183, 197 and 167 g/mol for Normet, Met and 3MT, respectively. The low molecular mass of METs increases the probability of interference issues. Additionally, structurally related compounds such as the bronchodilator isoetharine need to be chromatographically resolved because it co-elutes with Met and Normet causing significant ionisation suppression (24). Other structurally related interferences reported are the illicit drugs and metabolites such as 3,4-methylenedioxyamphetanine and 4-hydroxy-3-methoxymethamphetamine (HMMA) (39). Wright *et al.* described the use of multiple reaction monitoring (MRM) with multistage fragmentation as an approach to enhance selectivity and overcome isobaric interferences in LC-MS methods (40). However, this approach was reported in a follow-up paper not to resolve the interference of O-methyldopa, a metabolite of the Parkinson's disease medication L-dopa (41).

2.1.5.4 Current LC-MS methodologies

2.1.5.4.1 The chromatographic methods

Due to the polarity and structural similarity of METs, they are not easily resolved or retained on the commonly used C_{18} RP-LC as they elute near the solvent front. The use of hydrophilic interaction liquid chromatography (HILIC), generally better to separate such polar analytes, is therefore commonly applied. However, RP-LC is still used with ≤5 % organic starting mobile phase to enable initial retention on the column stationary phase (42). For RP-LC separations it is necessary to match the final extract composition with the initial mobile phase composition to achieve retention and often methods employ a solvent exchange step to achieve this.

Table 2.2 summarises some of the LC-MS plasma METs analytical methods that are published in the literature. While the methods vary in quality and assay performance, they are deemed fit-for-purpose for use in routine clinical laboratory. However, none of the papers summarised in the table mentioned anything about measurement uncertainty and traceability. In the UK, with the transition to UKAS medical laboratory accreditation (ISO 15189:2012), as part of the accreditation, hospital laboratories are now required to evaluate measurement uncertainty. The laboratories are required to consider the measurement uncertainty when interpreting results and to make the uncertainty estimate available to users of data if they request it (section 5.5.1.4 of the ISO standard). As for traceability, under section 5.3.1.4 of ISO 15189:2012, laboratories need to have documented procedures for the calibration of equipment that directly or indirectly affects results. Furthermore,

results should be *'reported in SI units, units traceable to SI units, or other applicable units'* (43). With the absence of a higher-order reference measurement procedure, it is challenging for the laboratories to avoid biased results due to the lack of metrologically traceable reference measurements to compare against. Furthermore, the need to understand the sources of uncertainty in the measurement procedure and how to best mitigate them to reduce the uncertainty is an essential part to developing a higher-order RMP.

Table 2.2 Chromatography Comparison - Summary of the most cited LC-MS Methods

Ref.	Sample volume (µL)	Injection volume (µL)	Chromatography Mode	LC Column (stationary phase chemistry)	Aqueous MP	Organic MP	Flow Rate (mL/min)	LLOQ	Reproducibility	Run Time (min)	Traceability & Measurement uncertainty
Adaway *et al.* (44)	150	NR – online SPE	HILIC – Gradient elution	Atlantis HILIC Silica 2.1 x 50 mm, 3 µm (silica)	100 mM Am F* solution pH 3.2	Acetonitrile	0.30	Met 37.5 pmol/L Normet 75 pmol/L	Inter-assay and intra-assay CV % reported up to 9.5 %	7.15	NR
Wright *et al.*(40)	100	15	HILIC – Gradient Elution	Ascentis Express HILIC 2.1 x 50 mm, 2.7 µm (bare silica)	25 mM Am F solution pH 3	Acetonitrile	NR	Met 0.05 pmol/L Normet 0.1 pmol/L	Inter-assay imprecision & % for Normet and 9.9 % Met	7.0	NR
Petteys *et al.*(24)	200	35	HILIC – Gradient Elution	Atlantis HILIC Silica 2.1 x 50 mm, 3 µm (silica)	100 mM Am F solution pH 3	Acetonitrile	0.35	0.1 nmol/L	Total %RSD up to 20 %	3.5	NR
Peaston *et al.*(45)	100	35	HILIC – Gradient Elution	Atlantis HILIC Silica 2.1 x 50 mm, 3 µm (silica)	100 mM Am F solution pH 3	Acetonitrile	0.35	Met 0.04 nmol/L Normet 0.05 nmol/L 3MT 0.06 nmol/L	Inter-assay %RSD Met 16 % Normet 13 % 3MT 18 %	3.5	NR

The table continues the following page

Graham *et al.*(46)	40	NR – online SPE	HILIC – Gradient Elution	Waters HILIC 2.1 mm x 50 mm, 3 µm (unbonded ethylene bridged hybrid substrate)	100 mM Am F solution pH 3	Acetonitrile	0.30	Met 0.04 nmol/L Normet 0.16 nmol/L	Inter-assay %RSD Met up to 14 % Normet 15 %	7.15	NR
Peitzsch *et al.*(42)	900	NR	RP-LC – Gradient Elution	Acquity UPLC HSS T3 2.1 x 100 mm, 1.8 µm (high strength silica with C_{18} ligands)	0.2 % FA in water	0.2 % FA in methanol	0.53	Met 0.020 nmol/L Normet & 3MT 0.024 nmol/L	Inter-assay %RSD Met up to 11.7 % Normet up to 8.4 % 3MT up to 11.4 %	5.0	NR
Gabler *et al.*(47)	500	20	RP-LC – Gradient Elution	Ultra PFP Propyl column *column dimensions not reported* (single silica with pentafluorophenyl propyl ligands)	1 mM Am F in water + 0.1 % FA	1 mM Am F in methanol + 0.1 % FA	0.5 ramp to 0.7	Met 0.03 nmol/L Normet 0.08 nmol/L	Inter-assay %RSD Met up to 9.6 % Normet up to 12.9 %	5.75	NR
Lagerstedt *et al.* (48)	1000	30	RP-LC – Isocratic Elution	Luna Cyano 4.6 x 150 mm, 5 µm (fully porous silica with cyanide ligands)	40:60 acetonitrile: water	NA	1.5	Met 0.2 nmol/L Normet 0.2 nmol/L	Inter-assay %RSD Met up to 9.2 % Normet up to 13.0 %	6.0	NR
* Am F = Ammonium formate – NR = Not reported											

2.1.5.4.2 The sample preparation method

Sample preparation prior to LC-MS analysis is an essential step when quantifying sub- ng/mL target analytes in complex matrices such as plasma to remove interferences and concentrate the target analyte if possible. Plasma is usually pre-treated (diluted with a buffer or water), internal standard (IS) added and cleaned up by solid-phase extraction (SPE). The SPE chemistry most reported in the literature for METs is mixed-mode cation exchange, incorporating weak cation exchange (WCX) with $R\text{-}COO^-$ group mixed with reversed-phase functions such as divinylbenzene (DVB) polymer in Waters™ Oasis® WCX SPE plates. These stationary phase chemistries are prominently selected due to the polarity and structures of METs as they are weak bases. At the suitable pH, the weak cation groups on the stationary phase would bind to the ionised METs, additionally, the mixed mode SPE containing the DVB functions allows for pi-pi interactions with the benzyl rings of the METs structures. Table 2.3 summarises some of the SPE methods that are in the literature.

Table 2.3 SPE Methods Comparison - Summary of the most cited SPE Methods

Ref.	LC	Plasma Volume (µL)	Sample Pre-treatment	SPE Phase	Conditioning Step	Loading Step	Wash Steps	Elution Step	Reconstitution Step
Adaway *et al.* (44)	HILIC	100	250 µL water	Oasis WCX	-200 µL acetonitrile +2 % FA -250 µL 80:20 acetonitrile:10mM Am F, pH 3 -250 µL 95 % acetonitrile	100 µL Sample +250µL Water	-200 µL water -200 µL 95 % acetonitrile	LC MP - Online SPE	Online SPE
Wright *et al.*(40)	HILIC	100	100 µL 10 mM ammonium phosphate (AM Ph) containing IS	µElution Oasis WCX	Not Reported	Not Reported	Not Reported	3x30 µL 95:3:2 acetonitrile: water:FA	No reconstitution step
Petteys *et al.* (24)	HILIC	200	200 µL aqueous IS	µElution Oasis WCX	-200 µL of methanol -200 µL water	200 µL plasma + 200 µL IS mixed 200 µL aliquot loaded on SPE	-200 µL water -200 µL methanol -200 µL acetonitrile +0.1 % FA	2x50 µL acetonitrile +2 % FA	No reconstitution step

Peaston *et al.* (45)	HILIC	100	100 µL 10 mM Am Ph + 25 µL Aqueous IS	Waters™ Oasis µElution WCX	-200 µL of methanol -200 µL 10mM Am Ph	100 µL Plasma +100 µL 10 mM Am Ph + 25 µL Aqueous IS	-200 µL water -200 µL methanol -200 µL acetonitrile +0.2 % FA	3x 25 µL 95:5 acetonitrile: water +2 % FA	No reconstitution step
Graham *et al.* (46)	HILIC	40	1:1 dilution with aqueous IS solution	Oasis WCX	1 mL acetonitrile	1 mL water	1 mL water	LC MP - Online SPE	Online SPE
Peitzsch *et al.* (42)	RP-LC	900	975 µL 10mM ammonium acetate (Am Ac) + 20 µL IS	Oasis MCX (mixed mode CX)	-500 µL methanol -500 µL water -500 µL 10mM Am Ac, pH 6.5	900 µL Plasma+ 975 µL 10mM Am Ac + 20 µL IS	-500 µL 2 % FA solution -500 µL methanol	2x 100 µL 5 % methanolic ammonia	lyophilised Recon. in 100 µL of 2 % acetonitrile 0.2 % FA
Gabler *et al.* (47)	RP-LC	500	500 µL of 10 mM Am Ph, pH 6.5 + 25 µL IS	Oasis WCX	-1 mL methanol -1 mL 10 mM Am Ph, pH 6.5	500 µL plasma +500 µL of 10 mM Am Ph pH 6.5 + 25 µL IS	-1 mL water -1mL methanol -1 mL acetonitrile +0.2 % FA	500 µL acetonitrile +2 % FA	100 µL 10 mM Am F +0.1 % FA
Lagerstedt *et al.* (48)	RP-LC	1000		Oasis HLB (RP hydrophilic-lipophilic balanced)	-1 mL methanol -1 mL water		2 mL water	1 mL methanol	100 µL methanol

2.1.6 Standardisation & Harmonisation of METs Analysis

The background on the need for standardisation of plasma METs was described in Chapter 1 (see 1.11). To reiterate briefly, there is only one EQAS for plasma METs available worldwide organised through the Royal College of Pathologists of Australia (RCPA). RCPA-EQA sends out both spiked and patient plasma lyophilised samples monthly. Due to a lack of higher-order certified reference materials (CRMs) and RMPs an all-laboratories consensus mean value is used to evaluate data. Figure 1.15 in Chapter 1 is an example of RCPA-EQA sample data (38).

The need for value assignment through recognised RMP to enable routine laboratories to determine the accuracy of their methods has been identified by PRESSOR, the global research organisation on PPGLs (2). This is critical to support the hospital laboratories measure their methods against a traceable reference value instead of an interlaboratory consensus mean that is used currently.

2.2 Experimental

2.2.1 Reagents and standards

SILIS of $^{2}H_{3}$-Met, $^{2}H_{3}$-Normet and $^{2}H_{4}$-3MT Sigma-Aldrich (Dorset, UK) were obtained at 1 µg/mL for $^{2}H_{3}$-Met, $^{2}H_{3}$-Normet and prepared at 40 ng/mL for $^{2}H_{4}$-3MT.

Methanol and acetonitrile used were HPLC Optigrade® (LGC Standards, UK). Ammonium acetate (NH_4OAc, >99.99 %) and ammonium formate (NH_4FA,

>99.99 %) (LGC Standard, Teddington, UK) and formic acid (>99.5 %) (Fisher Scientific, New Hampshire, USA) were used for different buffers preparation. High purity water (18.2 MΩ.cm) was generated from an ELGA PureLab Flex System.

2.2.2 Instrumentation

2.2.2.1 LC-MS/MS method

Agilent 6490 Tandem Mass Spectrometer with Agilent 1260 Infinity Capillary Pump liquid chromatography system (Agilent, Santa Clara, California, USA) was used. The mass spectrometer was operated in positive ionisation electrospray mode. There were two instruments that were used: the Agilent 6490 and the AB Triple Quad Sciex 6500^{++}.The conditions of each of the MS instruments that were used are in Table 2.4 and Table 2.5, respectively. All the conditions and SRM transitions for both instruments were optimised manually using flow injections and LC injections.

Table 2.4 Agilent 6490 MS parameters

Delta EMV (+)	500	Sheath Gas Flow (L/h)	6
Delta EMV (-)	0	Capillary (V)	3500
Gas Temperature	290 °C	Nozzle Voltage (V)	2000
Gas Flow (L/h)	18	High Pressure RF (+ & -)	150
Sheath Gas Temp	300 °C	Low Pressure RF (+ & -)	60
Dwell time (ms) – Normet & Met	50	Cell accelerator voltage (V)	5
Dwell time (ms) – 3MT	75	Nebulizer pressure (psi)	30

A quantitative SRM transition as well as a confirmation SRM transition were acquired for each compound and its SILIS. The SRM transitions are summarised in Table 2.5.

Table 2.5 Metanephrines quantification and confirmation SRM transitions

Analyte	Precursor ion (*m/z*)	Precursor ion anticipated formula	Product ion (*m/z*)	Collision Energy (eV)
Normet	166.2	$[(M-H_2O)+H]^+$	134.1*	17
			106.1	15
D3-Normet	169.2	$[(M-H_2O)+H]^+$	137.1*	17
			109s.1	15
Met	180.2	$[(M-H_2O)+H]^+$	148.1*	17
			165.1	17
D3-Met	183.2	$[(M-H_2O)+H]^+$	151.1*	17
			168.1	17
3MT	151.2	$[(M-NH_3)+H]^+$	91.0*	20
			119.0	13
D4-3MT	155.2	$[(M-NH_3)+H]^+$	95.0*	20
			123.0	13
*Quantification SRM				

The micro-flow liquid chromatography method used a mixed mode reversed-phase C_{18} with pentafluoropentyl (PFP) column; Acquity UPLC HSS PFP column with the dimensions of 150 mm x 1 mm and 1.8 µm particle size supplied by Waters™ (Wilmslow, UK). Mobile phase solvents were water with 0.1 % formic acid (bottle A) and acetonitrile with 0.1 % formic acid (bottle B). The flow rate was optimised to 30 µL/min. The gradient was started with 2 % B held for 3 minutes then ramped to 25 %B over 7 minutes and another quick ramp to 98 %B which is held to wash the column for 3 minutes before going back to starting conditions to equilibrate the column for 12 min. Sample injection volume was optimised to 8 µL.

2.2.2.2 LC-MS/MS method after further optimisation

At one point during this research the Agilent 6490 instrument was no longer available for use and the method was transferred to the AB Sciex Triple Quad™ 6500$^+$ Mass Spectrometer (AB Sciex UK Ltd, Macclesfield, UK). The

same LC system was used Agilent 1260 Infinity Capillary Pump liquid chromatography system (Agilent, Santa Clara, California, USA) was used. The mass spectrometer was operated in positive electrospray ionisation mode and the conditions of each of the MS instruments that were used are in Table 2.6.

Table 2.6 AB Triple Quad Sciex 6500^{+} optimised conditions

Condition	Value
Ion spray voltage (V)	4000
Source temperature (°C)	300
Ion source gas 1 (psig)	10
Ion source gas 2 (psig)	10
Collision gas (psig)	5
Curtain gas (psig)	20

A quantitative SRM transition as well as a confirmation SRM transition were acquired for each compound and its SILIS. The SRM transitions are summarised in Table 2.7.

Table 2.7 Metanephrines quantification and confirmation SRM transitions on the AB Triple Quad Sciex 6500^{++}

Analyte	Precursor ion (*m/z*)	Precursor ion anticipated formula	Product ion (*m/z*)	Collision Energy (eV)	Declustering Potential (eV)
Normet	166.2	$[(M-H_2O)+H]^+$	134.2*	27	100
			106.0	25	100
D3-Normet	169.2	$[(M-H_2O)+H]^+$	137.2*	27	100
			109.0	25	100
Met	180.2	$[(M-H_2O)+H]^+$	148.1*	30	80
			165.2	25	80
D3-Met	183.2	$[(M-H_2O)+H]^+$	151.1*	30	80
			168.2	25	80
*Quantification SRM					

2.2.2.3 MS instruments maintenance

All MS instruments used in this research were maintained regularly. There was a system suitability check run before each analytical batch to assess the instrument performance to eliminate the instrument condition (e.g. requiring cleaning) as a source of variability. When an instrument failed the system suitability, it was cleaned. In some cases, the instrument needed to be cleaned every other week. The cleaning covered the source, cone and ion optics. Any deeper cleaning was performed by an engineer when needed.

2.2.2.4 Infusion experiments

All infusion experiments performed were direct infusion using a syringe pump at flow rate of 30 µL/min with the standard used was in 98:2 water:acetonitrile at 100 pg/g.

2.2.2.5 LC method development instrumentation

Several LC columns were evaluated for the analysis of metanephrines, using standard flow and micro-flow UHPLC instruments. Table 2.8 lists the columns tested at the method development stage. The columns were selected among the columns available for testing based on their chemistries and their likelihood of retaining the target analytes. Several types of RP-LC columns were selected because, each of these columns had a unique aspect whether an added chemical function to the C_{18} chain on the stationary phase. The HILIC columns were selected to test the different available HILIC stationary phases. The expression “normal flow LC” was used throughout this thesis to

indicate LC with flow rates typically above 0.2 mL/min as opposed to “micro-flow LC” where the flow rates are typically less than 100 µL/min and for this method, it was 30 µL/min.

Table 2.8 LC columns that were tested for the analysis and the flow rates they were used.

Column	Stationary Phase	Particle size (µm)	Pore size (Å)	Column Diameter (mm)	Column Length (mm)	Carbon Load (%)	Flow Rate (mL/min)	Void Volume * (mL)	T_0** (min)
Acquity UPLC HSS T3 (Waters™)	High strength Silica with C_{18} chains	1.8	100	2.1	100	11	0.4	0.24	0.61
Excel 3 C_{18}-HL (ACE)	C_{18} with high surface area, high carbon load phase	3	90	2.1	150	20	0.4	0.36	0.91
Excel 2 C_{18}-AR (ACE)	C_{18} linked to phenyl ring	2	100	3.0	150	15.5	0.4	0.74	1.86
Excel3 CN-ES (ACE)	Extended Spacer alkyl chain between the silica surface and CN group	3	100	2.1	150	12.6	0.4	0.36	0.91
Triart PFP (YMC)	C_{18} linked to pentaflourophenyl ring	3	120	2.0	150	15	0.25	0.33	0.82
Excel 2 C_{18}-PFP (ACE)	C_{18} linked to pentafluorophenyl ring	3	100	3.0	150	14.3	0.4	0.74	1.86
Kinetex™ HILIC (Phenomenex)	Unbonded core-shell silica, phase details not specified by manufacturer	2.6	100	2.1	100	NR	0.3	0.25	0.81
SeQuant® ZiC® HILIC	Zwitterionic stationary phase which has hydrophilic partitioning and electrostatic interactions	5	100	2.1	150	NR	0.3	0.36	1.21
Acquity UPLC HSS PFP	C_{18} linked to pentafluorophenyl ring,	1.8	100	1	150	7 %	0.03	0.08	2.75

*Void Volume (V_m) was estimated using 70 % of cylinder volume formula: $V_m=0.7 \times \pi \times r^2 \times L$ *(r = internal diameter/2, L=column length)* (49–51).
** T_0 was estimated by T_0 column=V_m/flow rate
NR = Not reported

2.2.2.6 Method development instrumentation

LC-MS instruments used during method development

Several instruments were used in the method development stage including:

- Thermo Scientific TSQ Vantage Tandem Mass Spectrometer with TLX-1 liquid chromatography system (Thermo Fisher Scientific, Waltham, Massachusetts, USA)
- Agilent 6490 Tandem Mass Spectrometer with binary pump 1290 liquid chromatography system (Agilent, Santa Clara, California, USA)
- a Waters™ Xevo TQ-XS and a Xevo TQ-S with Waters™ M Class UPLC (Waters™, Wilmslow, UK).
- Sciex 4000 Q-Trap® triple quadrupole (A B Sciex, Warrington, UK) was used in the ionisation study

The mass spectrometers that were selected to for the assessment in the method development stage of the METs analysis work were selected due several reasons including: a) the availability for use; b) having triple quadrupole mass analysers for improved selectivity for quantitative analysis and, c) better sensitivity, details about why certain instrument models from the list above were found to be more sensitive are discussed in the results and discussion.

Other instruments

Mettler Toledo XP205 and Mettler Toledo micro balance XP6 were used. Table 2.9 lists the measurement uncertainties of the balances. Weighing steps were all aided by the U ionizer anti-static device by Mettler Toledo (Mettler Toledo, Leicester, UK). Eppendorf electronic pipette E3x single channel (Eppendorf, Hamburg, Germany) was used for most of the preparation work.

Table 2.9 Balances measurement uncertainties and details

Manufacturer	Model	Serial number	Standard uncertainty (g)	Maximum capacity (g)
Mettler	XP205	1127072041	0.0001595	250
Mettler	XP6	B512740561	0.0000247	6
Ohaus	AP250D	N08717	0.00017	200

2.2.3 Sample preparation

2.2.3.1 Gravimetric preparation procedure

Although most of the preparation of samples and standards in solvent at the method development stage was volumetric, the final assessment of the method prior to commencing samples analysis and method evaluation work was performed gravimetrically. The weighing process was performed using high accuracy Mettler Toledo XP205 balance mentioned above which were all UKAS calibrated and the balance calibration checked by weighing a series of reference standard weight ranging from 100 mg - 200 g prior of using the balances. The weighing was aided by an anti-static device through which the vial, in which the sample was being prepared, passes before each weighing

to remove static charges. Additionally, all elements that were weighed including vials, caps, samples, standards in solvents were left at room temperature for at least 2-3 hours prior to the weighing process to ensure equilibration to room temperature.

Moreover, each weighing step was repeated at least three times, if a trend in the measured masses occurred, the weighing continues and if the trend persisted, it was investigated and resolved before continuing. Ultimately, each mass was weighed repeatedly until three or more measurements were scattered (not trending) and were within ± 0.0001595 g (159.5 µg) of each other when using the XP205 balance and were within ± 0.0000247 g (24.7 µg) on the XP6 balance. Hence, whenever a mass is mentioned throughout this thesis, it is only an indicative of the nearest nominal mass of the additions rather than the actual mass of the addition that was used for the measurement.

2.2.3.2 Blends preparation

For sample blends, ~0.5 g of plasma sample was taken to which ~0.150 g of SILIS was added in 2 mL Eppendorf tubes and vortex mixed. Subsequently, the blend was diluted by 0.5 mL 50 mM ammonium acetate buffer to pH of about 7. The blend was then vortex mixed and centrifuged for 40 min at 18000 g using an Eppendorf 5810 R centrifuge at 4 °C.

2.2.3.3 Solid-phase extraction

The basis of the method (SPE plate and wash steps) was adapted and optimised from a method by Waters™ (52) where here the sample and solvent

volumes and a second SPE step were all optimised for this method. The SPE 96-well plate used was Oasis WCX 10 mg sorbent per well, 30 µm particle size. The SPE manifold used was a manual manifold that used negative pressure for elution. Each well was conditioned with 0.5 mL methanol, followed by 0.5 mL 50 mM ammonium acetate buffer. The supernatant of the blends was then loaded on the SPE plate, upon complete dryness the SPE wells were washed with 1 mL of 50 mM ammonium acetate buffer followed by 1 mL of 100 % methanol. Upon complete dryness of the well, the target compounds were eluted using 250 µL of methanol with 5 % formic acid twice. The extracts were transferred to high recovery autosampler vials to be dried down using TurboVap® evaporation system at 30 °C water batch under a gentle stream of nitrogen for 10 min. The Turbovap® system used was supplied by Biotage (Uppsala, Sweden). The extracts were reconstituted using 100 µL of the starting mobile phase conditions of 98:2 A:B.

2.2.3.4 Solid-phase extraction after further optimisation

After the initial assessment of the method's performance a constant shift in the retention time of the peaks occurred in long analytical runs hence a further sample clean-up was required. A two-step SPE was performed, where samples were prepared using a WCX SPE plate as described in 2.2.3.3 and 2.2.3.4. however, instead of evaporating the extracts and reconstituting them, a second SPE step was performed. The WCX extracts (0.5 mL methanol with 5 % formic acid) were diluted with 1.7 mL water to achieve <20 % organic solvent to enable retention on the C_{18} SPE cartridges. This second SPE step

used Strata C_{18} SPE cartridges (100 mg sorbent mass, 3 mL volume) from Phenomenex (Macclesfield, UK). The SPE cartridges were conditioned with 1 mL methanol, followed by 1 mL water. Then the diluted extracts were loaded on the cartridges. The eluent was collected in 15 mL falcon tubes. These were placed in the TurboVap™ for the evaporation and reconstitution step. When the extracts were ~ 1 mL they were transferred to the autosampler vials in which the rest of the evaporation and reconstitution in 100 µL of water took place to minimise losses upon transfer.

2.2.4 Measurement Equation and Uncertainty Calculation

2.2.4.1 The simplified double exact-matched IDMS equation

The calculated amount of METs in each of the sample blends was determined using the simplified double exact-matched IDMS (DEM-IDMS) equation (53,54).

Equation 2.1 Simplified double exact-matched IDMS equation

$$w_x = w_z \cdot \frac{m_z}{m_{yc}} \cdot \frac{m_y}{m_x} \cdot \frac{R'_B}{R'_{BC}}$$

Where:

w_x mass fraction of METs in the sample (pg/g)

w_z mass fraction of METs standard used to prepare the calibration blend (pg/g)

m_x mass of the sample used (g)

m_y mass of the isotopically-labelled METs standard added to the sample blend (g)

m_z mass of the METs standard added to the calibration blend (g)

m_{yc} mass of the isotopically-labelled METs standard added to the calibration blend (g)

R'_B measured ratio of natural peak area to the isotopically-labelled peak area in the sample blend

R'_{BC} mean measured ratio of the natural peak area to the isotopically-labelled peak area in the calibration blend injected before and after the sample

The uncertainty associated with the measured mass fraction was calculated by combining the relative standard uncertainties of the ratio measurements, the weights of the blend addition preparation, and the mass fraction of the compound in solution, as shown in Equation 2.2 (53,54).

2.2.4.2 The double exact-matched IDMS uncertainty equation

Equation 2.2 Exact-matching IDMS Uncertainty Calculation Equation

$$u(w_x) = w_x \sqrt{\left(\frac{u(w_z)}{w_z}\right)^2 + \left(\frac{u(m_x)}{m_x}\right)^2 + \left(\frac{u(m_y)}{m_y}\right)^2 + \left(\frac{u(m_z)}{m_z}\right)^2 + \left(\frac{u(m_{yc})}{m_{yc}}\right)^2 + \left(\frac{u(R'_B / R'_{BC})}{R'_B / R'_{BC}}\right)^2}$$

Where:

w_x mass fraction of METs in the sample (pg/g)

$u(w_x)$ standard uncertainty estimate of the mass fraction of METs in the sample blend (pg/g)

w_z mass fraction of METs standard used to prepare the calibration blend (pg/g)

$u(w_z)$ standard uncertainty of the mass fraction of METs in the calibration blend (pg/g)

m_x	mass of the sample used (g)
$u(m_x)$	standard uncertainty of the mass of the sample used (g)
m_y	mass of the labelled METs standard added to the sample blend (g)
$u(m_y)$	standard uncertainty of the mass of the labelled METs standard added to the sample blend (g)
m_z	mass of the METs standard added to the calibration blend (g)
$u(m_z)$	standard uncertainty of the mass of the METs standard added to the calibration blend (g)
m_{yc}	mass of the labelled METs standard added to the calibration blend (g)
$u(m_{yc})$	standard uncertainty of the mass of the labelled METs standard added to the calibration blend
R'_B/R'_{BC}	mean ratio of the measured ratio of natural peak area to the labelled peak area in the sample blend and in the bracketing calibration blend (n=5)
$u(R'_B)/R'_{BC})$	standard deviation of R'_B/R'_{BC}

2.2.4.3 Calculation of Combined Uncertainty

The combined uncertainty for a sample set was calculated using Equation 2.3 (53,54).

Equation 2.3 Combined standard uncertainty

$$u_{combined}(w_X) = \sqrt{\overline{u(w'_X)}^2 + b_{\mathrm{var}}{}^2}$$

Where:

$\overline{u(w'_x)}$	Mean of the standard uncertainty of the mass fraction w'_x of analyte in the sample blend

b_{var} Blend to blend variation (the standard deviation mean) which is calculated by taking the standard deviation of the mass fractions of a sample set and dividing by the square root of the total number of samples combined

$u_{combined}(w_x)$ Combined standard uncertainty

Standard uncertainties were converted to expanded uncertainties (95 % confidence interval) using a coverage factor k = 2 when enough degrees of freedom are available. See 01.9.3 for coverage factor definition.

2.3 Results & Discussion

The aim of the work from the outset was to understand the sources of variability in higher-order reference measurements by using the currently available modern technologies to achieve accurate quantification of METs in plasma by developing a higher-order RMP. The question of how low the measurement uncertainty of the RMP can be is based on the available technologies. Therefore, during this work the optimisation of all aspects of the method was crucial to achieve the lowest possible uncertainty by capitalising on the available resources to identify way to improve the measurement.

Most of the initial method development work was performed using standards in solvent at two different mass fractions 1 ng/g and 100 pg/g. The former was used for initial chromatography separation evaluation and the latter was used for quantification work and estimation of measurement uncertainty. With the absence of traceable reference cut-off limits, the decision was made to select a working mass fraction of 100 pg/g as a starting point after which plasma material was used to further develop the method.

2.3.1 Ionisation of METs

METs are very small compounds and are susceptible to in-source dissociation, this has been reported in the literature consistently (28,40,41,44–48). If such dissociation is not optimised and controlled, it could cause variability in the measurement due to the signal instability of the precursor ions. Because the aim of the work is to achieve high accuracy quantification, an initial aim to achieve a method with high specificity by using an adduct of the intact ion instead of an adduct of an in-source fragment ion was pursued.

An investigation of different MS instrumentation was performed. Each of the MS instruments from different vendors had a different ESI source design. Figure 2.4 shows the schematic drawings of the designs of four of the instruments' ESI sources that were evaluated. An infusion of the METs standards in solvent was performed on each of the instruments while optimising the different conditions of each source, see Figure 2.4. Full scans were captured over a range of *m/z* 50 - 250 to observe the different ions forming in the source, as adducts, protonated adducts or in-source fragments. The data in Figure 2.5 shows a comparison of four of the instruments that were evaluated by comparing the relative abundance of the three compounds' protonated adducts $[M+H]^+$ vs. the in-source fragments.

Firstly, it was found that the compounds behaved differently across ESI sources. However, no adducts of higher *m/z* (e.g., sodium or potassium adducts) than that of the molecular mass of the compounds were observed in any of the scans. This indicated that the compounds did not bind with alkali

metal ions in the environment such as sodium or potassium to form adducts of higher *m/z.* Evidently, it was not possible to eliminate the in-source fragment ion formation in any of the instruments evaluated even when the source temperatures were dropped as low as 150 °C. In the case of Normet the protonated adduct was not observed at all in the Agilent Jet Stream source and the Thermo Fisher HESI II source, only the in-source fragment was seen.

Overall, across all sources that were evaluated the conditions that were found most influential on the in-source fragmentation were: a) the temperature of the source main capillary needle/nebuliser needle b) the temperature and flow of the sources' gases. It was found that the sources where the nebuliser needles have higher temperatures, the in-source fragmentation was most pronounced, such as with the Agilent Jet Stream and the Thermo Fisher HESI II sources, see Figure 2.4. The source that was deemed to have the softest in-source fragmentation was the Waters™ Z-spray source however the fragmentation was minimised but not eliminated. Although the protonated adduct would be ideal for increased method specificity and reproducibility, but it was not achievable the exception to the Waters™ Z-Spray. This ESI source was not selected because the MS it was on had lower sensitivity. Further discussion of the selection of the MS instruments and their sensitivity evaluation is to follow.

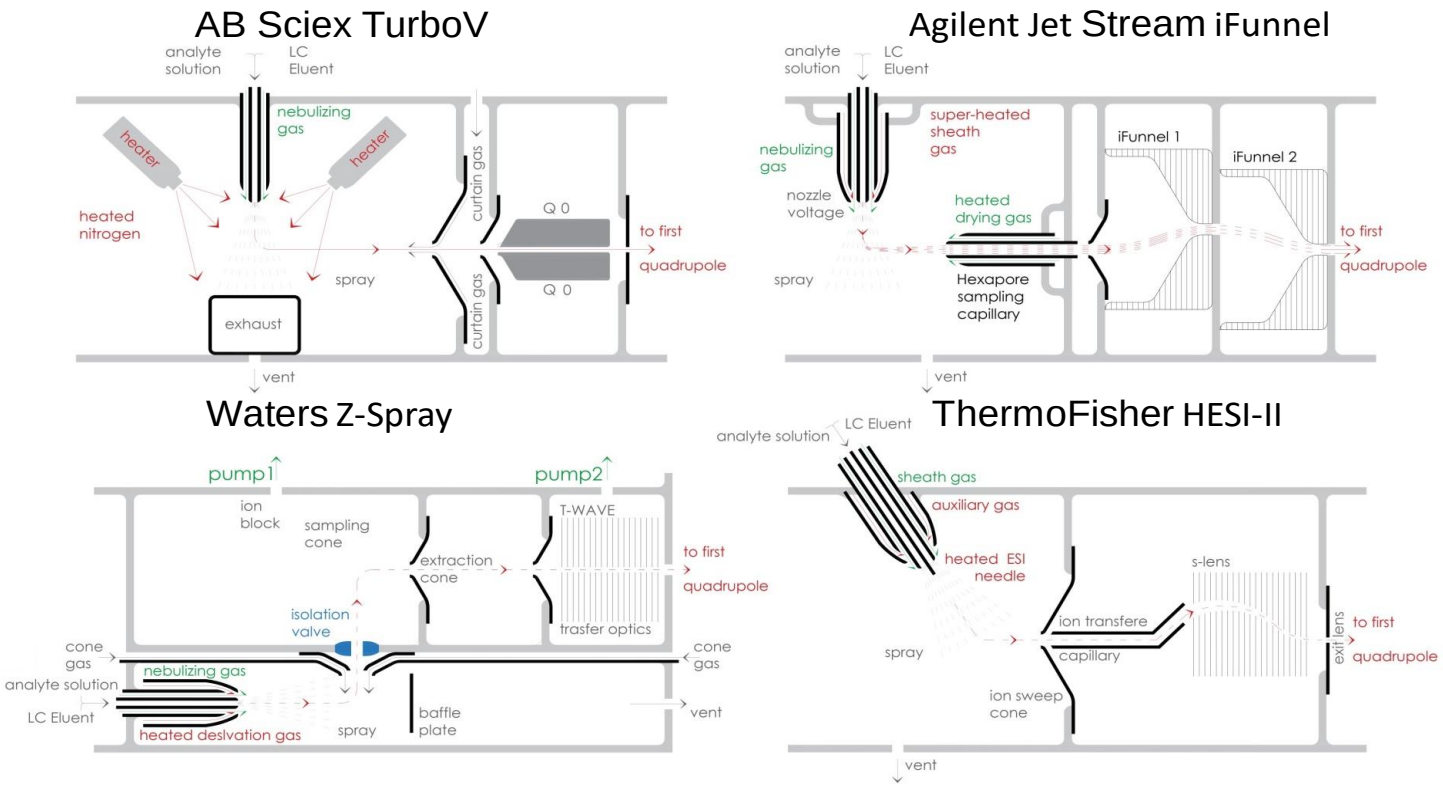


Figure 2.4 Schematics of the ESI sources and the ion optics compared in the evaluation. Schematics redrawn from vendors' catalogues (55–58).

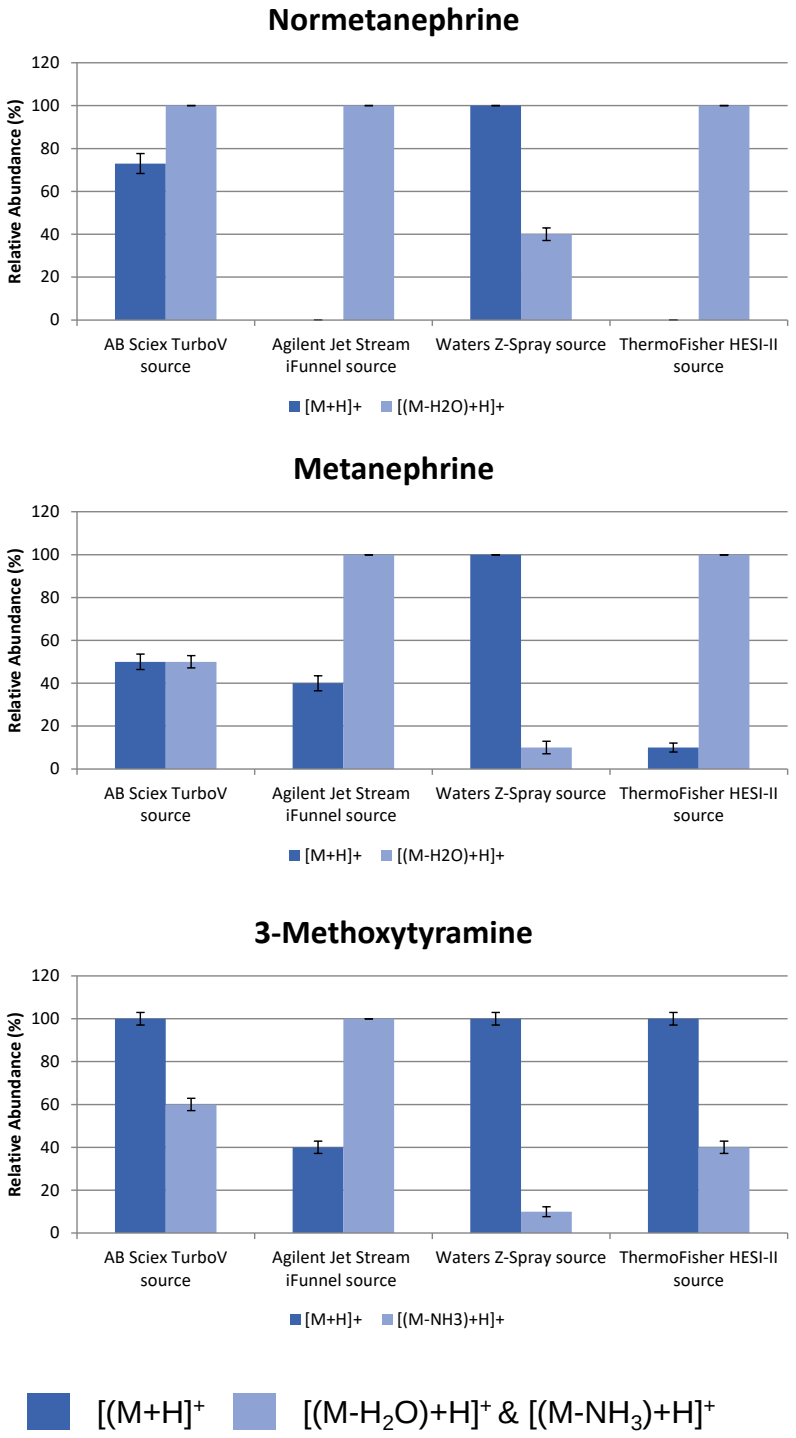


Figure 2.5 Comparison of the ionisation of the compounds; protonated molecule vs in-source dissociation ion for each of the METs by tuning of each of the sources. Error bars = standard deviation of 3 infusion experiments, each of 1 min acquisition.

Given the in-source fragment ions were the same ions formed across the different instruments and were not much smaller than the protonated adducts, indicating a loss of H_2O or NH_3, they were deemed to be robust and to have high assurance specificity wise. Because in-source fragmentation could impact the sensitivity as the compound signal would be distributed over several ions, it was decided to optimise the ESI source conditions to maximise the in-source fragment as a precursor ion for the SRM set up. The same in-source fragment precursor ions were observed in both the infusion experiments and the LC experiments. The structures of the in-source fragments were expected to be as shown in Figure 2.6, which were consistent with the anticipated structures by Eisenhofer *et al* (59). Optimising the ESI source conditions to maximise the in-source fragment in this case was crucial to reduce variability in the signal. This is important not only from a reproducibility aspect but additionally because it provides control over the method and eliminates a potential source of variability and uncertainty in the measurement.

Upon deciding the precursor ion masses, the SRMs of each of the compounds were optimised. Two SRMs were optimised for each compound. The SRM with the higher signal was used for quantification and the second SRM was used for confirmation. Table 2.7 includes the optimised SRM transitions.

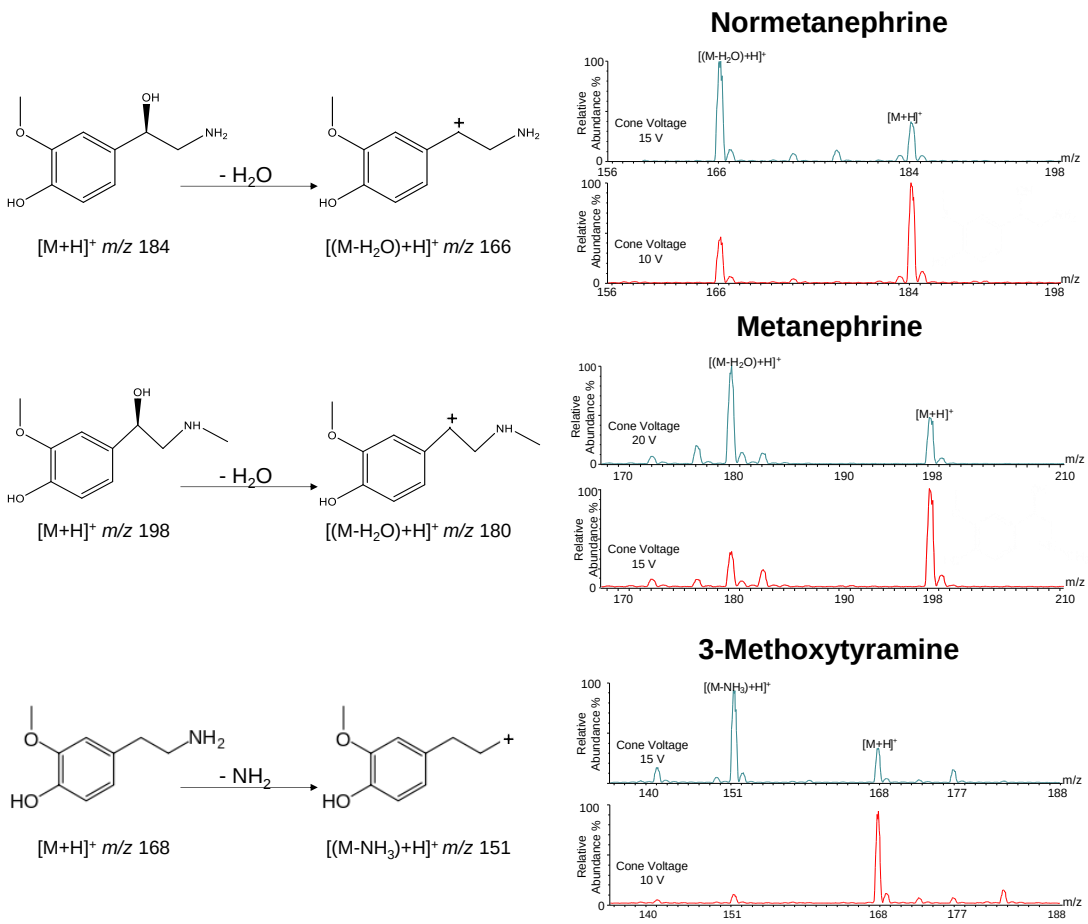


Figure 2.6 Example full scan mass spectra generated using the Z-Spray source on the Waters™ Premiere XE with the suspected structures of the ions generated by in-source dissociation.

2.3.2 LC columns selection

2.3.2.1 RP-LC columns screening

METs are small highly polar compounds (see structures in Figure 2.3) that are not highly retentive on the commonly used RP-LC stationary phases. Hence, reported methods are generally split between the use of HILIC and RP-LC, see Table 2.2. To select the chromatography for the development of the higher-order RMP that would achieve the highly accurate measurements required, both RP-LC and HILIC columns were tested.

Figure 2.7.E is the chromatogram with the best separation achieved by the Excel 2 C_{18}-PFP among the RP-LC columns tested as it provided the highest retention and complete separation of the compounds' peaks. A column screening experiment of RP-LC columns was performed to evaluate the retention and the chromatographic separation. Over eleven columns were evaluated, chromatograms of six of these columns are shown in Figure 2.7. The evaluation was performed using METs standards in solvent at higher mass fraction of 1 ng/g using the Thermo Vantage TSQ LC-MS.

Better retention and separation was achieved by mixed mode LC columns, specifically C_{18}-PFP, see Figure 2.7.D and Figure 2.7.E. C_{18}-PFP columns provided a dual retention mechanism where the benzene rings in the structures of METs bind to the benzene ring in the PFP function by pi-pi interactions while the chains in the structures would exhibit Van der Waals interactions bonding with the C_{18} chains.

The Excel 2 C_{18}-PFP (Figure 2.7.E) provided a better chromatographic separation and better peak shape that the Triart PFP (Figure 2.7.D). While the carbon loading of both columns was similar, the pore sizes were different, 100 Å and 120 Å, respectively. This smaller pore size of the Excel 2 column gave it a larger surface area which provided the better retention and separation. This was assessed across the columns by comparing the retention time of the first eluting peak with t_0 and the time between the retention times among the three peaks. However, column characterisation sometimes would not be as accurate as labelled by manufacturer to know what precisely the reason behind the better performance was.

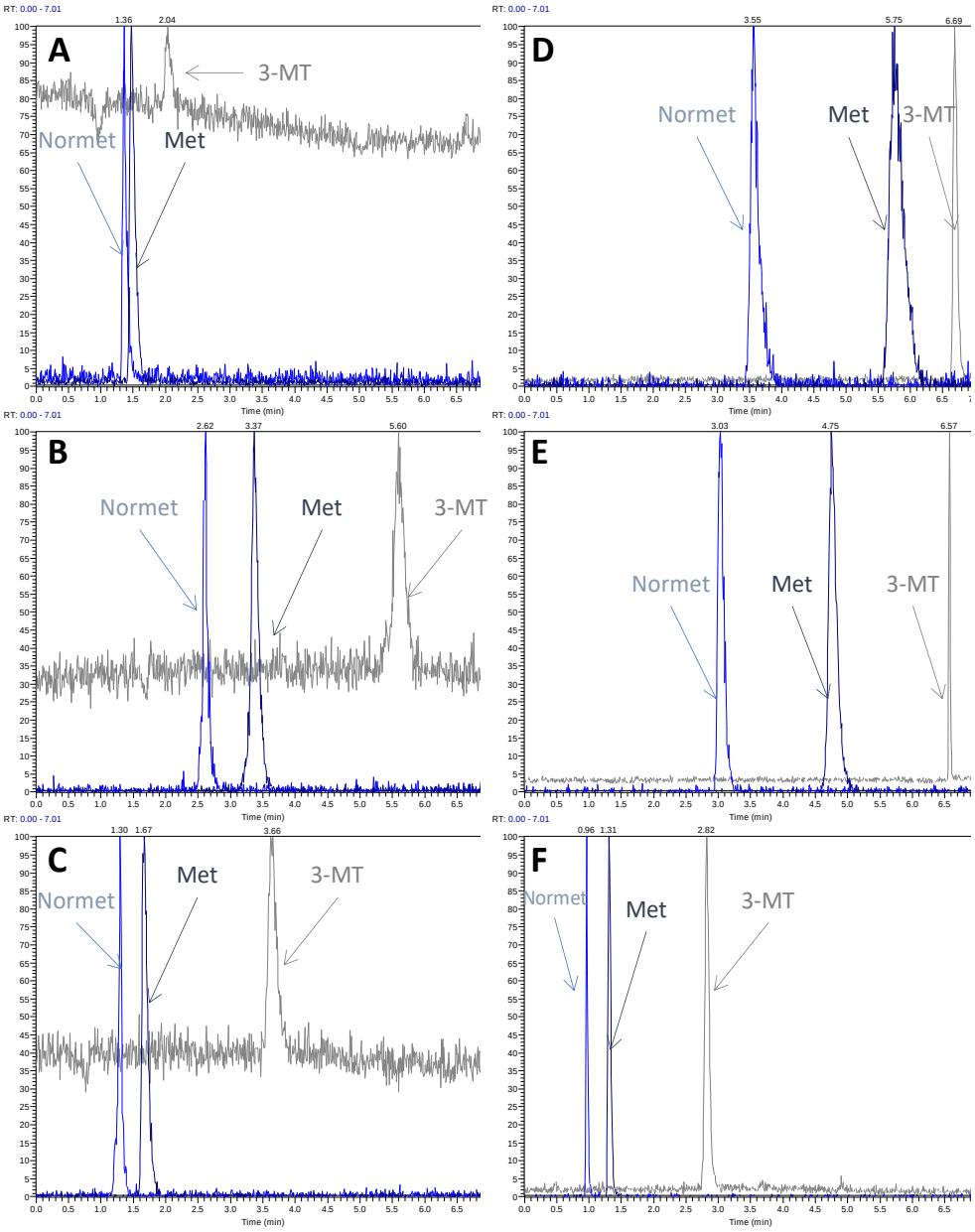


Figure 2.7 Separation of METs using 6 different RP-LC columns: A) Excel3 CN-ES B) Excel 2 C_{18}-AR C) Excel 3 C_{18}-HL D) Triart PFP E) Excel 2 C_{18}-PFP F) Acquity UPLC HSS T3.

2.3.2.2 HILIC columns assessment

On the other hand, due to the polarity of the METs HILIC was the other chromatographic mode to be evaluated. An assessment was performed using two different HILIC columns: Sequant® Zic®-HILIC and Kinetex HILIC columns (columns details are in Table 2.8). The Kinetex HILIC columns failed to achieve baseline separation whereas the Sequant® Zic®-HILIC column achieved a partial baseline separation (see Figure 2.8.B). However, Figure 2.8 shows the superior chromatographic separation achieved by the ACE C_{18}-PFP column (Figure 2.8.A) compared to the Sequant® Zic®-HILIC column under the experiment conditions. HILIC remained an option for the chromatography despite the poorer chromatographic separation under the conditions tested because the target analytes were separated spectrally using tandem MS (MS/MS). The HILIC and RP-LC chromatographic methods that were successful in separating the compounds were further investigated after the MS instrument was selected.

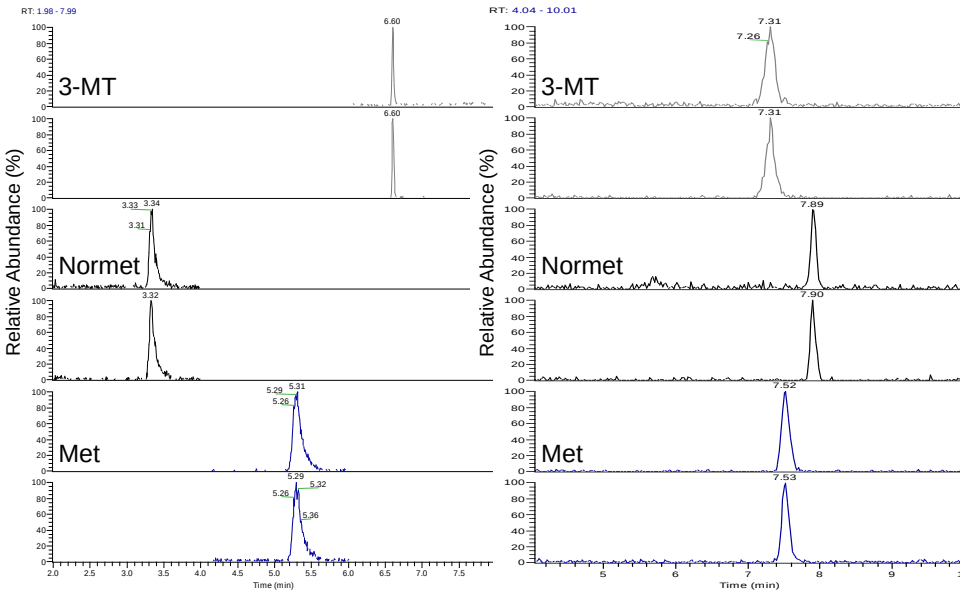


Figure 2.8 Separation of METs using two different chromatography approaches RP-LC and HILIC A) ACE C_{18}-PFP (3.0 x 150 mm, 2 µm) and B) SeQuant ZIC-HILIC (2.1 x 150 mm, 5 µm) includes sulfobetaine functional group.

2.3.3 HILIC vs RP-LC

The HILIC method was assessed and found to be not robust under the method conditions tested. The retention times were shifting from one week to another, jeopardising the precision of the chromatography. The retention times shifted by 0.5 minute additionally, the peak areas were not consistent for the same mass fraction standard. Moreover, when the method was tested using plasma extracts it was found to be not reproducible. Figure 2.9 shows a comparison of chromatograms of plasma extracts analysed on different weeks, the peaks (definition and retention time) were not reproducible from batch to batch despite maintaining the same column temperature and all LC conditions. For example, Met eluted at 7.97 min in the first batch but eluted a minute earlier (7.08 min) in the second batch. This could be due to a contamination, insufficient equilibration of the HILIC column, minor mismatch between the 100 µL extract make-up and the starting mobile phase solvents which were

10:90 10 mM ammonium acetate:acetonitrile or variability in mobile phase solvents preparation. This indicated that the HILIC method suffered from poor method robustness which is not suitable for a higher-order RMP. Additionally, upon examining the Joint Committee for Traceability in Laboratory Medicine and Clinical Chemistry (JCTLM) database LC-MS methods of small molecules, no higher-order RMP on the database used HILIC (60). Finally, the retention factor on the RPLC-column was higher than on the HILIC column.

Given the aim of the work was to achieve high accuracy quantification, method robustness and reproducibility were key requirements of the method. Hence, the decision was made to select RP-LC based chromatography instead of HILIC to ensure chromatography robustness and precision. The method selected is described in section 2.2.2.1and is RP-LC based using 98 % aqueous starting mobile phase conditions to achieve initial retention of the compounds on the column stationary phase. RP-LC is desirable to the use of HILIC not only because of its reproducibility but additionally due to the use of water and organic solvent as mobile phase solvents instead of the use of buffers in the case of HILIC which could be prone to reproducibility issues in the preparation procedure (29).

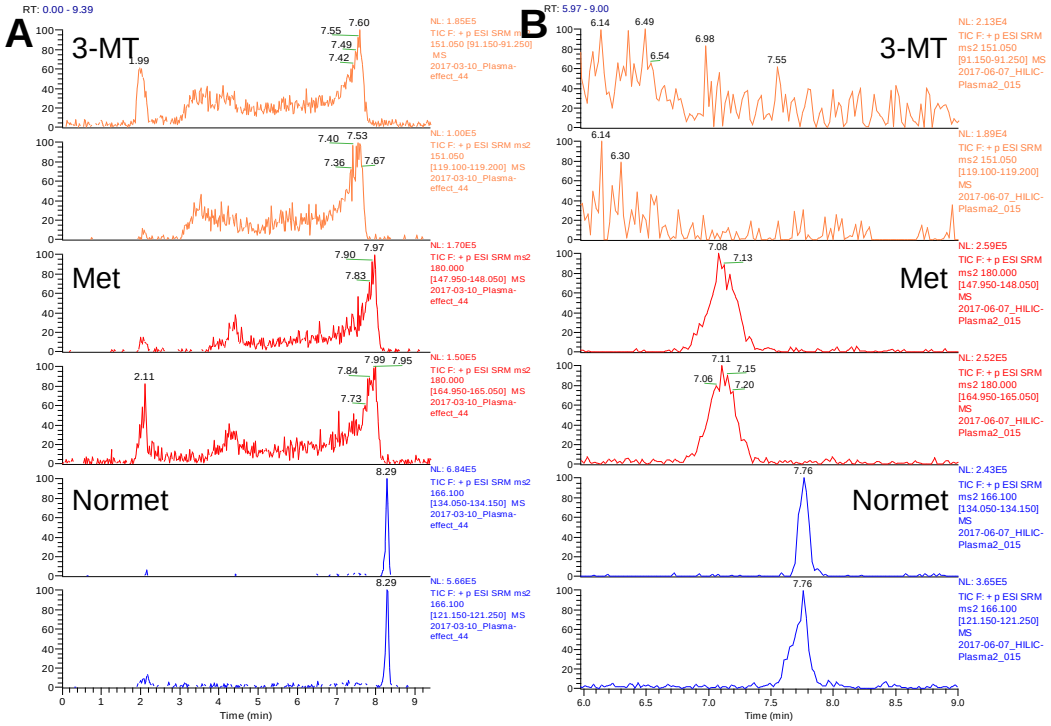


Figure 2.9 HILIC method robustness, the chromatograms were generated using the same HILIC method on two different weeks A) in March 2017 and B) in June 2017.

2.3.4 MS instruments sensitivity comparison

To obtain high accuracy quantification with low measurement uncertainty, the signal of the analytes in the detector needs to be the highest possible within the dynamic range of the instrument to avoid signal saturation. The signal sensitivity could differ between different MS instruments according to instrument design, condition, age and capabilities. Therefore, an evaluation of different triple quadrupole MS instruments was performed using the same LC method and same METs standards in solvent. The evaluative tool to assess the instrument sensitivity for this study was the estimated signal-to-noise ratio peak to peak ($S:N_{p/p}$) of each compound's quantitative peak. Figure 2.10.A shows the estimated $S:N_{p/p}$ achieved with each instrument using a METs standard in solvent at 100 pg/g. The higher $S:N_{p/p}$ the better the instrument

sensitivity which could provide a lower measurement uncertainty. The analysis of the standards in solvent on each of the instruments was repeated five times to assess variability. Although $S:N_{p/p}$ as an assessment tool is not highly accurate and precise but, for the purposes of comparing different platforms it was found sufficient to assess differences in sensitivity. Figure 2.10.B shows the variability of the five replicate analysis that could be attributed to the instrument sensitivity and/or the variability in using $S:N_{p/p}$ as a tool for assessment.

The triple quadrupole MS instruments evaluation study found the Xevo-TQXS to be the most sensitive with the lowest %RSD of all the instrument tested. The Xevo-TQXS was the most modern instrument on the market at the time and had the newest and most sensitive electron multiplier detectors of all the other instruments. The upgrade in the instrument manufacturing, according to the vendor, uses the StepWave XS™ ion guide technology which allegedly provides increased sensitivity as well as an enhanced detection system with six orders of linear dynamic range to better improve sensitivity (61). However, the standard deviation of the $S:N_{p/p}$ values of the five replicate analysis on the Xevo TQ-XS (see error bars in Figure 2.10.B) was higher than both the Xevo-TQS and the Agilent 6490. Such high variability is to be considered with some scepticism, as indicated above because it could be due to the instrument or the assessment of the $S:N_{p/p}$ and the difference in noise regions between analysis. Unfortunately, the instrument was evaluated at the Demo Laboratory at Waters™ facilities in Wilmslow and was not available to perform the research it hence, no further experiments were performed using the Xevo TQ-

XS. The Waters™ Xevo-TQS and the Agilent 6490 both demonstrated similar level of sensitivity, with no significant difference between the two instruments for both Normet and Met ($p > 0.05$) while for 3MT the $S:N_{p/p}$ between the two instruments was significantly different for 3MT ($p < 0.05$). The final choice of instrument was the Agilent 6490 due to its better availability for use. More work on increasing instrument sensitivity to improve the measurement quality and drive the measurement uncertainty as low as possible is described later.

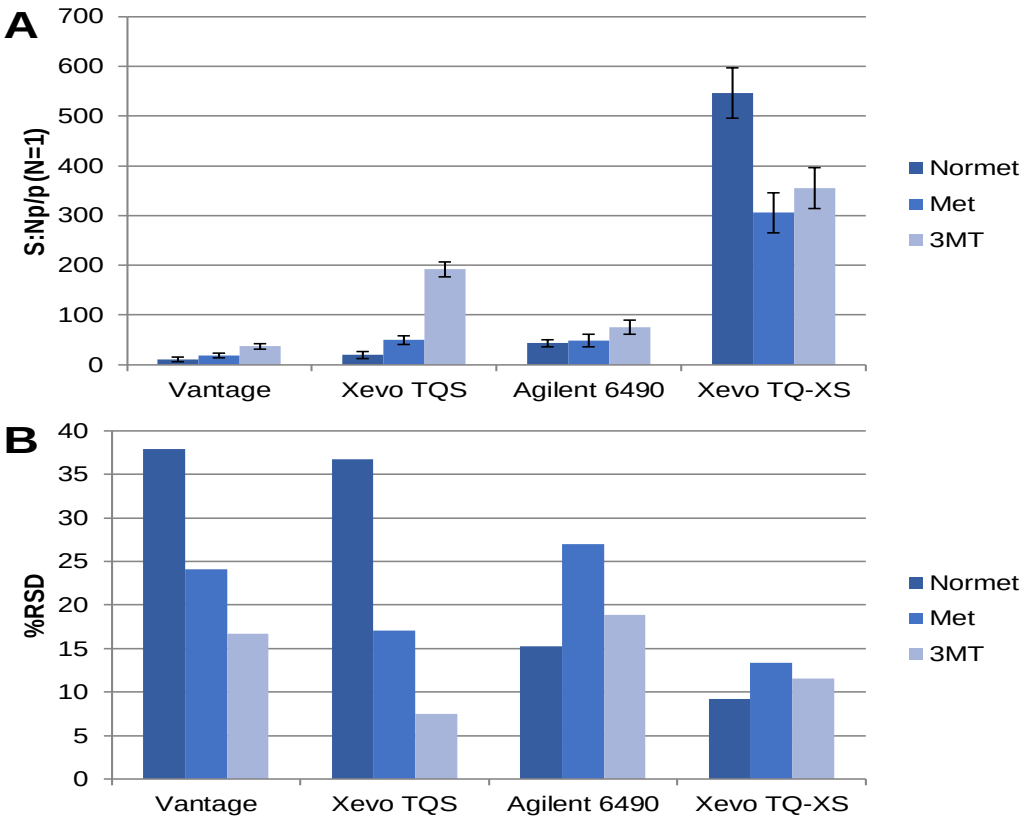


Figure 2.10 Comparison of four different triple quadrupole mass spectrometers using the same RP-LC-MS/MS method and same METs standards in solvent. A) Shows the signal-to-noise peak to peak ratio ($S:N_{p/p}$) where the noise is normalised to 1, error bars = standard deviation of 5 injections. B) demonstrates the relative standard deviation of the $S:N_{p/p}$ of the five injections.

2.3.5 Micro-flow liquid chromatography

Low measurement uncertainty and traceability are the key aspects of higher-order reference measurements. While traceability is attainable using the correct reference standards and preparation procedures, the measurement uncertainty estimate remains very much dependent on the instrumental analysis. The main component of the measurement uncertainty budget is the ratio precision of the instrument. The term ratio precision refers to the relative standard deviation (%RSD) of the ratio of ratios in DEM-IDMS calibration. The ratio of ratios is the ratio of the peak areas of the compound to its SILIS in the sample blend to that of the calibration blend, described in 1.8.4. The variability in instrument measurements was reflected in variability in the ratios hence it was evaluated by ratio precision i.e., %RSD of the ratios of repeat measurements, which is typically the driving component of the measurement uncertainty.

Better ratio precision is achieved by improved sensitivity (within the dynamic range of the instrument), therefore micro-flow liquid chromatography was assessed to improve overall detection sensitivity. Improved instrument sensitivity would be anticipated with micro-flow LC due to the reduction in the LC flow rate going through the ESI source, and such flows can be as low as 10 µL/min as opposed to 400 µL/min in standard LC. At such lower flow rates the ionisation efficiency of the ESI source would be improved as the dissolution of the mobile phase solvent would be more efficient resulting in higher ionisation hence more of the compound is ionised and passed through to the mass spectrometer (62,63).

Improved sensitivity with the use of the micro-flow LC was reported in numerous studies, as an example it enabled the measurement of vitamin D metabolites which are also measured at trace levels (pg/mL) (64). Similar increase in sensitivity in measurement of pesticides was achieved using micro-flow LC (65). Both the vitamin D study and the pesticides analysis study used a column with internal diameter (i.d.) 0.5 mm and injected 9.5 µL and 3 µL, respectively.

When developing the micro-flow LC method two narrower i.d. columns were tested 1.0 mm and 0.3 mm. The 1.0 mm i.d. column provided the best compromise between high injection volume and sensitivity without increasing signal suppression. In agreement with the work of Quigley *et al.* (66) the most critical factors in the optimisation of the micro-flow LC method included the flow rate, internal diameter and injection volume. The optimisation of these factors was critical to achieving improved signal and for the capitalisation on the mass spectrometer capabilities.

Uniquely, the electric current on the nebuliser needle on the Agilent 6490 Jet Stream ESI source is grounded by design, which allowed direct connection of the LC column onto the ESI source, see Figure 2.11.B. This minimised dead volume between the column end and the ESI source which reduced any potential band broadening and lowered peak heights i.e., lower sensitivity. Figure 2.11 shows pictures of the unique set up of the micro-flow LC system. The set up used narrow pre-cut polyether ether ketone with fused silica lined (PEEKsil™) tubing to minimise dead volume. The system had a micro-autosampler with an injection loop up to 8 µL that is designed to minimise

dead volume at the autosampler. The binary LC pump Agilent 1290 is designed to run the LC program and perform the mixing of solvents at higher flow rates then splitting the flow to the micro-flow rate via a micro-flow sensor, this ensures accurate and robust mixing of mobile phase solvents in addition to robustness of the low flow rate delivered. A micro-filter was set up inline before the column to reduce the chances of blockages. Finally, the column temperature was maintained at 30 °C using an external hot pocket by Thermo Fisher because the column was mounted directly to the ESI source and therefore could not be placed in the LC column oven.

Figure 2.11 micro-flow LC-MS/MS set up; A) the connection of the micro-autosampler to the column was optimised to minimise dead volume by using PEEKsil tubing with internal diameter of 25 µm. An inline micro-filter was also used. B) A close-up photo of the unique column connection set up directly connected to the ESI course using a coupler. C) An overview over the instrument set up showing the column temperature control using a hot pocket set to 30 °C.

The challenges with micro-flow LC are technical as unlike normal flow LC it is harder to detect leaks in the system with such low flow rates because they would not be visible and leak detectors often require larger volumes to alert user. Moreover, the chances of blockages are higher due to the use of micro scale system including tubing and fittings. For example, the use of autosampler vials with pre-slit caps proved essential to reduce blockages in the micro-autosampler valve. The back pressure was maintained at 270 bar throughout the sequence. An increase in back pressure was observed with the gradient mixing back to starting mobile phase condition but it was consistent to 300 bar to then drop back to 270 bar. The maximum back pressure for the micro-flow LC pump was 400 bar.

Once the set up was completed and demonstrated acceptable robustness a comparison between the LC methods was performed using the same MS, same samples, same column stationary phase particles and column dimensions. The only difference in the columns between the micro-flow LC method and the normal flow LC was the internal diameters of the columns which were 1.0 mm and 3.0 mm, respectively. The narrower column internal diameter reduces the band broadening in the column and would result in sharper peaks. Similarly, the same gradient program was used while considering the delay in the lower flow. The injection volume was optimised for both methods to be the highest attainable which was 40 µL and 8 µL for normal flow and micro-flow LC, respectively.

The injection volume optimisation was performed using both standards in solvents and plasma extracts. The increase in injection volume showed

significant increase in signal for all three compounds in standards solvent ($p < 0.05$). However, in plasma extracts only Normet and 3MT demonstrated significant increase in signal ($p < 0.05$). Overall, the increase in signal was higher in standards in solvents than in plasma extracts due to the matrix effects where the other matrix components could cause additional matrix suppression. Additionally, the increase in injection volume showed a small shift in the retention time but this shift was repeatable throughout the runs.

When analysing METs standards in solvent using micro-flow LC method higher sensitivity was observed compared to the normal flow LC method (see Figure 2.12). Assessing the increase in sensitivity between the two methods was performed by analysing a METs standard in solvent at 100 pg/g, five times using each method. The compounds peak areas were significantly bigger in micro-flow LC analysis than in normal flow LC with $p < 0.05$ across all three compounds (see Figure 2.13.A) and $S:N_{p/p}$ was also improved using micro-flow LC.

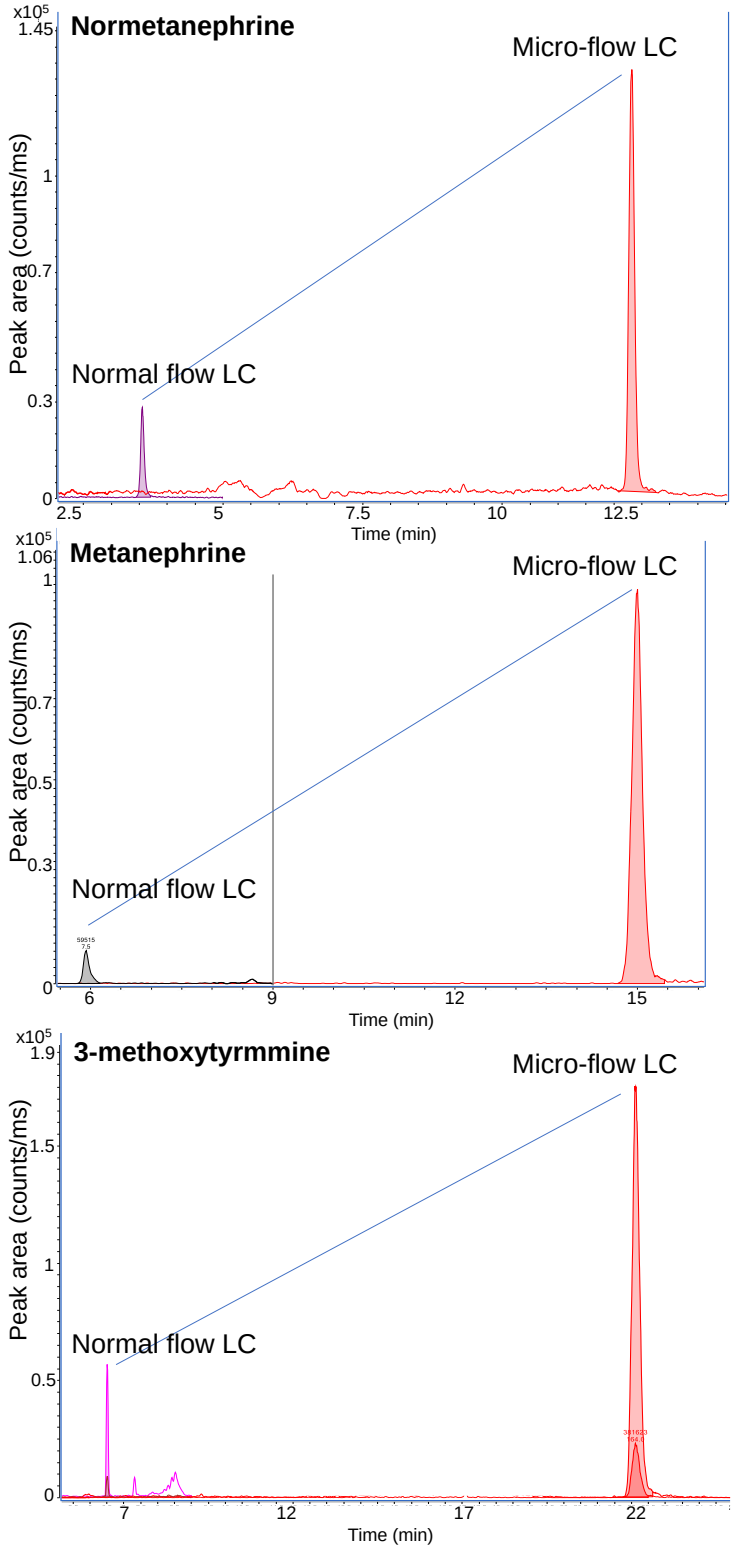


Figure 2.12 Comparison of normal flow LC (300 µL/min) vs micro-flow LC (30 µL/min); Peaks of the SRM transitions are overlaid, the left peak is the normal flow, and the right side (red) peak is the micro-flow LC.

One of the driving factors of the measurement uncertainty budget is the variability in the instrumental analysis, this is assessed by replicate analysis. This variability that can be evaluated by calculating the %RSD of the ratios of five replicate analysis, the smaller the %RSD the lower the measurement uncertainty would be. Therefore, a comparison of the %RSD of the micro-flow LC vs the normal flow LC method was performed. Figure 2.13.B demonstrates the reduced variability using the micro-flow LC method as its %RSD is significantly lower than the normal flow LC. This indicated the micro-flow LC-MS/MS method could provide the measurement uncertainty required for higher-order reference measurement and worth pursuing as a technique of choice for the method. Several experiments were therefore undertaken to pursue further gains in sensitivity.

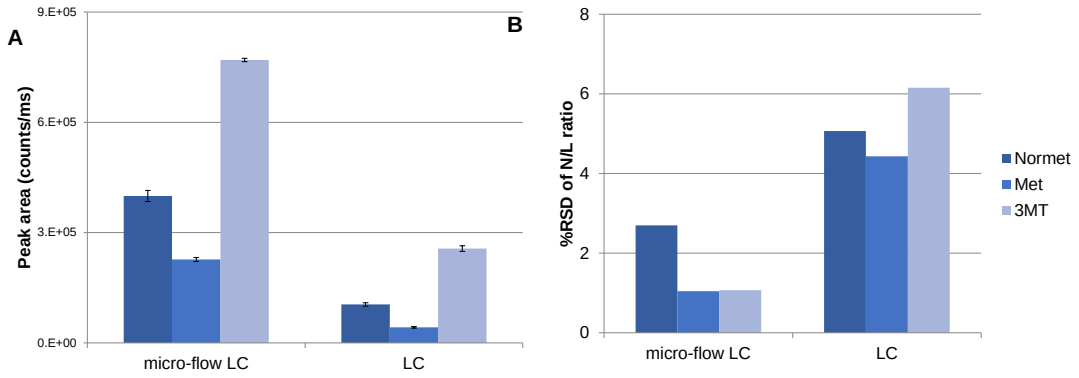


Figure 2.13 Comparison of micro-flow LC vs normal flow LC A) comparison between peak areas of the three compounds acquired using both methods, each peak area is an average of 5 replicate analysis, Error bars = standard deviation of 5 replicate analysis B) comparison of %RSD (n = 5) of the ratio of the compound to its isotopically-labelled internal standard.

2.3.6 Micro-flow LC-MS/MS method optimisation

The aim was to understand variability in measurement by developing the best performing method using available technology to tackle the challenges of the measurement of small polar compounds at trace levels in plasma. Additionally, when repeating the analysis using METs standards in solvent, it was observed that the %RSD of the compound to its SILIS ratio (N/L) was variable from one week to another. Because of these two reasons, a thorough optimisation of the method was performed to a) understand the sources of the variability in the method and b) increase the sensitivity of the micro-flow LC method to improve method's performance and reduce the variability in measurement i.e., to improve the %RSD of the N/L ratio, the driving factor of the uncertainty.

2.3.6.1 MS conditions optimisation

Different MS conditions were optimised to improve the signal including gas flow and temperatures, nozzle voltage, sheath gas flow and temperature. Compound specific conditions of the MS method such as collision energy of each SRM and dwell times were also further optimised.

The peak shape is determined by the data-points across the peak, each data-point represents an MS/MS scan. Given that the dwell time is the duration during which the signal collected for each scan, it was found to be a critical MS condition to optimise. Poorly optimised dwell time resulted in poorly characterised chromatographic peak, hence the variability increased. Furthermore, the optimisation of the MS dwell time proved to be the most influential on the measurement uncertainty estimate.

A variety of MS methods were set up. A group of methods were set up with continuous acquisition of all SRM transitions which would result in less scans per compound. A second group of methods had the SRM transitions time-segmented on three time segments each starts a minute before the peak retention time and ends a minute after it. With both groups of methods (continuous SRM table and segmented SRM tables) a variety of dwell times were tested. One of the segmented methods was the ‘dynamic’ dwell time setup is a software built-in setting where the software assigns a dwell time automatically and does not provide the value assigned. Each of these methods was repeated five times using the same sample. The five replicated were used to calculate the %RSD of the ratios of each method. These were then compared to evaluate the variability of each of the method and the impact on the measurement uncertainty of the methods.

Figure 2.14.A shows the %RSD of five replicate analysis using some of the methods that were evaluated. The optimum dwell time that provided lowest %RSD (subsequently lower measurement uncertainty) was the segmented method with ‘dynamic’ dwell time for Normet and Met and three windows with 75 ms dwell time for 3MT. Because reproducibility and robustness are key features of a higher-order reference measurement, full control over the method conditions is essential hence the second-best dwell time for Normet and Met (50 ms) was selected as the optimum, see Figure 2.14.A.

The scans across each of the peaks were counted manually using the MussHunter™ software function of “walk across the peak” (see Figure 2.14.B). This was done on three of the chromatograms to assess variability.

All the different methods had high number of scans/data-points across the peaks of all three compounds as shown in Figure 2.14.B. Normet and Met peaks had over 50 datapoints per peak at dwell time of 50 ms while 3MT which had the sharpest peak had over 30 datapoints at dwell time of 75 ms. This is much higher than the minimum of seven datapoints across the peak for identification and fifteen datapoints across the peak for quantification which are recommended by the European Commission decision 2002/657/EC (67) and adopted by Kruve *et al.* (68) as the clearest guidance on the subject. This high number of scans ensured accurate definition of the peak and it was concluded that the method dwell time setup was optimum for quantification at the lowest measurement uncertainty possible.

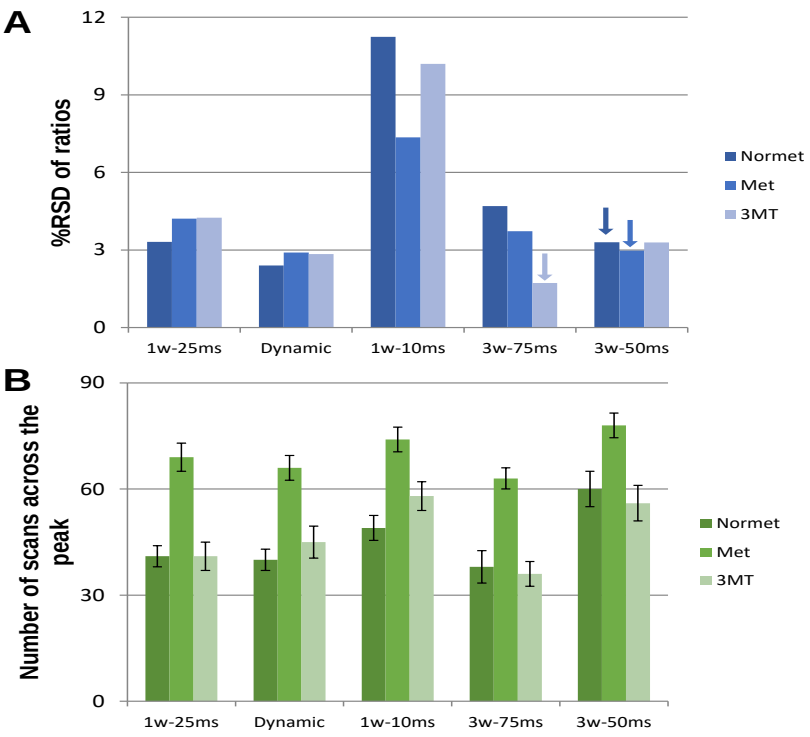


Figure 2.14 MS dwell time optimisation impact on measurement uncertainty estimate A) %RSD of five replicate (n = 5) analysis using methods with different dwell time settings; 1w-25 ms: is a continuous SRM table method using 25 ms for each transitions, Dynamic: is a software set up that assigns the dwell time automatically, 1w-10 ms: is : is a continuous SRM table method using 10 ms for each transitions, 3w-75 ms: is segmented SRM tables method using 75 ms dwell time for each SRM, 3w-50 ms: is segmented SRM tables method using 50 ms dwell time for each SRM. The arrows indicate the dwell times that were selected as the optimum set up B) the number of scans across the peak of every method, the scans were counted manually using the MS hunter walk through chromatogram function, three chromatograms for each method were counted and the average is plotted in the bars with error bars = standard deviation of three chromatograms count.

2.3.6.2 Post LC column dead volume reduction

Reduced dead and spray volume should give better chromatography and less dispersion and therefore more efficient ionisation due to reduced volume for desolvation. Therefore, additional attempts to reduce the dead volume even post the LC column were performed by optimising two main parts: a) the coupler sleeve i.d. from 250 µm to 25 µm (see Figure 2.15.A) and b) the

nebuliser needle i.d. from the standard needle to the micro-needle (see Figure 2.16.A). These optimisations were performed sequentially and resulted in increase of signal. When using the micro-nebuliser needle, it was found that the signal improved as the peak areas increased significantly for Normet and Met ($p < 0.05$) while the peak area of 3MT was not significantly affected ($p > 0.05$), see Figure 2.15.B. Furthermore, the earliest eluting Normet peak (at 98% aqueous) was the most improved of the three compounds (see Figure 2.15.B and Figure 2.16.B) with this optimisation. This supports the theory that the narrower route into the ESI source (the smaller sleeve coupler i.d. and the smaller micro-nebuliser needle i.d.) reduced the dispersion and focused the flow, which improved the linear velocity of the spray hence improved the signal.

Figure 2.15.C and Figure 2.16.C show that %RSD of the ratios of five replicate injections dropped by about half in Normet and Met whereas it remained similar in the case of 3MT. This resulted in improved measurement uncertainty to half of what it was when using the standard size i.d. for both the coupler sleeve and the nebuliser needle. The improvement in measurement uncertainty as the %RSD drops from up to 5.5 % for Normet to <2 % and from 2.3 % to <2 % for Met. The uncertainty on the 3MT was not improved as the %RSD between the two experiments was the similar. This is because the optimised parts had most pronounced effect on the aqueous flow ionisation rather than the gradient in which 3MT eluted. This was inline with what the work set out to achieve to ultimately reduce the measurement uncertainty and mitigate the sources of variability.

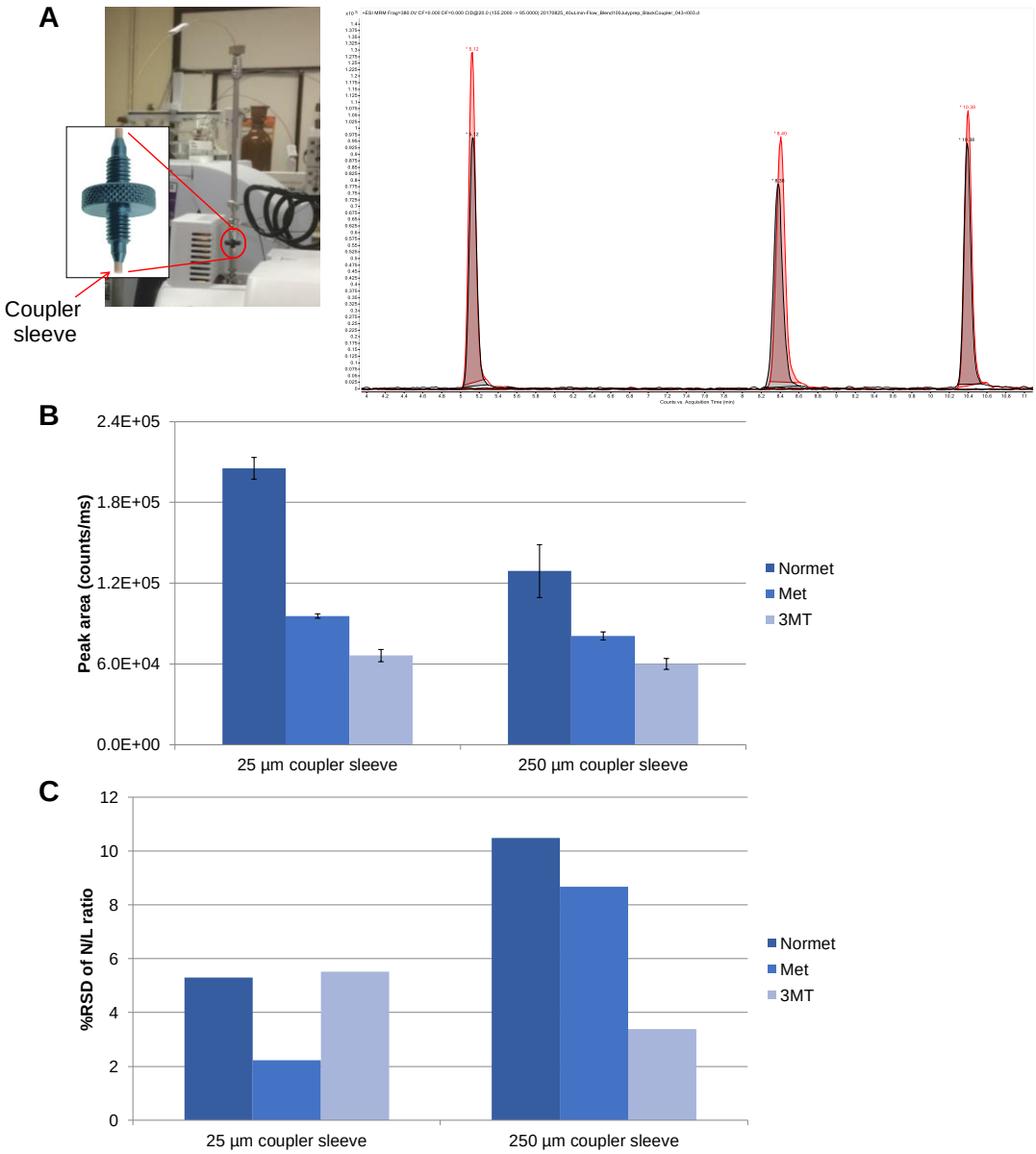


Figure 2.15 The coupler sleeve i.d. optimisation impact on sensitivity; A) an example chromatogram of overlaid SRM transitions of METs using 250 µm PEEK tubing for the coupler sleeve i.d. (black) and 25 µm i.d. (red). B) comparison of peak areas of METs using the different coupler sleeve internal diameters, significant difference for Normet and Met ($p < 0.05$) whereas no significant difference for 3MT peak area ($p > 0.05$). Error bars = standard deviation of 5 replicate analysis of METs standards in solvent at 100 pg/g. C) %RSD of the N/L ratios of 5 replicate analysis of METs standards in solvent at 100 pg/g.

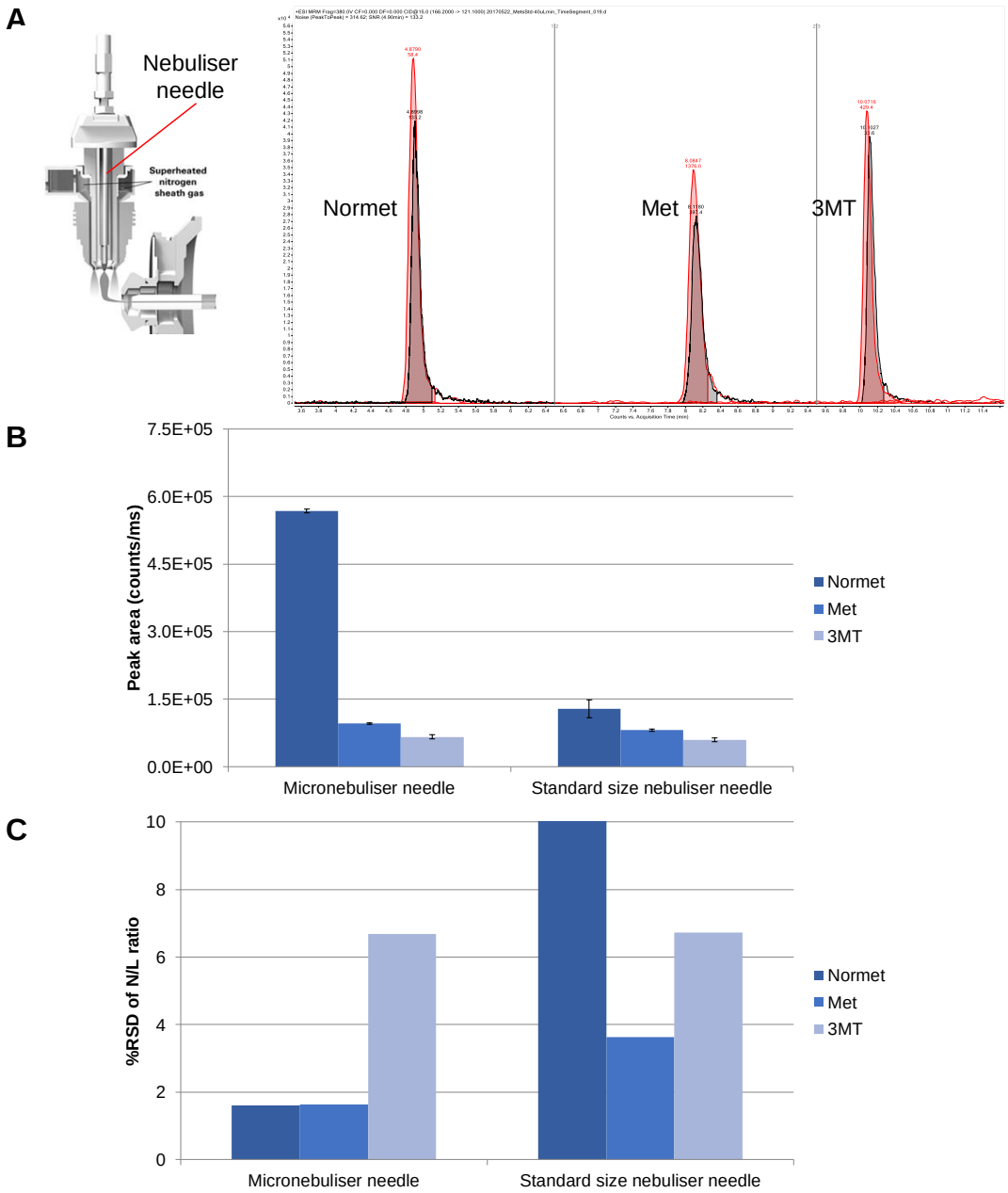


Figure 2.16 The ESI nebuliser needle i.d. optimisation impact on sensitivity; A) an example chromatogram of overlaid SRM transitions of METs using standard size nebuliser needle (black) and micro nebuliser needle (red). B) comparison of peak areas of METs using the different needles, significant difference for Normet and Met ($p < 0.05$) whereas no significant difference for 3MT peak area ($p > 0.05$). Error bars = standard deviation of 5 replicate analysis of METs standards in solvent at 100 pg/g. C) %RSD of the N/L ratios of 5 replicate analysis of METs standards in solvent at 100 pg/g.

2.3.7 Instrument equilibration time

Unexpectedly, it was found that the relative standard deviation of the ratios acquired by the instrument in its first few hours of operation was much higher than later in the analytical sequence. This could be due to the time required for the instrument parts to equilibrate to temperatures and voltages and gases from being on standby in the laboratory. The temperature of all the laboratories that the work was conducted in both sample preparation and the LC-MS analysis were controlled and maintained at 20 ±2 °C. This was important to reduce the instrument's variability in measurement hence the overall measurement variability and uncertainty.

2.3.8 Plasma sample preparation method development

Parallel to the development of the LC-MS/MS method, the sample preparation method was also in development. The method was an SPE based method using a mixed-mode WCX sorbent, the selection of this sorbent chemistry is due to the polarity of METs where the weak cation exchange functions would retain molecules at the correct pH. WCX was chosen over strong cation exchange (SCX) as it could provide a higher level of selectivity by retaining more matrix than SCX because SCX would require a strong reagent for elution. One of the work limitations was not performing a comparison experiment between SCX and WCX.

Because the pK_a of the mixed mode WCX sorbent is about 5 whereas the pK_a of METs ranges between 9-10, see Figure 2.3, the SPE was performed in the middle of the two pK_a values at pH of 7 hence, the sample preparation process included buffering the sample to pH of 6.5-7. At pH 7 both the stationary

phase and the analytes were largely fully ionised. The development of the sample preparation method was lengthy and included a wide range of experiments, some of which are summarised in the diagram in Figure 2.17.

The method used for the initial performance evaluation was adapted and developed from a method used by Waters™ (52) produced satisfactory results. The results of the plasma analysis using this method provided measurements that had expanded measurement uncertainty estimates of ±6 % for Normet at 98 pg/g and ±6.6 % for Met at 46 pg/g and ±20 % for 3MT at 3 pg/g, using k = 2 coverage factor, at the 95 % confidence interval. An example chromatogram is shown in Figure 2.18.

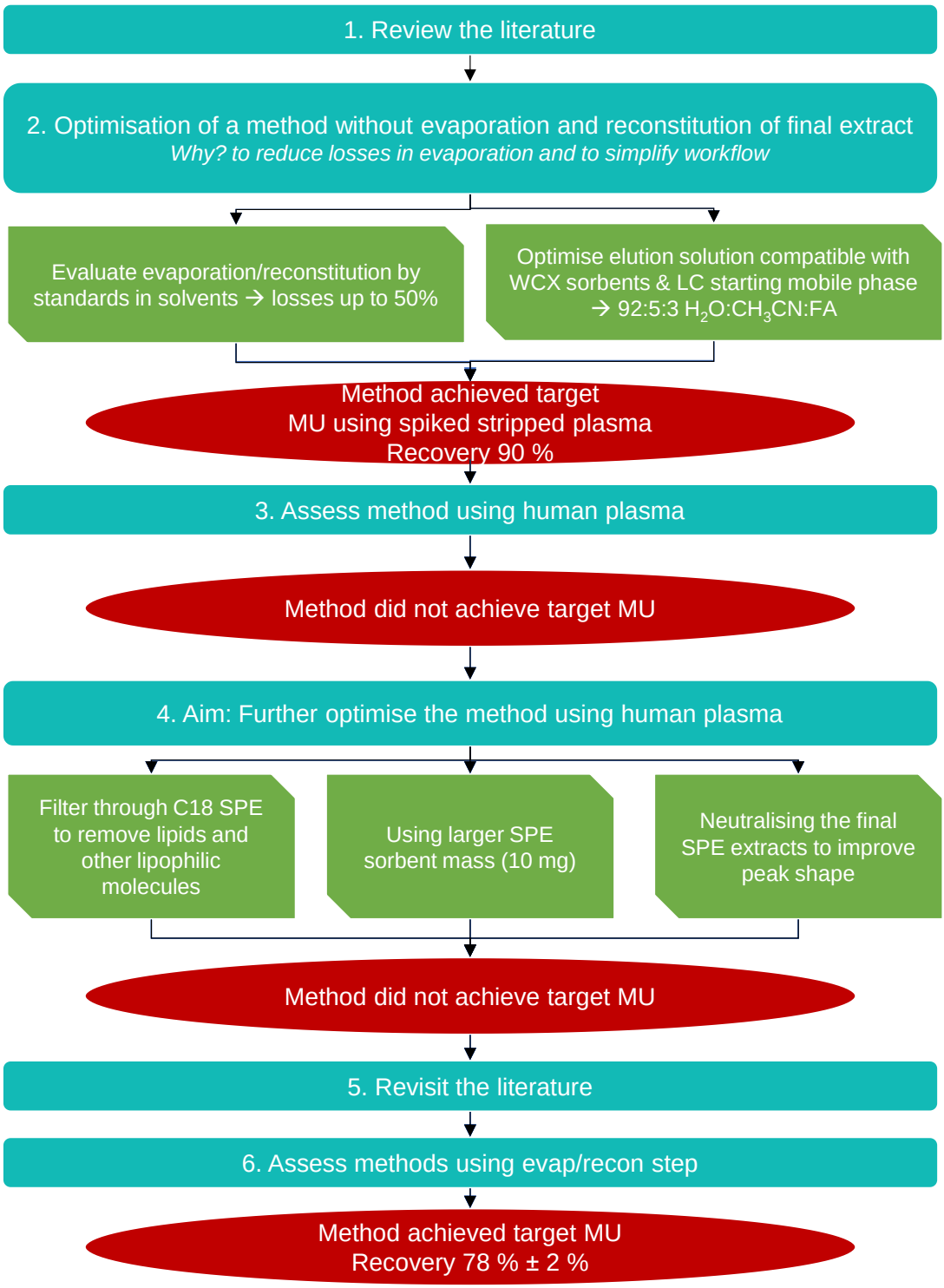


Figure 2.17 Sample preparation method development workflow diagram.

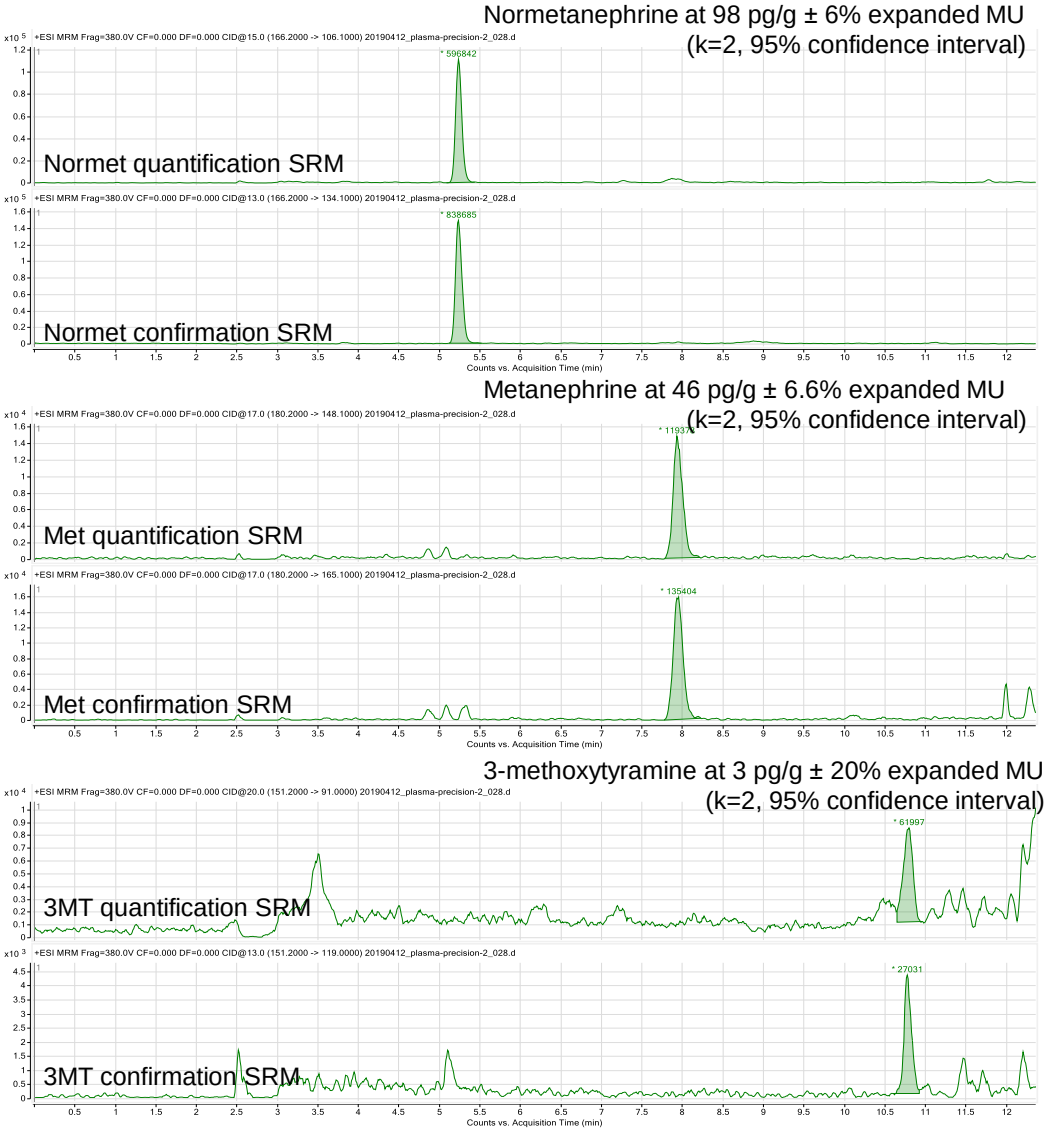


Figure 2.18 Example chromatogram of human plasma analysis using the optimised SPE method and micro-flow LC-MS/MS using DEM-IDMS calibration achieving measurement uncertainty.

As discussed in Chapter 1 (see 1.9), the ideal measurement uncertainty estimate for an analytical method is what is fit for the purpose of the analysis. In the assignment of higher-order reference measurement, the measurement uncertainty needs to be the lowest possible analytically as this method/measurement would be at the top of the traceability chain (see Figure 1.1). Typically, expanded measurement uncertainty estimates <7 % are reported for higher-order reference materials depending on the technical

challenges of the analysis (60). This uncertainty needs to be small because subsequent secondary reference measurement procedure would be using reference values assigned by the higher-order reference measurement and this measurement uncertainty estimate will take part of their measurement uncertainty budget hence it needs to be as small as possible (69).

The expanded measurement uncertainty of 6 % and 6.6 % for Normet and Met was deemed fit for purpose. As for the 3MT, the measurement uncertainty estimate was 20 %. The low mass fraction of 3MT (3 pg/g) resulted in such higher uncertainty estimate, further work is needed to develop a method of lower uncertainty for 3MT measurement. This could include developing an independent method using a different analytical technique or a separate sample preparation method which would start with a much larger sample volume to concentrate the sample further and gain better signal for 3MT which would reduce the variability in the measurement. However, given that 3MT is not part of the clinical panel of the plasma test (see Table 1.1) and was only investigated for potential clinical relevance as a clinical diagnostic biomarker for specific types of pheochromocytoma (42), it was decided to not include it in the higher-order reference method submission to the JCTLM-database, whenever the method is submitted.

The performance testing of this method and the analysis of example samples was to be performed at this stage of the method.

2.3.9 Method transfer & further optimisation

After the method was deemed fit to start the method performance testing, the LC-MS instrument was no longer available for use at the NML. The method was transferred to the AB Sciex Triple Quad™ 6500+ with an Agilent 1260 Infinity Capillary Pump liquid chromatography system. The same LC method gradient program, flow rate, injection volume and column temperature were used. The method was transferred successfully, all the optimisation experiments performed on the Agilent 6490 and described in this chapter were performed on the AB Sciex Triple Quad™ 6500+ instrument. Initial method performance starting when two issues arose that required further optimisation of the method before embarking on the performance testing experiments.

2.3.9.1 Interferences

At the preparation step the amount of SILIS to be added to the sample and calibration blends was calculated to obtain a response ratio of the compound to SILIS of one ±10 %. However, the measured ratios Normet were up to 1.3. This required further investigation because previous indicative attempts at preparing the sample blends provided a ratio of one. Upon examination of the blank blends, interference peaks at similar retention times to those of the quantitative peak of Normet were detected in what was supposedly blank plasma. These interferences biased the ratios of Normet by 20-30 % (see the ratios in Figure 2.19).

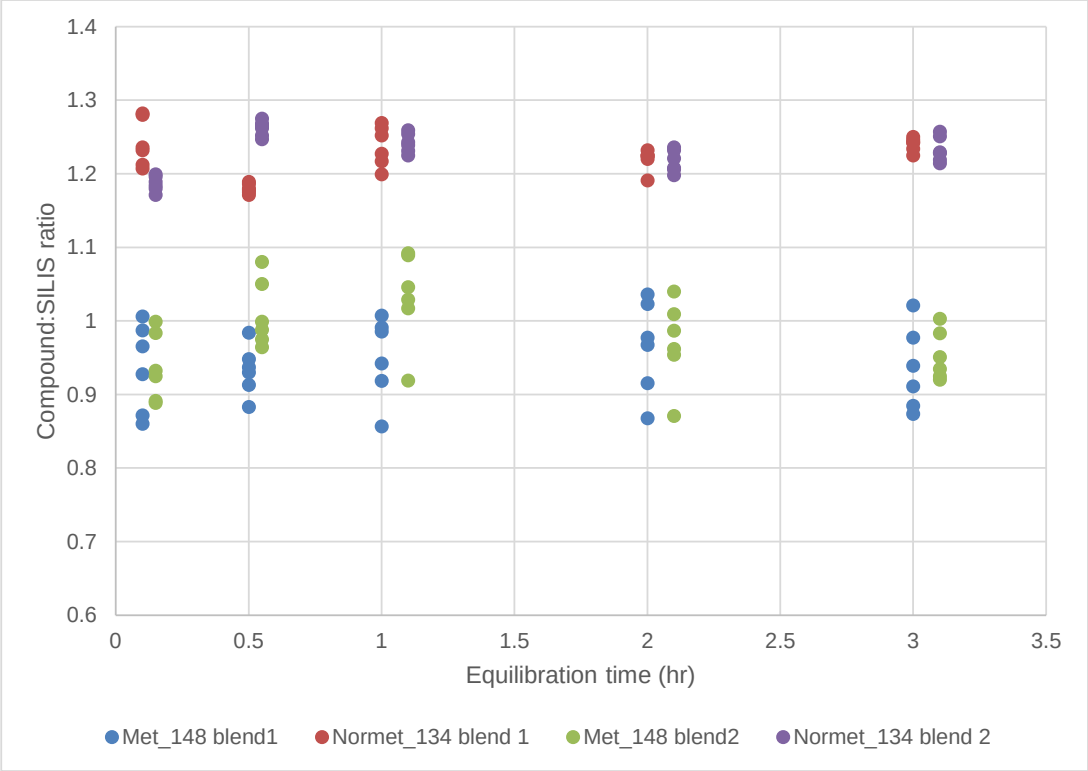


Figure 2.19 Comparison of the biased measured ratio of Normet to Met, the plotted data points are the ratio of the response of the compound to its corresponding SILIS, each blend was injected five times.

A thorough investigation to find the sources of these sudden interferences was required because remained unchanged. The blank plasma that was used had been assessed several times and confirmed as being free of Normet and Met or any isobaric interference. Upon investigating the different potential sources of these interferences, including assessing the SILIS potential exchange of deuterium to hydrogen, it was found that the interferences were coming from the SPE plate. Other SPE plates of different batches from the same manufacturer were tested and found to have these interferences despite having had used the same type of plates for over 18 months. Only upon using a different branded WCX SPE plate from a different manufacturer which had the same stationary phase chemistry did the interferences disappear.

Therefore, the WCX SPE plates used thereafter were the Thermo WCX SPE plates.

2.3.9.2 Chromatography troubleshooting

With the performance testing of the method, it was the first batch to be 4.5 days of running time. Previous running analysis during the method development stage were not as long because they did not include as many samples. Each injection running time was 37 minutes and the replicate analysis of all the different samples and calibration blends increased the running time significantly. Under these circumstances a new issue came up, the chromatographic method showed poor robustness as the peak shape deteriorated throughout the run and the retention times kept shifting throughout the analysis. This poor chromatography introduced variability to the peaks' integration process. Additionally, the peak shape was deteriorating from the beginning of the analysis to the end.

Several experiments were performed to resolve the issue and improve the chromatography including a longer column wash step with acetonitrile in the LC gradient program, testing new columns and trying a PFP guard column. It was found that the LC column's performance was deteriorating across long batches as the retention times were shifting constantly to earlier times. This indicated that the plasma extracts were not clean enough and the co-extractives were accumulating on the LC column causing severe deterioration of the stationary phase. Here, a second SPE step was introduced to the sample preparation method where after the completion of the WCX SPE step. The WCX SPE extracts were diluted down to 20:80 methanolic extract:water

then were loaded onto C_{18} SPE cartridges. The loaded diluted extracts were collected and evaporated down to dryness and then reconstituted with 100 µL of water. The combination of a second C_{18} SPE step and the PFP guard cartridge provided a robust chromatographic separation, with %RSD of the retention time of Normet dropping from 4.5 % to 1.01 % and for Met from 2.0 % to 0.4 %. This optimised two-step SPE procedure was used thereafter.

The optimisation work described in this chapter enabled a better understanding of the sources of variability in measurement and established a higher-order RMP. This higher-order RMP would ultimately support the standardisation of plasma METs measurements in UK hospital laboratories and globally.

2.4 Conclusions

A study to investigate the sources of variability in higher-order measurement of ultra-low level small molecules in plasma was performed using plasma METs as the case study. It was found that the key factors to reduce variability in the measurement included: the optimisation of instrument sensitivity, optimisation of MS dwell time, the use of appropriate chromatography and appropriate sample preparation. It was also found that the use of micro-flow liquid chromatography improved the performance of the MS instrument by increasing the ionisation efficiency resulting in increased signal which subsequently reduced the variability in the measurement to achieve high accuracy measurements of the compounds.

Therefore, the hypothesis of this chapter holds true as the objectives were achieved and all the major sources of measurement uncertainty for the measurements of clinically relevant small molecules at the lower end of the molecular weight range were determined and additionally mitigated to improve the RMP performance.

2.5 Novelty & Value

Better understanding of the sources of variability in measurement of ultra-low level small molecules in plasma was established. Moreover, a candidate higher-order reference measurement procedure for the analysis of plasma METs was developed. This would ultimately support the standardisation of plasma METs analysis to enable assigning standardised clinical cut-off limits based on traceable measurements and support hospital laboratories assess their methods through EQAS samples that are assigned traceable reference values instead of the consensus mean value currently used.

2.6 Future work

Improvement of measurement uncertainty estimates could be achieved by further investigating the sources of variability and possibly by using a more sensitive instruments if available.

Developing a method with a smaller measurement uncertainty estimate for the analysis of 3MT requires further work, that could include developing a sample preparation method that would start with a larger sample volume to enable further concentration of the extract to give a higher signal on the MS and subsequently smaller uncertainty.

The following chapter, Chapter 3, covers the performance testing of the method developed in this chapter and analysis of example samples. In addition to the traceability aspect of the measurements using this method.

2.7 References

1. Veillet-Chowdhury M, Hassan SF, Stergiopoulos K. Takotsubo cardiomyopathy: A review. Acute Card Care. 2014 Mar 1;16(1):15–22.
2. PRESSOR - Pheochromocytoma Research Support Organization [Internet]. [cited 2016 Jan 4]. Available from: http://www.pressor.org/
3. McNeil AR, Blok BH, Koelmeyer TD, Burke MP, Hilton JM. Phaeochromocytomas discovered during coronial autopsies in Sydney, Melbourne and Auckland. Aust N Z J Med. 2000 Dec 1;30(6):648–52.
4. Amar L, Eisenhofer G. Diagnosing phaeochromocytoma/paraganglioma in a patient presenting with critical illness: biochemistry versus imaging. Clin Endocrinol (Oxf). 2015 Sep;83(3):298–302.
5. Eisenhofer G, Huynh T-T, Hiroi M, Pacak K. Understanding Catecholamine Metabolism as a Guide to the Biochemical Diagnosis of Pheochromocytoma. Rev Endocr Metab Disord. 2001 Aug;2(3):297–311.
6. Lenders JW, Eisenhofer G, Keiser HR, Willemsen JJ, Thien T, Kloppenborg PWC, et al. 114 Plasma metanephrines in the diagnosis of pheochromocytoma. J Hypertens. 1993;11:S446.
7. Longo D, Fauci A, Braunwald E, Hauser S, Jameson J, Neumann HP. Harrison's Principles of Internal Medicine. 18th ed. McGaw-Hill; Chap 343.
8. Eisenhofer G, Keiser H, Friberg P, Mezey E, Huynh T-T, Hiremagalur B, et al. Plasma metanephrines are markers of pheochromocytoma produced by catechol-O-methyltransferase within tumors. J Clin Endocrinol Metab. 1998;83(6):2175–85.
9. Goldstein David S., Eisenhofer Graeme, Flynn John A., Wand Gary, Pacak Karel. Diagnosis and Localization of Pheochromocytoma. Hypertension. 2004 May 1;43(5):907–10.
10. Adler JT, Meyer-Rochow GY, Chen H, Benn DE, Robinson BG, Sippel RS, et al. Pheochromocytoma: Current Approaches and Future Directions. The Oncologist. 2008 Jan 7;13(7):779–93.
11. Lenders JWM, Duh Q-Y, Eisenhofer G, Gimenez-Roqueplo A-P, Grebe SKG, Murad MH, et al. Pheochromocytoma and Paraganglioma: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2014 Jun 1;99(6):1915–42.
12. Lenders JW m., Pacak K, Eisenhofer G. New Advances in the Biochemical Diagnosis of Pheochromocytoma. Ann N Y Acad Sci. 2002 Sep 1;970(1):29–40.
13. Lenders JM, Pacak K, Walther MM, et al. Biochemical diagnosis of pheochromocytoma: Which test is best? JAMA. 2002 Mar 20;287(11):1427–34.
14. Lenders JW, Eisenhofer G, Mannelli M, Pacak K. Phaeochromocytoma. The Lancet. 2005 Aug;366(9486):665–75.
15. Lenders JW, Willemsen JJ, Eisenhofer G, Ross HA, Pacak K, Timmers HJ, et al. Is supine rest necessary before blood sampling for plasma metanephrines? Clin Chem. 2007;53(2):352–4.
16. Eisenhofer G, Därr R, Pamporaki C, Peitzsch M, Bornstein S, Lenders JW. Supine or Sitting? Economic and other considerations for use of plasma

metanephrines for diagnosis of phaeochromocytoma. Clin Endocrinol (Oxf) [Internet]. 2014 [cited 2015 Apr 22]; Available from: http://onlinelibrary.wiley.com/doi/10.1111/cen.12602/abstract
17. Därr R, Pamporaki C, Peitzsch M, Miehle K, Prejbisz A, Peczkowska M, et al. Biochemical diagnosis of phaeochromocytoma using plasma-free normetanephrine, metanephrine and methoxytyramine: importance of supine sampling under fasting conditions. Clin Endocrinol (Oxf). 2014 Apr 1;80(4):478–86.
18. Chortis V, Bancos I, Crowley RK, Arlt W. Supine or sitting? Economic considerations regarding patient position during plasma metanephrine analysis for the exclusion of chromaffin tumours. Clin Endocrinol (Oxf). 2015 Mar;82(3):462–3.
19. Eisenhofer G, Därr R, Pamporaki C, Peitzsch M, Bornstein S, Lenders JWM. Supine or Sitting? Economic and other considerations for use of plasma metanephrines for diagnosis of phaeochromocytoma. Clin Endocrinol (Oxf). 2015;82(3):463–4.
20. Rao D, Peitzsch M, Prejbisz A, Hanus K, Fassnacht M, Beuschlein F, et al. Plasma methoxytyramine: clinical utility with metanephrines for diagnosis of pheochromocytoma and paraganglioma. Eur J Endocrinol. 2017 May 5;177(2):103–13.
21. Human Metabolome Database [Internet]. [cited 2019 Nov 11]. Available from: http://www.hmdb.ca/
22. Yi Z, Brown PR. Chromatographic methods for the analysis of basic neurotransmitters and their acidic metabolites. Biomed Chromatogr. 1991 May 1;5(3):101–7.
23. de Jong WHA, de Vries EGE, Kema IP. Current status and future developments of LC-MS/MS in clinical chemistry for quantification of biogenic amines. Clin Biochem. 2011 Jan;44(1):95–103.
24. Petteys BJ, Graham KS, Parnás ML, Holt C, Frank EL. Performance characteristics of an LC–MS/MS method for the determination of plasma metanephrines. Clin Chim Acta. 2012 Oct 9;413(19–20):1459–65.
25. Determination of Homovanillic Acid and Vanillylmandelic Acid in Neuroblastoma Screening by Stable Isotope Dilution GC-MS - Fauler - 1998 - Journal of Mass Spectrometry - Wiley Online Library [Internet]. [cited 2016 Jan 4]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1096-9888(199705)32:5%3C507::AID-JMS503%3E3.0.CO;2-9/abstract
26. Rapid Analysis of Metanephrine and Normetanephrine in Urine by Gas Chromatography-Mass Spectrometry [Internet]. [cited 2016 Jan 4]. Available from: http://www.clinchem.org/content/48/2/332.short
27. Kema IP, Meiborg G, Nagel GT, Stob GJ, Muskiet FAJ. Isotope dilution ammonia chemical ionization mass fragmentographic analysis of urinary 3-O-methylated catecholamine metabolites: Rapid sample clean-up by derivatization and extraction of lyophilized samples. J Chromatogr B Biomed Sci App. 1993 Aug 11;617(2):181–9.
28. Jong WHA de, Graham KS, Molen JC van der, Links TP, Morris MR, Ross HA, et al. Plasma Free Metanephrine Measurement Using Automated Online Solid-Phase Extraction HPLC–Tandem Mass Spectrometry. Clin Chem. 2007 Jan 9;53(9):1684–93.
29. Current role of liquid chromatography–mass spectrometry in clinical and forensic toxicology - Springer [Internet]. [cited 2016 Jan 4]. Available from: http://link.springer.com/article/10.1007/s00216-007-1248-5
30. An Electrochemical Detector for Liquid Chromatography with Picogram Sensitivity - Analytical Letters - Volume 6, Issue 5 [Internet]. [cited 2016 Jan4].Availablefrom:

http://www.tandfonline.com/doi/abs/10.1080/00032717308058694#.VoqBl1e5jJs
31. Lenders JW, Eisenhofer G, Armando I, Keiser HR, Goldstein DS, Kopin IJ. Determination of metanephrines in plasma by liquid chromatography with electrochemical detection. Clin Chem. 1993 Jan 1;39(1):97–103.
32. Roden M, Raffesberg W, Raber W, Bernroider E, Niederle B, Waldhäusl W, et al. Quantification of Unconjugated Metanephrines in Human Plasma without Interference by Acetaminophen. Clin Chem. 2001 Jan 6;47(6):1061–7.
33. Davidson FD. Paracetamol-associated interference in an HPLC-ECD assay for urinary free metadrenalines and catecholamines. Ann Clin Biochem. 2004 Jan 7;41(4):316–20.
34. Madhavaram H, Woollard GA. Interference from Indian diet on the internal standard in a commercial method for the measurement of urinary metanephrines by high-performance liquid chromatography with electrochemical detection. Ann Clin Biochem Int J Biochem Lab Med. 2014 May 1;51(3):400–5.
35. Eisenhofer G, Peitzsch M. Laboratory Evaluation of Pheochromocytoma and Paraganglioma. Clin Chem. 2014 Jan 12;60(12):1486–99.
36. Ito T, Imai T, Kikumori T, Shibata A, Horiba T, Kobayashi H, et al. Adrenal Incidentaloma: Review of 197 Patients and Report of a Drug-Related False-Positive Urinary Normetanephrine Result. Surg Today. 2006 Nov;36(11):961–5.
37. Bouhanick B, Fauvel J, Pont F. BIochemical misdiagnosis of pheochromocytoma in patients treated with sulfasalazine. JAMA. 2010 Nov 3;304(17):1898–901.
38. Keevil B. Personal Communication with Head of Biochemistry Department at University Hospital of South Manchester. 2015.
39. Dunand M, Donzelli M, Rickli A, Hysek CM, Liechti ME, Grouzmann E. Analytical interference of 4-hydroxy-3-methoxymethamphetamine with the measurement of plasma free normetanephrine by ultra-high pressure liquid chromatography–tandem mass spectrometry. Clin Biochem. 2014 Aug;47(12):1121–3.
40. Wright MJ, Thomas RL, Stanford PE, Horvath AR. Multiple Reaction Monitoring with Multistage Fragmentation (MRM3) Detection Enhances Selectivity for LC-MS/MS Analysis of Plasma Free Metanephrines. Clin Chem. 2015 Jan 3;61(3):505–13.
41. Peitzsch M, Adaway JE, Eisenhofer G. Interference from 3-O-methyldopa with ultra-high performance LC-MS/MS measurements of plasma metanephrines: chromatographic separation remains important. Clin Chem. 2015 Jul;61(7):993–6.
42. Peitzsch M, Prejbisz A, Kroiß M, Beuschlein F, Arlt W, Januszewicz A, et al. Analysis of plasma 3-methoxytyramine, normetanephrine and metanephrine by ultraperformance liquid chromatography–tandem mass spectrometry: utility for diagnosis of dopamine-producing metastatic phaeochromocytoma. Ann Clin Biochem. 2013 Jan 3;50(2):147–55.
43. ISO 15189:2012 - Medical laboratories -- Requirements for quality and competence [Internet]. [cited 2018 Aug 13]. Available from: https://www.iso.org/standard/56115.html
44. Adaway JE, Peitzsch M, Keevil BG. A novel method for the measurement of plasma metanephrines using online solid phase extraction-liquid chromatography tandem mass spectrometry. Ann Clin Biochem Int J Biochem Lab Med. 2014 Jul 21;0004563214546560.
45. Peaston RT, Graham KS, Chambers E, van der Molen JC, Ball S. Performance of plasma free metanephrines measured by liquid chromatography–tandem

mass spectrometry in the diagnosis of pheochromocytoma. Clin Chim Acta. 2010 Apr 2;411(7–8):546–52.
46. Graham KS, van der Molen J, Stob G, Cooper DP, Morris MR, Keevil BG, et al. Measurement of plasma free metanephrines using on-line solid phase extraction high performance liquid chromatography tandem mass spectrometry. In: CLINICAL CHEMISTRY [Internet]. AMER ASSOC CLINICAL CHEMISTRY 2101 L STREET NW, SUITE 202, WASHINGTON, DC 20037-1526 USA; 2006 [cited 2015 Apr 22]. p. A90–A90. Available from: http://www.waters.com/webassets/cms/library/docs/wa43228.pdf
47. Gabler J. A Sensitive and Interference-Free Liquid Chromatography Tandem Mass Spectrometry Method for Measuring Metanephrines in Plasma. J Chromatogr Sep Tech [Internet]. 2013 [cited 2016 Jan 6];04(07). Available from: https://www.longdom.org/open-access/a-sensitive-and-interference-free-liquid-chromatography-tandem-mass-spectrometry-method-for-measuring-metanephrines-in-plasma-2157-7064.1000195.pdf
48. Lagerstedt SA, O'Kane DJ, Singh RJ. Measurement of Plasma Free Metanephrine and Normetanephrine by Liquid Chromatography–Tandem Mass Spectrometry for Diagnosis of Pheochromocytoma. Clin Chem. 2004 Jan 3;50(3):603–11.
49. Anderson J, Berthod A, Pino V, Stalcup AM. Analytical Separation Science, 5 Volume Set. John Wiley & Sons; 2016. 2148 p.
50. Coskun O. Separation techniques: Chromatography. North Clin Istanb. 2016 Nov 11;3(2):156–60.
51. Snyder LR, Kirkland JJ, Dolan JW. Introduction to modern liquid chromatography. 3rd ed. John Wiley and Sons Inc.; 1974.
52. Danaceau J. Plasma metanephrines SPE discussion. 2018.
53. Bedson P. Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS) [Internet]. 2002 [cited 2018 Jun 25]. Available from: http://pubs.rsc.org/en/Content/eBook/978-0-85404-418-4 pages 1-34.
54. Alonso J, Gonzalez P. Isotope Dilution Mass Spectrometry [Internet]. 2013 [cited 2019 Nov 26]. Available from: https://pubs.rsc.org/en/content/ebook/978-1-84973-333-5 - pages 10-150
55. turbo-v-operator-guide-en.pdf [Internet]. [cited 2019 Nov 14]. Available from: https://sciex.com/Documents/Downloads/Literature/turbo-v-operator-guide-en.pdf
56. Momoh P, Fandino A, Aisawa E, Schlabach T, Miller K, Stafford G, et al. Now available on the New Agilent 6490 Triple Quadrupole LC/MS System. :8.
57. Xevo G2 QTof Operator's Overview and Maintenance Guide Revision B. :248.
58. HESI-II Probe - User Guide. :36.
59. Eisenhofer G, Peitzsch M, McWhinney BC. Impact of LC-MS/MS on the laboratory diagnosis of catecholamine-producing tumors. TrAC Trends Anal Chem. 2016 Nov 1;84(Part B):106–16.
60. Database of higher-order reference materials, measurement methods/procedures and services [Internet]. [cited 2018 Feb 6]. Available from: https://www.bipm.org/jctlm/
61. Xevo TQ-XS Triple Quadrupole Mass Spectrometry : Waters [Internet]. [cited 2019 Nov 15]. Available from: https://www.waters.com/waters/en_US/Xevo-TQ-XS-Triple-Quadrupole-Mass-Spectrometry/nav.htm?locale=en_US&cid=134889751
62. Abian J, Oosterkamp AJ, Gelpí E. Comparison of conventional, narrow-bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry: practical considerations. J Mass Spectrom. 1999;34(4):244–54.

63. Shen Y, Zhao R, Berger SJ, Anderson GA, Rodriguez N, Smith RD. High-Efficiency Nanoscale Liquid Chromatography Coupled On-Line with Mass Spectrometry Using Nanoelectrospray Ionization for Proteomics. Anal Chem. 2002 Aug 1;74(16):4235–49.
64. Duan X, Weinstock-Guttman B, Wang H, Bang E, Li J, Ramanathan M, et al. Ultrasensitive Quantification of Serum Vitamin D Metabolites Using Selective Solid-Phase Extraction Coupled to Microflow Liquid Chromatography and Isotope-Dilution Mass Spectrometry. Anal Chem. 2010 Mar 15;82(6):2488–97.
65. Uclés Moreno A, Herrera López S, Reichert B, Lozano Fernández A, Hernando Guil MD, Fernández-Alba AR. Microflow Liquid Chromatography Coupled to Mass Spectrometry—An Approach to Significantly Increase Sensitivity, Decrease Matrix Effects, and Reduce Organic Solvent Usage in Pesticide Residue Analysis. Anal Chem. 2015 Jan 20;87(2):1018–25.
66. Legido-Quigley C, Smith NW, Mallet D. Quantification of the sensitivity increase of a micro-high-performance liquid chromatography–electrospray ionization mass spectrometry system with decreasing column diameter. J Chromatogr A. 2002 Nov 8;976(1):11–8.
67. European U. European Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. Off J Eur Coomun. 2002 Aug 12;L221:8–36.
68. Kruve A, Rebane R, Kipper K, Oldekop M-L, Evard H, Herodes K, et al. Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. Anal Chim Acta. 2015 Apr 22;870:29–44.
69. Ellison S, Williams A. EURACHEM/CITAC Guide CG4 - Quantifying Uncertainty in Analytical Measurement [Internet]. Eurochem & CITAC - Co-Operation on International Traceability in Analytical Chemistry; 2012 [cited 2017 Nov 14]. Available from: https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf

3 Sources of variability in method performance testing – small polar molecules in plasma

3.1 Introduction

3.1.1 Higher-order reference measurement

When a national measurement institute (NMI) such as the National Measurement Laboratory (NML) at LGC embarks on developing a higher-order RMP it would be for one or more objectives that include a) the production of a higher-order certified reference material (CRM); b) provide a measurement service that could be to CRM producers, external quality assurance schemes (EQAS) and clinical reference measurement laboratories, where only specific samples are assigned reference value rather than a wide range of routinely tested samples. For NMIs to provide such services, they aim to enlist their service on to the Joint Committee for Traceability in Laboratory Medicine and Clinical Chemistry (JCTLM) database of RMPs and CRMs for the clinical laboratory and *in vitro* diagnostics (hereafter mentioned as “the JCTLM database”). The method validation required for a RMP to be listed on the JCTLM database is the compliance with ISO standard 15193:2009 (1). The work described in this chapter is in compliance with ISO 15193:2009 (1,2) following the method performance characteristics of the ICH Tripartie Guideline; Validation of Analytical Procedures (3), referred to hereafter as the ICH guidelines.

3.1.2 Analytical methods validation

Method validation is *'the process of defining an analytical requirement, and confirming that the method under consideration has performance capabilities consistent with what the application requires'* (4). Measurement validation is based on detailed processes that are well-defined and generally accepted concepts and terms. However, historically there had been a lack of common terminology to cover the practical basis and concepts of validation (5). There are numerous guidelines on the validation of analytical methods. These include: the Eurachem guidelines (4), the Food and Drug Agency (FDA) guide on bioanalytical method validation (6), the International Council for Harmonisation (ICH) (3) and the European Medicine Agency (EMEA) guidelines (7). While the FDA and EMEA use harmonised terminology, some of these guidelines have different terminologies for different concepts. For example the concepts of specificity and accuracy in the ICH guidelines are the same as those of selectivity and trueness in the Eurachem guidelines, respectively (3,4). However, efforts for the harmonisation of these terms have been made and their use confirms that the key concepts of validation of bioanalytical methods are similar (5,8,9).

Fundamentally, the performance characteristics that all these guidelines recommend are the same. Ultimately, they aim to assess the error in an analytical method, describe and where possible quantify it to assess the method's suitability for its intended application. Menditto *et al.* using the internationally accepted terminology, simplified the key concepts of method validation in a straightforward figure that was adapted in Figure 3.1 (10). In

Figure 3.1, the total error which comprises both the systemic and random error, can be explained by the qualitative tools (e.g., accuracy and precision) and could be quantified as the standard deviation or measurement uncertainty.

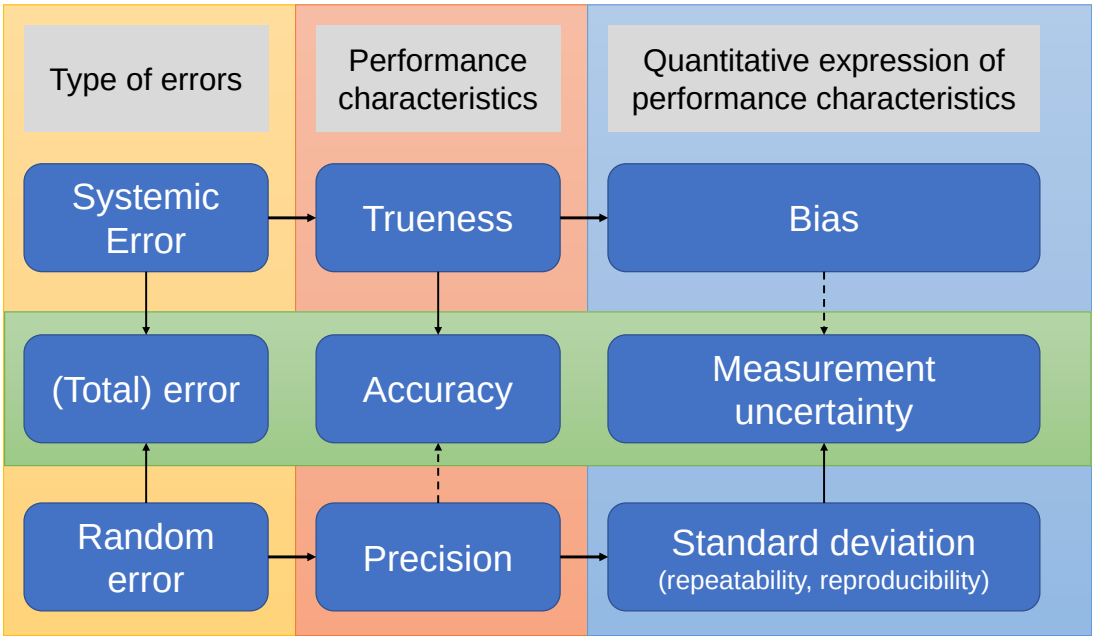


Figure 3.1 The expressions of error and their combination showing the qualitative and quantitative expressions. Given that bias should be removed when possible the line connecting it to the measurement uncertainty is dotted. Adapted from (10).

3.1.3 What measurement uncertainty estimate is fit-for-purpose?

Currently there is no clear guideline on the fit-for-purpose measurement uncertainty estimate for the analysis of plasma metanephrines (METs). Therefore, an estimation of what could be an acceptable measurement uncertainty could be derived from the EQAS data. The example EQAS data set described in (Figure 1.15) and the figure repeated below as Figure 3.2 was used to set this estimation. It was found that the estimated current acceptable range is 38.3 % of the overall range of measurements for both Normet and Met. As for the acceptable range's width, it was estimated to be 26.3 % and

33.7 % for Met and Normet, respectively. Therefore, for a RMP to add value to the EQAS it needs to provide an expanded measurement uncertainty that is smaller than 26.3 % and 33.7 % for Met and Normet, respectively. This estimation was used to assess the measurement uncertainty of the method.

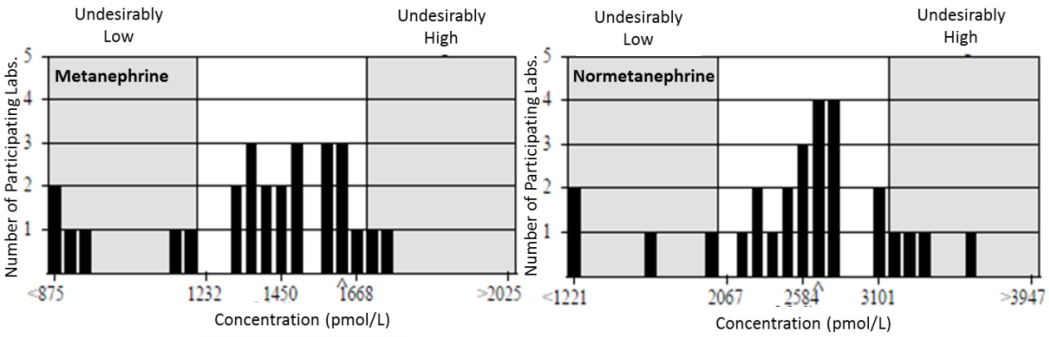


Figure 3.2 Example of RCPA-EQA plasma metanephrines analysis data. Data shown is of one sample analysed by 27 participating labs, 23 of which are using LC-MS. Charts in the figure used with permission from The Consultant Clinical Scientist of the Liverpool University Hospital, Mr Andrew Davison, from their laboratory's participation in the EQAS report.

3.1.4 Hypothesis and objectives

In line with the aim of this thesis expressed in Chapter 1 (section 1.13.2.1), the hypothesis for this chapter is that "the candidate RMP developed in Chapter 2 is fit-for-purpose, where the RMP can achieve traceable measurements with low measurement uncertainties".

In order to test this hypothesis, the following objectives were set:

- An assessment to define what measurement uncertainty is "fit-for-purpose" based on the EQAS data due to the absence of a clear guideline defining the measurement uncertainty for plasma METs RMP.
- Evaluate the accuracy, specificity and recovery of the candidate RMP.

- Evaluate the calibration mode and measurement range of the candidate RMP.
- Evaluate precision of the candidate RMP, repeatability and intermediate precision.
- Calculate the measurement uncertainty estimate of the candidate RMP.

3.2 Experimental

3.2.1 Reagents and standards

In addition to standards and reagents mentioned in section 2.2.14.2.1, the following METs standards were used and were assigned traceable purity values by qNMR: a) DL-normetanephrine hydrochloride that was purchased from Sigma Aldrich (Merk, Feltham, UK) and b) DL-metanephrine hydrochloride purchased from Clear Synth (Mumbai, India). Both solid standards were assigned traceable reference purity values by qNMR analysis performed by the purity analysis at the National Measurement Laboratories (LGC, Teddington, UK), the details of the purity analysis can be found in the certificate of analysis in appendix 7.1. The assigned traceable reference purity at values were 82.09 % ±0.61 % (at 95 % CI) and 81.49 % ±0.63 % (at 95 % CI) for Normet and Met, respectively. The certificates of analysis of the standards that were supplied from the vendors provided a non-traceable purity value of the salt rather than the free base. The purity reference values assigned at the NML were of the free base and were traceable to the SI unit (see 3.2.9.1.1 and full certificate of analysis in the Appendix 7.1.3).

The blank plasma that was used to prepare the calibration blend was SeraCon-II Normal Human Plasma, Double stripped, delipidated, Mat No. 1800-0058 - Batch 10455817 – Manufacture date 11Dec2019 from SeraCare Life Sciences (LGC Group, Milford, USA).

3.2.2 Instrumentation

3.2.2.1 LC-MS/MS method

The LC-MS/MS method and Other instruments used were described in 2.2.2.1 and 2.2.2.2.

3.2.3 Material production (sample T)

A candidate reference material was produced to assess the method performance. A frozen pooled human plasma from both sexes in potassium ethylenediaminetetraacetic acid (K_2EDTA) tubes, was purchased from Sera Laboratories - BioIVT (Sussex, UK). The material was received frozen and once defrosted was mixed, sub-aliquoted into 1.5 mL Eppendorf vials and then stored at -80 °C. The vials of the materials were then used for the analysis performed to assess the method performance. This material is referred to in this work as the T sample.

3.2.4 Preparation of standards in solvent

The preparation of standards in solvent was performed gravimetrically. Individual stock standards and subsequent dilution standards of Normet and Met were prepared in the same way. The traceable solid Normet and Met standards were taken out of the freezer at least 2 hours before the preparation

to equilibrate to room temperature. The solid powder standards were mixed thoroughly, by rotating the vial then mixing with the spatula, before an aliquot of 38.004 mg and 63.571 mg of the Normet and Met powders, respectively, was accurately weighed in a clean pre-weighed glass cup. This glass cup was custom-made (Southampton University glassblower, Southampton, UK) with dimensions of 1.5 cm height and 1.0 cm radius and was designed specifically to fit on the Mettler XP6 balance pan. The cup was then dropped in an amber 40 mL vial, capped and weighed using the Mettler XP205. The addition of methanol was also weighed being 23.32 g and 23.73 g methanol for Normet and Met, respectively. The vial was capped immediately after the addition then weighed to avoid any evaporation effects. Evaporation checks of standards were performed and are described later.

Because the balance with the smallest uncertainty (Mettler XP6) has a maximum capacity of 6 grams it was not used for weighing the methanol in the standards preparation. This balance was therefore used to for the weighing of the reference material which was between 38-65 mg. However, the mass of methanol that was required for the stock preparation was 23 g which required the use of the Mettler XP205 which has a maximum capacity of 250 g albeit a larger uncertainty, see Table 2.9 for uncertainties and maximum weighing capacity. To minimise evaporation losses, steps were taken in the preparation procedure including: a) the use of 40 mL amber vials with solid plastic caps lined with a PTFE-faced rubber liner and a long thread to ensure efficient closure and minimise evaporation; b) once methanol was added, all weighing steps were performed on the closed vials; c) constant

monitoring of potential evaporation during storage and use was performed and is described later.

The subsequent dilution working standards were also prepared gravimetrically using the Mettler XP205 balance. The smallest mass taken from the stock standards and higher mass fraction standards was >1 g to keep the uncertainty low. The reason to do the several weighing steps of the same addition was to ensure there was no evaporation taking place during the process. In the cases where the mass was observed to be trending downward, the vial was examined for capping properly or if a small drop of methanol had landed in the rim of the vial cap or on the outer surface for example. In such cases the standard would be disposed of and a fresh one prepared.

The mass fraction of each of the standards in methanol was accurately obtained by the gravimetric preparation procedure. The measurement uncertainty estimates of the mass fractions of the standards included the variation in the weighing process, the balance uncertainty and, most importantly, the purity value of the traceable solid standards. All standards in methanol were stored at -20 °C and were taken out of the freezer at least two hours before use to ensure equilibration to room temperature. Each standard was tightly capped and wrapped in foil.

To monitor any evaporation losses of the standards during storage and use an evaporation check was performed by weighing the standards in methanol before and after use. Before the use of any standard in solvent, it was weighed, and the mass was compared to the mass of the standard from after the previous use to check if any significant evaporation had taken place. If a

standard had more than 0.50 % loss in mass during the freezer storage and temperature equilibration time, it would not be used, and a new standard would be prepared.

Fresh working dilutions standards in water (sub 0.5 ng/g) were prepared on the day of the sample preparation.

3.2.5 Blends preparation and pre-treatment

The sample and calibration blends were prepared gravimetrically and the buffer added to them as described in sections 2.2.3.1 and 2.2.3.2. Each sample T and QC sample was analysed in duplicate, two independent blends.

3.2.6 Solid phase extraction

A two-step SPE was performed, where samples were prepared using a WCX SPE plate as described in 2.2.3.3 and 2.2.3.4.

3.2.7 DEM-IDMS sequence running order

Each sequence started with a single injection of each of the blank blends followed by the samples and another single injection of each of the blanks at the end of the sequence. For DEM-IDMS calibration; each sample blend (SB) was injected five times on the LC-MS/MS, each time bracketed by its ‘matching’ calibration blend (CB); [CB-SB-CB] x5 for each sample blend and each QC sample blend. The order of samples was randomised to vary from the preparation order to remove any possible trend in data due to the preparation order. Each of the blank blends (described below) were injected once at the beginning of the analytical sequence and once at the end.

3.2.8 Quality Control

3.2.8.1 Analysis of Quality Control Material

Two QC plasma METs materials were purchased from Australian Scientific Enterprise (Hornsby, Australia). The plasma METs materials were two levels and were named by the supplier as Level I QC and Level II QC. Level I QC (QC1) was certified at 0.5-0.7 nmol/L (91-128 pg/g) and 0.05-0.25 nmol/L (9-49 pg/g) for Normet and Met, respectively. Level II QC (QC2) was certified at 2.7-3.1 nmol/L (494-568 pg/g) and 1.0-1.3 nmol/L (193-256 pg/g) for Normet and Met, respectively. Both materials were lyophilised plasma and were reconstituted with 1.5 mL of water on the day of use as per the supplied instructions. The lyophilised materials were stored at 4 °C.

3.2.8.2 Blanks analysis

The blank blends prepared and analysed by LC-MS as listed below and were analysed at the beginning and the end of the LC-MS sequence:

- B-SPE: blank water that goes through the SPE procedure to assess for interferences from the SPE.
- B-S: One sample T that naturally contains METs without the addition of the SILIS.
- B+N: Blank plasma with the addition of METs standards in solvent only.
- B+L: Blank plasma with the addition of SILIS only.
- B-P: Blank plasma without any addition.
- B-QC1: Blank Level I QC sample without the addition of the SILIS.
- B-QC2: Blank Level II QC sample without the addition of the SILIS.

3.2.9 Method performance testing experiments

3.2.9.1 Accuracy

3.2.9.1.1 Primary traceable calibrator

The solid standards were assigned reference purity values that are traceable to the SI unit for them to be used as a primary calibrator. This work was commissioned to the UKAS accredited NML purity team. The details of their work are found in the certificates of analysis in Appendix 7.1.

3.2.9.1.2 Assessing the reliability of preparation of standards in methanol

This study included preparing two or more stock standards in methanol followed by three subsequent dilutions down to the working mass fraction level, on two different days by two different analysts for comparison. These standards were then measured against each other using the candidate RMP. The measured mass fraction of the standard should agree with the gravimetric preparation mass fraction. In the standards in solvent validation study, one standard was used as the sample to prepare the sample blends and the other was used as a standard to prepare the calibration blends.

3.2.9.2 Specificity

This was performed by two experiments of MS infusion and LC-MS injections of the SILIS standards in pure methanol and the primary calibration standard in methanol. The MS was operated in full scan to check for the presence of the precursor ions of the unlabelled compound in the SILIS and for the presence of the SILIS precursor in the unlabelled calibration standard.

The method's selectivity was assessed using several tools including: a) the assessment of compound:SILIS ratios; b) the analysis of blank reagents and blank samples (described in 3.2.8.2); and c) comparing quantification and confirmation SRMs. To assess the presence of interferences in the samples, a comparison of the quantitative and confirmation SRMs ratio was performed by comparing the ratio of standards in methanol to those in samples with an acceptable tolerance of ±10% between the standards in methanol and samples was considered acceptable.

3.2.9.3 Calibration model

SILIS equilibration study

Six vials of the sample T were analysed in duplicate. The SILIS addition was made to the sample blends at different times prior to the sample treatment and SPE. The times were -17 h, -3 h, -2 h, -1 h, -30 min and T_0 immediately before adding the buffer to all sample and calibration blends. Upon analysis of the samples, the ratios of natural compound to SILIS ratios across the different time points were compared. Each sample blend was analysed five times by LC-MS/MS.

3.2.9.4 Limit of detection and limit of quantification

The technical limit of detection (LOD) and lower limit of quantification (LLOQ) of the candidate RMP of plasma METs were estimated visually using signal-to-noise ($S:N_{p/p}$) of 3:1 and 10:1 for LOD and LOQ, respectively (3,4,11,12). The assessment was carried out by injecting smaller injection volumes of the

T sample, after it was assigned a reference value, to take into account the matrix effect on the assessment.

3.2.9.5 Recovery

The recovery was assessed by a post-extraction spike experiment using sample T. The sample blends R1 and R2 were prepared where the sample was added and no SILIS was added. These blends along with typical sample blends (SBs) (that include the SILIS) underwent the sample preparation method. Upon the completion of the SPE steps, the SILIS was added to R1 and R2 blends. The SILIS underwent the evaporation and reconstitution step. Then the ratios of compound:SILIS for the typical sample blends were compared to those of R1 and R2. The calculate percentage difference between the pre- and post-extraction spiked blends was the % recovery for the day.

The recovery was assessed in duplicate on three different days. The average recovery ± the standard deviation of the three recovery estimates was assigned as the overall recovery assessment.

3.2.9.6 Processed samples stability

The processed samples stability was assessed by injecting the same samples that were stored in the autosampler vial at 7 °C on different days and compare the peak areas. To account for the variability of the MS signal an acceptable tolerance of 20 % was accepted given no downward trend was observed in the peak area.

3.2.9.7 Precision

For the assessment of the methods precision three types of samples were used: sample T, QC1 and QC2 described in 3.2.3 and 3.2.8.1 to cover a range of mass fractions and two types of samples; frozen and lyophilised plasma. Each sample was analysed in duplicate in one batch and this was repeated over four weeks. The calculations of the within day repeatability and intermediate precision are described below.

3.2.9.7.1 Repeatability (within-day)

The reported measured was the mean mass fraction of five repeat LC-MS/MS analysis of the same extracts (i.e. five LC-MS/MS injections) which gave an indication of the within-day repeatability. Additionally, the repeatability was calculated by a single factor analysis of variance (ANOVA) statistical tool by taking the square root of the within group mean squares. Because ANOVA was used for evaluating the intermediate precision using the between group variation (day-to-day) the statistical analysis could also be used to evaluate the within group (within day) variation which is the repeatability.

3.2.9.7.2 Intermediate precision (day-to-day)

The intermediate precision to assess the variability between day to day without the within day variability was calculated by the equation below from the single factor ANOVA data.

Equation 3.1 Combined Uncertainty

$$IP = \sqrt{\frac{MS_{bg} - MS_{ig}}{n_d}}$$

Where:

IP	intermediate precision
MS_{bg}	Mean squares of all measurements on all days
MS_{ig}	Mean squares of measurements within the same day
n_d	number of measurements in a day

3.2.10 Measurement equation and uncertainty calculation

See section 2.2.4 in addition to the top-down standard measurement uncertainty was calculated using the equation below based on the precision data (13).

Equation 3.2 Intermediate Precision

$$u_{combined}(w_X) = \sqrt{\frac{IP}{n_d} + \frac{Rep}{n_d \times D}}$$

Where:

$u_{combined}(w_x)$	Combined standard uncertainty
IP	Intermediate precision
n_d	number of measurements in a day
Rep	within day repeatability
D	number of days

Standard uncertainties were converted to expanded uncertainties (95 % confidence interval) using a coverage factor. See 1.9.3 for coverage factor definition.

3.3 Results and Discussion

In the absence of method validation guidelines specific to higher-order RMP, the ICH guidelines were adopted. However, some additional considerations were important even when using these guidelines including: a) the use of a primary calibrator that is traceable to the SI unit; b) the assessment of SILIS and standards interferences and contribution; c) SILIS equilibration in sample; d) assessment of the compound:SILIS ratio; and e) the assessment of measurement uncertainty, etc. These additional considerations will be dealt with in turn and specified where appropriate in the following sections in order to satisfy higher-order metrological requirements.

3.3.1 Accuracy

The accuracy of the plasma METs RMP was assessed by assigning accepted true values of mass fractions using a primary calibrator that is traceable to the SI unit and by assigning measurement uncertainty to measurements. This was due to the absence of commercially available high-order traceable reference materials of METs in plasma at the time of conducting this work.

3.3.1.1 The primary traceable calibrator

The absence of commercially available high-order traceable reference materials of METs was not only in plasma but also in the pure solid form. Therefore, to have an accurate calibration standard, a traceable purity value of the solid standards was required. The manufacturers of the available solid standards provided a certificate of analysis of the ‘pure’ standards. However, these certificates of analysis certified the purity of the compound without using

higher-order reference measurements of the purity and with no traceability (see Appendix 7.1). Therefore, after the purchase of the solid standards, these standards were sent to the UKAS accredited purity analysis team at the NML (LGC, Teddington, UK) to be certified, by a traceable primary reference method (qNMR) before they were fit to be used (see Appendix 7.1). The purity value assigned to these standards is critical because it determines the mass fractions of the calibration standards in methanol prepared from these solid standards and every standard in methanol prepared subsequently. Moreover, a traceable purity value is essential to obtain traceability of the measurement along with the gravimetric preparation process.

Table 3.1, summarises the traceable purity values reported by the NML purity team against the purity values stated in the manufacturer's certificate of analysis. Full details could be found in the certificates of analysis and the analysis performed by qNMR are in Appendix 7.1. The traceable purity values and their measurement uncertainties were used in the mass fraction assigned to stock standards in solvent that were prepared gravimetrically. Subsequently, these values linked directly to the final measurement of the compounds in the plasma samples.

Table 3.1 Solid standards traceable purity reference values.

Solid standard	Manufacturer's certified purity (as salt)[+]	NML traceable purity value[a] (as free base)
Normetanephrine	≥ 98 %	82.09 % m/m ± 0.61 %*
Metanephrine	98.38 %	81.49 % m/m ± 0.63 %**

[+] purity assigned by HPLC
* at the 95 % confidence interval (k= 1.98)
** at the 95 % confidence interval (k= 2.01)
[a] for more details of the analysis see Appendix 7.1

The standards that were used to prepare the calibration blends are the cornerstone of the measurement. This is because they provide the accuracy and contribute to the traceability to the SI unit. Therefore, assessing these standards and their preparation as described below was crucial to the measurement's accuracy.

3.3.1.2 Assessing the reliability of preparation of standards in methanol

The gravimetric preparation of standards in solvent described in section 4.2.4.1 shows how the use of high accuracy balances is important for both accurate preparation and for maintaining the traceability to the SI. However, validating these standards and the preparation process is important to maintain the integrity of the standards. Experimental details of this study are described in 3.2.9.1.2.

For Met the two standards agreed with a percentage difference of 0.60 % between the measured mass fraction and the gravimetric preparation mass fraction, see Figure 3.3.C. This showed that the preparation of the standards was sufficiently accurate and agreed regardless of the variation that could be introduced by preparation by different analysts on different days. However, for Normet the first standards comparison did not agree; with the measured mass fractions being 2.91 % higher than the gravimetric mass fraction, see Figure 3.3.A. A series of experiments was therefore performed to check where the error had occurred whether during the preparation of the diluted standards in solvent or at the stock standard preparation level. After several comparison studies, it was found that the error was at the stock standard in solvent preparation stage, where the preparation of a new fresh stock standard and

its subsequent dilutions by a third analyst was necessary. The comparison study results of the preparation of three stock standards by three different analysts showed that two of the standards agreed confirming that the error was in the preparation of one stock standard, see Figure 3.3.B. This would result in having a lower mass of powder than what was dissolved in the solvent hence when that standard was used as a calibration standard it could result in higher measured mass fraction. The standards in solvent validation study proved the necessity of such a study. Although the process could work accurately, like in Met, the space for human variability would always be there (as shown in Normet). Hence, there always needs to be procedures and checks in place to capture that potential variability and correct it. This would eliminate any human error in the calibration standards preparation and ultimately ensure the accuracy of the measurement.

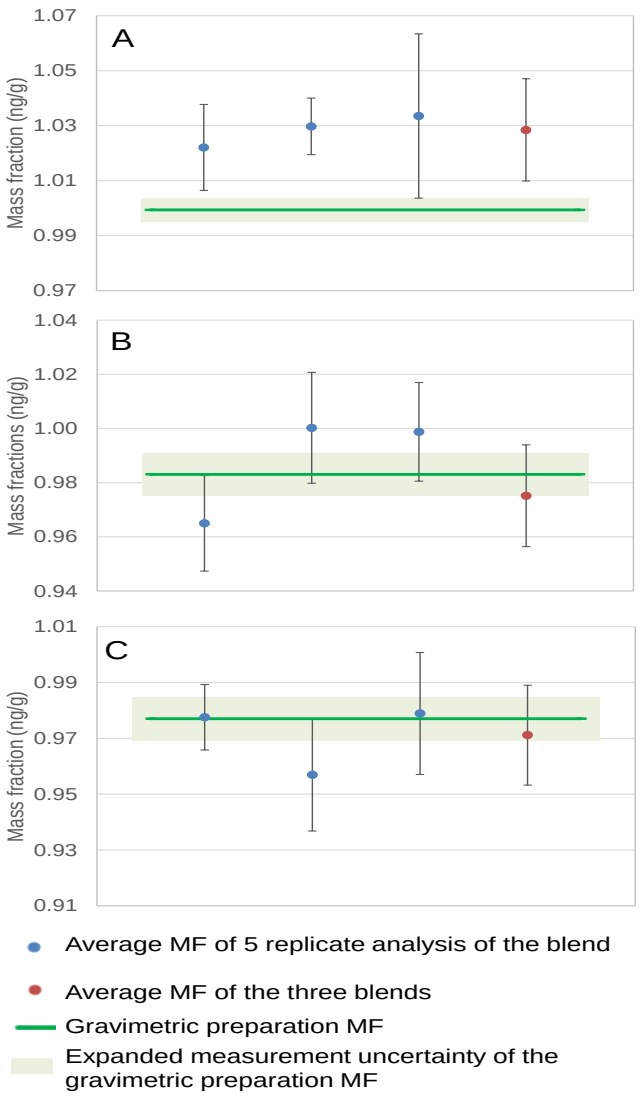


Figure 3.3 Standards in methanol comparison studies results: A) failed Normet standard in solvent comparison, 2.91 % difference between measured mass fraction (MF) and gravimetric MF. B) successful Normet standard in solvent comparison, 0.5 % difference between measured MF and gravimetric MF. C) successful Met standard in solvent comparison, 0.61 % difference between measured MF and gravimetric MF.

3.3.2 Specificity

Analytical selectivity is “the extent to which the method can be used to determine particular analytes in mixtures or matrices without interferences from other components of similar behaviour” (4,14). Selectivity is the recommended IUPAC term and the term used by the FDA however some industries (e.g. pharmaceuticals) and some guidelines (e.g. ICH guidelines)

use the term ‘specificity’ or ‘analytical specificity’ (3). In the work described here the selectivity was assessed by several tools including: a) the assessment of interferences and contribution in the SILIS and the calibration standards; b) the assessment of compound:SILIS ratios; c) the analysis of blank reagents and blank samples; and d) comparing quantification and confirmation SRMs.

3.3.2.1 SILIS interferences and contribution

The measurement of quantity of compound is based on the measurement of the ratio of the ratios R'_B and R'_{BC}; where the former is the measured ratio of the response of the compound to the response of its SILIS in the sample blend and R'_{BC} is the measured ratio of the calibration blend. Therefore, it is crucial that the monitored SRM channels are free from interference, especially that of the compound in the SILIS SRM channels. This is referred to as the ‘contribution of natural into SILIS SRM’. Such contribution would bias the results. The first step which was performed at the method development stage and again in the method performance testing stage was checking for the presence of the natural compound in the SILIS. This was performed by two experiments of MS infusion and LC-MS injections of the SILIS standards in pure solvent. Figure 3.4 demonstrates that the SILIS is pure from the unlabelled natural compound and that the compound standards do not contain an impurity that would interfere with the SILIS precursor ion.

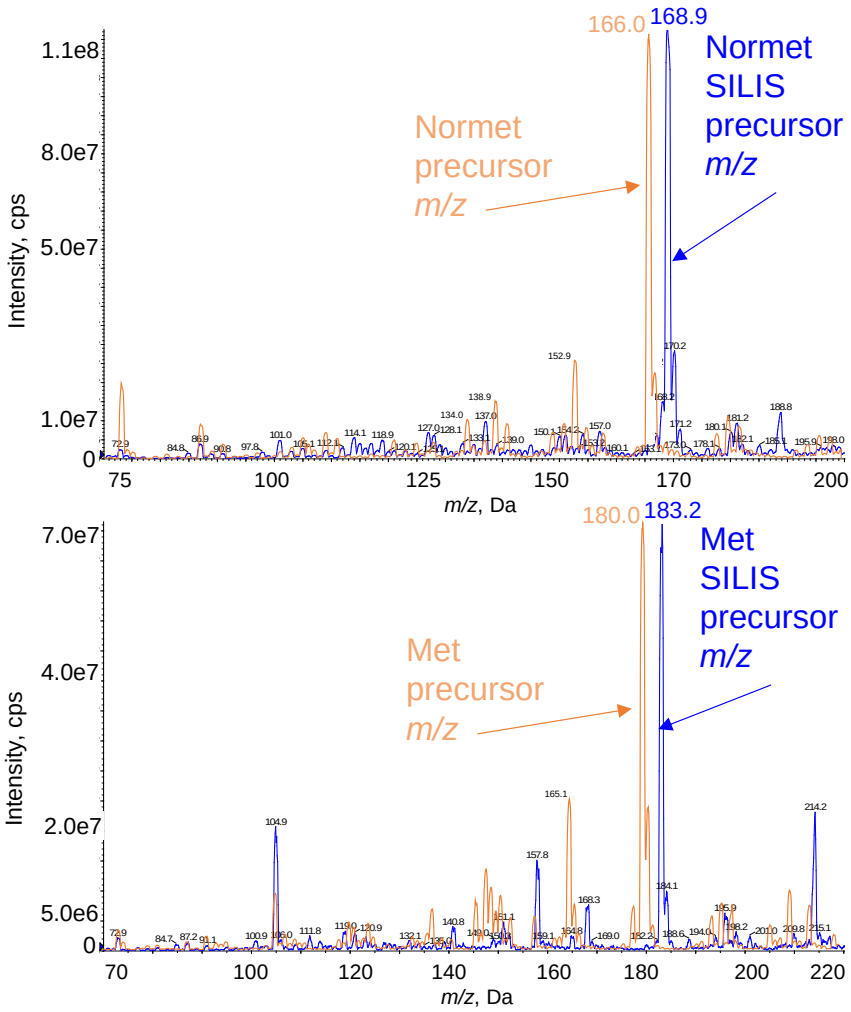


Figure 3.4 Overlayed two MS full scan spectra of Normet and Met and their respective SILIS showing no masses of the compound were present in the SILIS and vice versa. The spectra were generated by direct infusion experiments of the standards in methanol.

3.3.2.2 Other selectivity assessment

In addition to ensuring the compound:SILIS ratios are one ±10 %, the selectivity was also assessed by analysing blanks. All the blanks tested (described in 3.2.8.2) were found to be free of interference. Additionally, to assess the presence of interferences in the samples, a comparison of the quantitative and confirmation SRMs ratio was performed by comparing the

ratio in standards in methanol to those in samples. All ratios matched and no interferences were detected.

Generally, the measurement of an accurate ratio of a SI traceable pure standard to the SILIS under the DEM-IDMS measurement conditions is an effective tool to identify when an interference comes up in the process. This is because the preparation of the calibration blends is done to achieve a ratio of one ±10 %. The ±10 % criteria is an accordance with the isotope dilution mass spectrometry standards of operations procedure at the NML to work under the ISO 17025 accreditation. This tolerance was assigned to take into account any instrument response variability. The example described in the earlier chapter when an interference was suddenly present in the SPE plates reflects the effectiveness of the ratios to monitor specificity along with the blank blends analysis with every batch.

3.3.3 Calibration model

The calibration model typically used in routine methods is linear calibration with internal standardisation. Therefore, typically the assessment of the calibration is done by assessing the linearity of the calibration standards. However, for the candidate RMP of plasma METs, the calibration model was DEM-IDMS calibration (described in 1.9.3) which is a double single-point calibration where the calibration blend (CB) is injected twice before and after the sample blend (this is also called ‘tight bracketing with CB’) and the mean ratio of the two CB injections is used for the ratio of ratio of the sample to calibration blend. Hence, the critical aspect of the DEM-IDMS calibration is the ratio and is not based on a linear calibration, therefore this needs to be

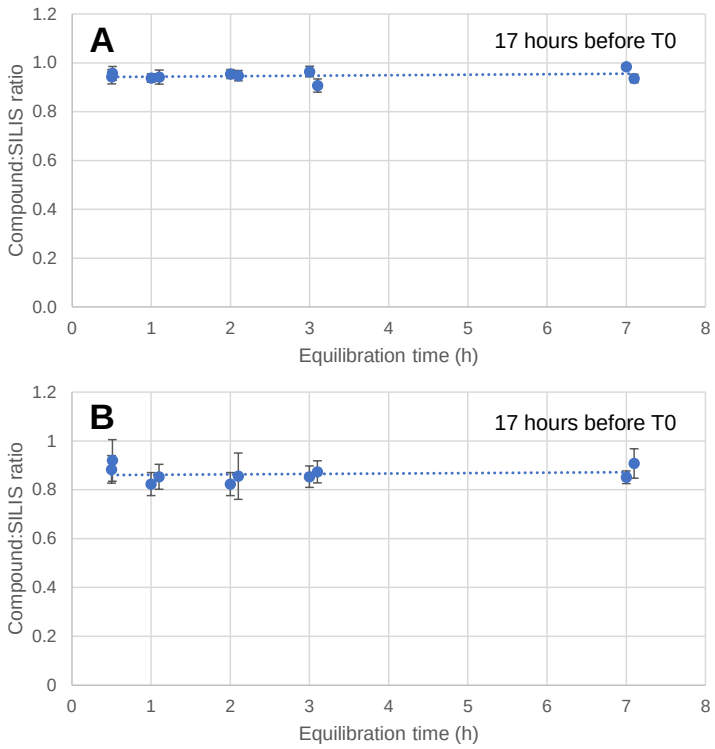
treated differently to traditional method validation. In addition to the interferences and purity assessment described earlier, to validate the accuracy of the ratio, another key element to assess was the equilibration of the SILIS in the sample and calibration blends.

In addition to the SILIS equilibration study described below, the DEM-IDMS calibration was assessed by monitoring the compound:SILIS ratios of the calibration blends and sample blends. The acceptable criteria were for the ratios to be equal to one ±10 %, any higher or lower ratios required investigation to check for interferences.

3.3.3.1 SILIS equilibration study

As mentioned earlier the measured ratio of responses of the natural compound to its SILIS is the cornerstone for the DEM-IDMS measurements. These ratios were measured on the premise that the SILIS behaved similarly to the natural compound and was added to the sample and calibration blends at the earliest step of sample preparation; the blends preparation. Therefore, as part of assessing the accuracy of the measurements and the reliability of the calibration model, it was important to evaluate the SILIS equilibration to the sample, see section 0. The results of the second equilibration study shown in Figure 3.5 demonstrated that the ratios of the natural compound to its SILIS were consistent regardless of the equilibration time for both Normet and Met. This indicated that the behaviour of the SILIS in the human plasma was similar to that of the natural unlabelled compounds, Normet and Met. Additionally, it indicated that if the sample preparation method were to be performed all in one day or on two consecutive days the results would not be affected. The

results showed that a minimal equilibration time for the SILIS with the material would be sufficient.



Error bars = standard deviation of five replicate analysis

Figure 3.5 SILIS equilibration study 2, the plotted data points are the average ratios of the response of the compound to its corresponding SILIS of five replicate analysis. Data points at 17 hours was plotted at 7 hours to improve visibility of the data points on the chart.

3.3.3.2 Measurement range

The measurement range is defined as '*the interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy and linearity*' (3). However, for higher-order RMPs that use DEM-IDMS calibration applied to a traceable reference material, linearity is less important. Here, the measurement range was assessed experimentally using the Level I and

Level II QC materials within which the reference value assignment should lie. The measurement uncertainty's estimate of the method was initially found to increase when measuring the Level II QC that is at the pathological level with higher mass fractions. The peaks were tailing and eluting later due to overloading of the LC column. This resulted in poor peak shape and subsequently inaccurate peak integrations resulting in higher measurement uncertainties. When the injection volume for these samples was lowered from 8 µL to 4 µL the performance improved and was similar to that of the Level I QC and sample T. When the 4 µL injection volume was used for the Level I QC samples it was found that the measurement uncertainty doubled.

Therefore, the candidate RMP of plasma METs was estimated to measure Normet at 76.3 - 260 pg/g with an average expanded measurement uncertainty of 5.92 % and up to 10.5 %. For Met, the candidate RMP was estimated to measure at 19.7 - 100 pg/g with an average expanded measurement uncertainty of 9.38 % and up to 15.3 %. Measurements outside these ranges would result in larger measurement uncertainty estimates. Whilst this is useful, a more in-depth characterisation of range would be beneficial and could be explored as part of future work. However, it is not standard protocol to define the range for a higher-order RMP that measures a specific sample at a specific mass fraction and so was not considered further here.

3.3.4 Limit of detection and limit of quantification

The technical LOD were estimated at 1.0 pg/g and 1.5 pg/g for Normet and Met, respectively. The technical LLOQ was estimated at 3.2 pg/g and 4.6 pg/g for Normet and Met, respectively.

It is very important to note that the LOD and LLOQ are not method parameters that would be relevant for the use of the candidate RMP because the method is designed to measure with high accuracy and at the smallest measurement uncertainty estimates possible. Measuring at the LLOQ would result in much higher uncertainties. All measurements acquired (shown in Figure 3.6) had a minimum $S:N_{p/p}$ of 40:1 and 100:1 for Met and Normet, respectively. Therefore, in line with ICH guidelines, this method can be applied for assigning reference values (i.e., $S:N_{p/p} > 10:1$) (3).

3.3.5 Recovery and processed samples stability

The recovery was assessed on three different days and in duplicate on each day, see 3.2.9.4 for experimental details. The recovery for Normet and Met was measured at 78 ± 2 % and 78 ± 6 %, respectively, across all days. This was acceptable based on the assessment of the measurement uncertainty described later. If the measurement uncertainty estimate achieved was larger and needed reduction, further work to improve the recovery and obtain more signal would have been needed.

For the processed samples stability, the peak areas of Normet and Met were found to be within 20 % of the freshly prepared samples up to five days which was sufficient as the analytical sequence was 4.5 days.

3.3.6 Precision

Precision is measured in different ways under different guidelines. Under the ICH guidelines, precision is *"the assessment process of the degree of dispersion of measurements acquired from multiple samples of a parent homogenous sample under specified conditions"*. The ICH guidelines consider precision at three levels: repeatability, intermediate precision and reproducibility (3). Repeatability '*expresses the precision under the same operating conditions over a short interval of time. Repeatability is also termed intra-assay precision*' (3). In this work, the within-day repeatability was calculated and shown in Table 3.2. The repeatability is also indicated by the standard measurement uncertainty of each measurement as it is the standard deviation of the five replicate measurements by LC-MS, see Table 3.3.

The intermediate precision expresses the precision within-laboratories variations (3), here the day-to-day variability was assessed for intermediate precision and is shown in Table 3.2. Reproducibility expresses the precision between laboratories (3), this is not described in this thesis and will take part in future work as conversations with other national measurements institutes to start an interlaboratory study have commenced.

The precision data are summarised in Table 3.2, the individual measurements and their expanded measurement uncertainties of each sample across the different batches are shown in Figure 3.6. With the exception to one sample data set, the intermediate precision was larger than the within-day repeatability. This indicated that there was a day effect on the method performance. If the measurement uncertainty estimate of the RMP was found

not fit-for-purpose, more work would have been required to improve the precision of the method. To begin with, a larger data set across longer period would be required to have a more accurate estimate of the day-to-day variability and its impact on the measurement uncertainty. However, the measurement uncertainty of the RMP was found to be fit-for-purpose as described later in 3.3.7.3.

Table 3.2 Precision data of the candidate RMP of plasma METs

Normet Precision						
Sample	Days	Mean MF (pg/g)	Within day repeatability (pg/g)	Within day repeatability (%)	Intermediate precision (pg/g)	Intermediate precision (%)
T	4	77.65	1.37	1.76	1.60	2.06
QC1	4	106.5	1.06	1.00	5.10	4.76
QC2	4	517.4	5.56	1.08	10.6	2.05
Average				1.28		2.96
Met Precision						
Sample	Days	Mean MF (pg/g)	Within day repeatability (pg/g)	Within day repeatability (%)	Intermediate precision (pg/g)	Intermediate precision (%)
T	4	22.97	1.37	5.97	1.07	4.67
QC1	4	19.69	0.65	3.28	1.45	7.37
QC2	4	201.8	6.20	3.07	6.64	3.29
Average				4.11		5.11

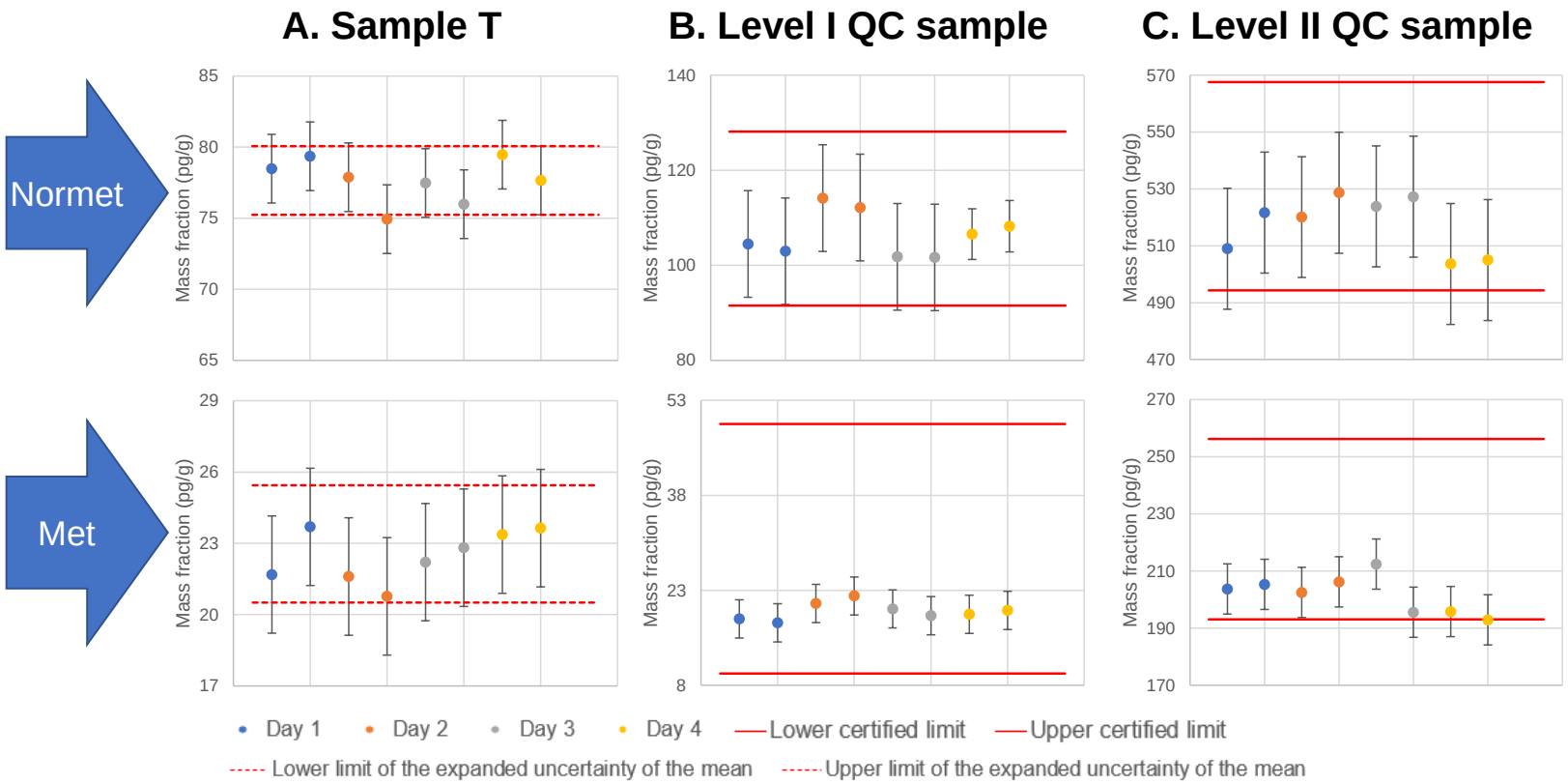


Figure 3.6 Analysis of samples over different weeks. A) Sample T measurements, B) Level I QC measurements, C) Level II QC measurements. Each data point is the mean mass fraction of five LC-MS analysis blends per sample vial. The error bars are the expanded measurement uncertainty at the 95 % confidence interval (coverage factors k = 3.18).

3.3.7 Measurement uncertainty and sources of variability

Measurement uncertainty in metrology is different to the closest variable in the ICH guidelines which is accuracy. In this work, the measurement uncertainty of the measurement and the sources of uncertainty were identified and quantified. This is critical for the assessment of higher-order RMPs that are designed to ultimately assign traceable reference values with the lowest measurement uncertainty.

3.3.7.1 Sources of variability in higher-order reference measurement

A clearer and more in depth understanding of the sources of variability in higher-order reference measurements of these small molecules in plasma was obtained through this work. The key sources of variability that were found to be evident in the measurement process are summarised in the fishbone diagram, Figure 3.7. The different factors described accumulatively increase the variability in measurement however, some factors have higher impact on the measurement uncertainty budget while others were interlinked and were difficult to individually quantify. For example, the issue of chromatographic robustness was influenced by several factors among them: a) the sample preparation and the necessity to improve the SPE method by adding a second SPE step to improve the chromatography and reduce variability; b) the injection volume in the case of disease level and; c) the LC system itself. Throughout the work described in chapters 2 and 3 the LC system suffered numerous faults that required days to be resolved by the analyst or an engineer. Another source of variability was found to be the MS instrument condition where the instrument performance improved, and the ratio of ratios

variability was reduced by cleaning the MS instrument parts e.g., the cone and the ion optics. Such factors of variability could not be quantified, nevertheless were crucial to resolve in order to achieve the performance required and were investigated prior to measurement uncertainty characterisation.

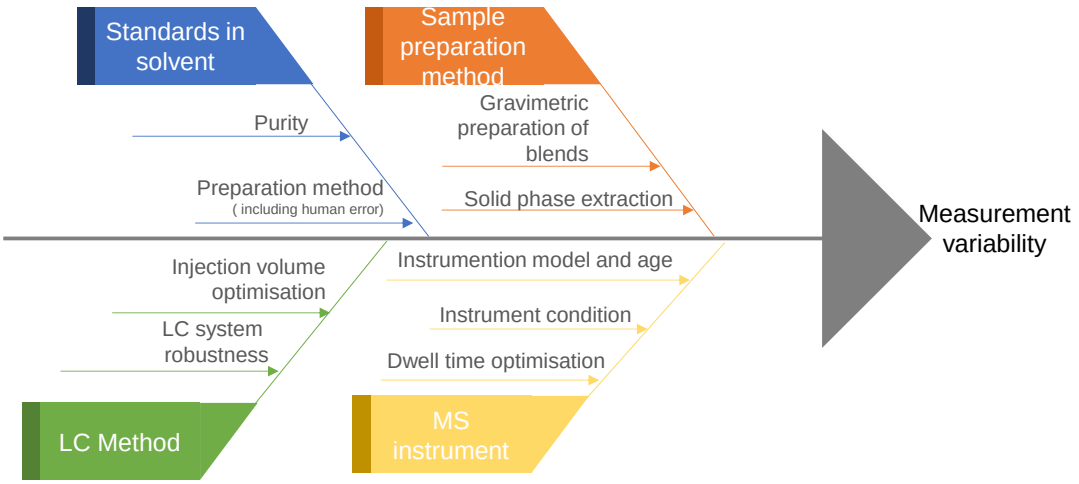


Figure 3.7 Fishbone diagram of the sources of variability in measurement

3.3.7.2 The measurement uncertainty budget of a reference measurement: the "bottom-up" approach

The fishbone diagram in Figure 3.8 does not reflect to what extent each of these factors contribute to the measurement uncertainty. However, the measurement uncertainty budget of a single measurement was evaluated as an example and the contribution of each source of variability in the measurement was calculated. This approach to measurement uncertainty is the "bottom-up" approach where every individual component to the uncertainty from the different stages of the method is estimated (15,16). In this approach, each component of the uncertainty is systematically evaluated and measured. These sources of uncertainties all contributed to the final uncertainty using the combined uncertainty equations (see Equation 2.3) (16–19).

The measurement uncertainty budget components stem from the DEM-IDMS mass fraction equation and its corresponding combined measurement uncertainty equation (see Equation 2.2). The three directly influencing components that are part of the uncertainty equation are: a) the calibration standard mass fraction, which is influenced by its purity and its gravimetric preparation; b) the gravimetric preparation of sample and calibration blends and; c) the instrument variability captured by the variation of the DEM-IDMS ratio of ratios. The individual uncertainty budgets of each of the other measurements performed of all samples demonstrated a similar trend in the budget. Overall, the largest component of the measurement uncertainty budget of a measurement on the same day was found to be the instrument variability which was up to 67 % and 98 % of the total uncertainty for Normet and Met, respectively. The second largest component was the standard in solvent at 31.4 % and 1.9 % for Normet and Met, respectively, see example data of a measurement for sample T in Table 3.3. The high proportion of the measurement uncertainty budget arising from instrumental variability demonstrates the importance of trying several different instruments in order to minimise this uncertainty. In this thesis a total of eight instruments were evaluated across several partnering sites (e.g., NML, King's College London, University of Southampton, Waters® UK Headquarters), the comparison data of four of these instruments is shown in Figure 2.10. This was a rare opportunity to evaluate the latest and best instruments for this project.

Table 3.3 Measurement uncertainty budget, using example data of a T sample measurement.

Normet						
Variables category	Variable	Mass fraction (pg/g)	u (pg/g)	Derivative	Variance	**Uncertainty budget (%)**
Calibration standard (pg/g)	Calibration standard mass fraction w_z (pg/g)	250.0	2.004	0.0080	6.43E-05	**31.42**
Gravimetric preparation (g)	Mass of sample in SB m_x (g)	0.513	0.00017	0.0003	1.10E-07	**0.05**
	Mass of SILIS in SB m_y (g)	0.149	0.00017	0.0011	1.31E-06	**0.64**
	Mass of sample in CB m_{zc} (g)	0.149	0.00017	0.0011	1.30E-06	**0.64**
	Mass of SILIS in CB m_{yc} (g)	0.149	0.00017	0.0011	1.31E-06	**0.64**
Instrumental variability (R'_B/R'_{Bc})	Instrumental variability (R'_B/R'_{Bc})	1.070	0.0125	0.0117	1.36E-04	**66.61**

Table continues on the following page

Met						
Variables category	Variable	Mass fraction (pg/g)	u (pg/g)	Derivative	Variance	**Uncertainty budget (%)**
Calibration standard (pg/g)	Calibration standard mass fraction w_z (pg/g)	74.61	0.591	0.0079	6.28E-05	**1.89**
Gravimetric preparation (g)	Mass of sample in SB m_x (g)	0.512	0.00017	0.0003	1.10E-07	**0.003**
	Mass of SILIS in SB m_y (g)	0.149	0.00017	0.0011	1.31E-06	**0.039**
	Mass of sample in CB m_{zc} (g)	0.149	0.00017	0.0011	1.31E-06	**0.039**
	Mass of SILIS in CB m_{yc} (g)	0.149	0.00017	0.0011	1.30E-06	**0.039**
Instrumental variability (R'_B/R'_{Bc})	Instrumental variability (R'_B/R'_{Bc})	1.028	0.058	0.0570	3.25E-03	**97.99**

3.3.7.3 The measurement uncertainty estimate of the RMP; the "top-down" approach

There is a distinct difference between a measurement uncertainty budget of a reference value of a sample or a CRM and a measurement uncertainty budget of a candidate RMP. Chapter 4 illustrates the measurement uncertainty budget of a candidate CRM and shows how the uncertainty components would be directly related to the material itself in addition to the RMP used to characterise it. Here, the estimation of the measurement uncertainty of the candidate RMP of plasma METs is described.

It was reported that when comparing the "bottom-up" and the "top-down" approaches of estimating the measurement uncertainties, no statistically significant differences were found (15,18). Therefore, the top-down approach was selected to estimate the initial measurement uncertainty of the candidate RMP of plasma METs. Typically, the top-down approach to estimating the measurement uncertainty evaluates the quality control samples data and/or method validation experiments data (15,20). This approach is usually more straight forward, cost-effective and more practical. Additionally, this approach also allows the opportunity to update the measurement uncertainty estimate whenever more data becomes available.

A measurement uncertainty estimation is influenced by the number of replicate measurements (13,20). Hence, the preliminary estimated measurement uncertainty was calculated using the replicate analysis of the three samples over four weeks as described in the precision calculations (see 3.2.9.4 and 3.3.3.2). The top-down approach used statistical equations to calculate the uncertainty based on the dispersion of measurements, the

equations are in 3.2.10. It was observed that intermediate precision was larger than the within-day repeatability. Therefore, the largest component of the overall estimated measurement uncertainty of the candidate RMP was the day-to-day variability of the measurements for both Normet and Met (see Table 3.2). This was observed in all samples analysed, regardless of mass fractions or whether the plasma sample is frozen or lyophilised.

From the current data acquired, it was found that the driving factor was the instrumental variability where the %RSD of the ratio of ratios was variable from week to week. Further work to reduce this variability could be performed such as cleaning the inner instrument parts before each batch of analysis.

The candidate RMP of plasma METs was estimated to have an average expanded relative measurement uncertainty of 5.92 % and 9.38 % for Normet and Met, respectively. This measurement uncertainty was found to be the largest at 10.5 % and 15.3 % for Normet and Met, respectively. These uncertainties are deemed suitable, see Table 3.4. As mentioned earlier (see 3.1.2), in the absence of any agreed guidelines of what measurement uncertainty is fit-for-purpose. The EQAS data acceptable range was used as an assessment tool of the measurement uncertainty. In the example EQAS data set, the acceptable range for a measurement was estimated at 33.7 % and 26.3 % for Normet and Met, respectively. The estimated RMP measurement uncertainty improves upon the acceptable range of the EQAS data even at its highest estimation. The EQAS data is likely more variable due to several potential reasons among them the use of different methods, instruments and calibrators across different participating laboratories. For Normet the highest estimated RMP measurement uncertainty is less than a

third of that of the EQAS acceptable range. For Met, the difference was smaller however still the largest estimated uncertainty for Met (15.3 %) was smaller than the EQAS acceptable range (26.3 %). This means that this RMP could be used to assign reference values to the EQAS samples with added traceability and better precision. This was considered a major success for this project.

Table 3.4 The measurement uncertainty estimate of plasma METs candidate RMP

Normet Measurement Uncertainty (top-down approach)							
Sample	Days	Mean MF (pg/g)	u (pg/g)	df	k*	U (pg/g)	%U
T	4	77.65	0.76	3	3.18	2.41	3.11
QC1	4	106.5	3.53	3	3.18	11.2	10.5
QC2	4	517.4	6.68	3	3.18	21.3	4.11
Average							5.92
Met Measurement Uncertainty (top-down approach)							
Sample	Days	Mean MF (pg/g)	u (pg/g)	df	k*	U (pg/g)	%U
T	4	22.97	0.61	3	3.18	1.95	8.47
QC1	4	19.69	0.95	3	3.18	3.01	15.3
QC2	4	201.8	2.76	3	3.18	8.79	4.35
Average							9.38
*k: coverage factor, two sided at 95% confidence interval							

3.4 Conclusions

The aim of this Chapter was to investigate the sources of variability in higher-order reference measurements by assessing the performance of a candidate higher-order RMP. The performance of the plasma METs candidate RMP was assessed in compliance with ISO 15193:2009 following the method validation guidelines of the ICH (2,3). The candidate RMP was considered suitable for reference value assignment for Normet and Met in plasma. This was based on a systematic performance assessment of accuracy, precision, specificity,

LOD, LOQ, calibration model and recovery. In addition to this, the measurement uncertainty was characterised for this candidate RMP. For both Normet and Met, the largest component of the measurement uncertainty was the instrument variability. This underlines the importance of choosing the most appropriate instrument for the candidate RMP application. A major benefit of this work was the opportunity to evaluate eight instruments in order to better understand their capabilities for this application despite their high specifications.

Therefore, the hypothesis of this chapter holds true in part, as there are no current guidelines to how low the measurement uncertainty needs to be. While the measurement uncertainty achieved was found of an added value to the existing EQAS ranges, the submission of the RMP to the JCTLM database is still pending. This requires a long peer-review process and inter-laboratory comparison study with at least one other national measurement institute. The NML at LGC will embark on in the next round of funding from the UK Department for Business, Energy and Industrial Strategy.

3.5 Novelty and value

The development of a candidate RMP for plasma METs described in this work is the first to be performed at a higher-order reference measurement level. Upon examining the JCTLM database for RMPs of small molecules analysed for clinical analysis, it was found that only three methods measure at such ultra-low levels (<100 pg/g), the analytes of these methods are: estrone 270.36 g/mol, total 17β-estradiol 272.4 g/mol and aldosterone 360.44 g/mol (1).

The knowledge and understanding of the sources of variability described here could simplify the assessment of the sources of uncertainty in developing higher-order RMPs of similar compounds in the future.

Clinically, this method provided lower measurement uncertainty than the accepted range for EQAS samples. Therefore, it is a starting point that will underpin the plasma METs measurements in the UK hospitals and globally through assigning reference values to EQAS samples and higher-order CRMs. When the hospital laboratories have these tools, they would be able to assess their methods and assign measurement uncertainties to their methods using higher-order CRMs. This will improve the quality of the routinely measured results and reduce the rate of false negatives and false positives of the routine methods (21,22). This could ultimately assist the hospital to improve the diagnosis of pheochromocytoma and paraganglioma, improving the patients' quality of care. Additionally, it would relieve the distress a patient lives through and the unnecessary costs of imaging resulting from false positive measurements.

3.6 Future work

This work established the basis of a candidate RMP of plasma METs. Further work is required for future submission to the JCTLM database such as interlaboratory comparison of samples is required with another national measurement institute. Ultimately, the method would also be used to assign reference values to EQAS samples and higher-order CRMs.

3.7 References

1. BIPM - About the BIPM [Internet]. [cited 2018 Aug 13]. Available from: https://www.bipm.org/en/about-us/
2. 14:00-17:00. ISO 15193:2009 [Internet]. ISO. [cited 2020 Apr 3]. Available from: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/04/20/42021.html
3. ICH harmonised tripartite, guideline. Validation of analytical procedures: Text and methodology Q2(R1) [Internet]. www.ich.org; 2005. Available from: https://www.gmp-compliance.org/files/guidemgr/Q2(R1).pdf
4. Magnusson B, Örnemark U, editors. Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics [Internet]. 2nd ed. http://www.eurachem.org/: Eurachem; 2014 [cited 2015 Dec 10]. Available from: https://www.eurachem.org/images/stories/Guides/pdf/MV_guide_2nd_ed_EN.pdf
5. Theodorsson E. Validation and verification of measurement methods in clinical chemistry. Bioanalysis. 2012 Feb;4(3):305–20.
6. FDA. Bioanalytical Method Validation Guidance for Industry. 2018;44.
7. EMEA. Draft guideline validation bionanalytical methods [Internet]. UK; 2009 [cited 2020 Dec 15]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-validation-bionanalytical-methods_en.pdf
8. Barwick V, Prichard E. International vocabulary of metrology -- Basic and general concepts and associated terms (VIM) 3rd ed [Internet]. 2012 [cited 2017 Nov 14]. Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf
9. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, et al. Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. Pharm Res. 2007 Oct;24(10):1962–73.
10. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accreditation Qual Assur. 2007 Jan 1;12(1):45–7.
11. Careri M, Mangia A. Validation and qualification: the fitness for purpose of mass spectrometry-based analytical methods and analytical systems. Anal Bioanal Chem. 2006 Sep;386(1):38–45.
12. Kruve A, Rebane R, Kipper K, Oldekop M-L, Evard H, Herodes K, et al. Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. Anal Chim Acta. 2015 Apr 22;870:29–44.
13. Ellison SLR, Barwick VJ, Farrant TJD. Practical Statistics for the Analytical Scientist [Internet]. 2009 [cited 2020 Dec 11]. Available from: https://pubs.rsc.org/en/content/ebook/978-0-85404-131-2
14. Vessman J, Stefan RI, van Staden JF, Danzer K, Lindner W, Burns DT, et al. Selectivity in analytical chemistry (IUPAC Recommendations 2001). Pure Appl Chem. 2001;73(8):1381–6.
15. Dabalus Islam M, Schweikert Turcu M, Cannavan A. Comparison of methods for the estimation of measurement uncertainty for an analytical

method for sulphonamides. Food Addit Contam Part Chem Anal Control Expo Risk Assess. 2008 Dec;25(12):1439–50.
16. Martinello F, Snoj N, Skitek M, Jerin A. The top-down approach to measurement uncertainty: which formula should we use in laboratory medicine? Biochem Medica [Internet]. 2020 Jun 15 [cited 2020 Dec 7];30(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7138004/
17. Rigo-Bonnin R, Blanco-Font A, Canalias F. Different top-down approaches to estimate measurement uncertainty of whole blood tacrolimus mass concentration values. Clin Biochem. 2018 Jul;57:56–61.
18. Lee JH, Choi J-H, Youn JS, Cha YJ, Song W, Park AJ. Comparison between bottom-up and top-down approaches in the estimation of measurement uncertainty. Clin Chem Lab Med. 2015 Jun;53(7):1025–32.
19. GUM - English - Annex B. General metrological terms [Internet]. [cited 2020 Dec 4]. Available from: https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e_FILES/MAIN_C045315e/AB_e.html
20. ISO/IEC Guide 98-3:2008 - Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) [Internet]. [cited 2017 Nov 14]. Available from: https://www.iso.org/standard/50461.html
21. Eisenhofer G, Peitzsch M. Laboratory Evaluation of Pheochromocytoma and Paraganglioma. Clin Chem. 2014 Jan 12;60(12):1486–99.
22. Eisenhofer G, Peitzsch M, McWhinney BC. Impact of LC-MS/MS on the laboratory diagnosis of catecholamine-producing tumors. TrAC Trends Anal Chem. 2016 Nov 1;84(Part B):106–16.

4 Understanding measurement uncertainty in higher-order reference material characterisation: nonpolar small molecules in whole blood samples

4.1 Introduction

While the previous two chapters investigated the sources of variability in the process of higher-order reference method development and method validation, this chapter investigated the sources of variability when using an existing candidate higher-order reference measurement procedure (RMP) for the characterisation of higher-order **candidate** certified reference materials (cCRM), see Figure 4.1. Reference materials that have not been fully certified are called candidate reference materials instead of **certified** reference materials (CRM) which would be fully characterised and certified.

These materials may not be challenging to produce but they require complex analytical work and detailed understanding of the sources of variability in the measurement process to achieve accurate traceable measurement of the lowest measurement uncertainty estimate possible. Once the analytical work to characterise and certify the cCRMs is completed, they would be submitted to the Joint Committee for Traceability in Laboratory Medicine and Clinical Chemistry (JCTLM) database to become higher-order CRMs. These materials would then be used to obtain traceability to the SI unit (see Figure 4.1). Higher-order CRMs would be used by secondary standards manufacturers to provide them with traceability and feed into their measurement uncertainty estimate.

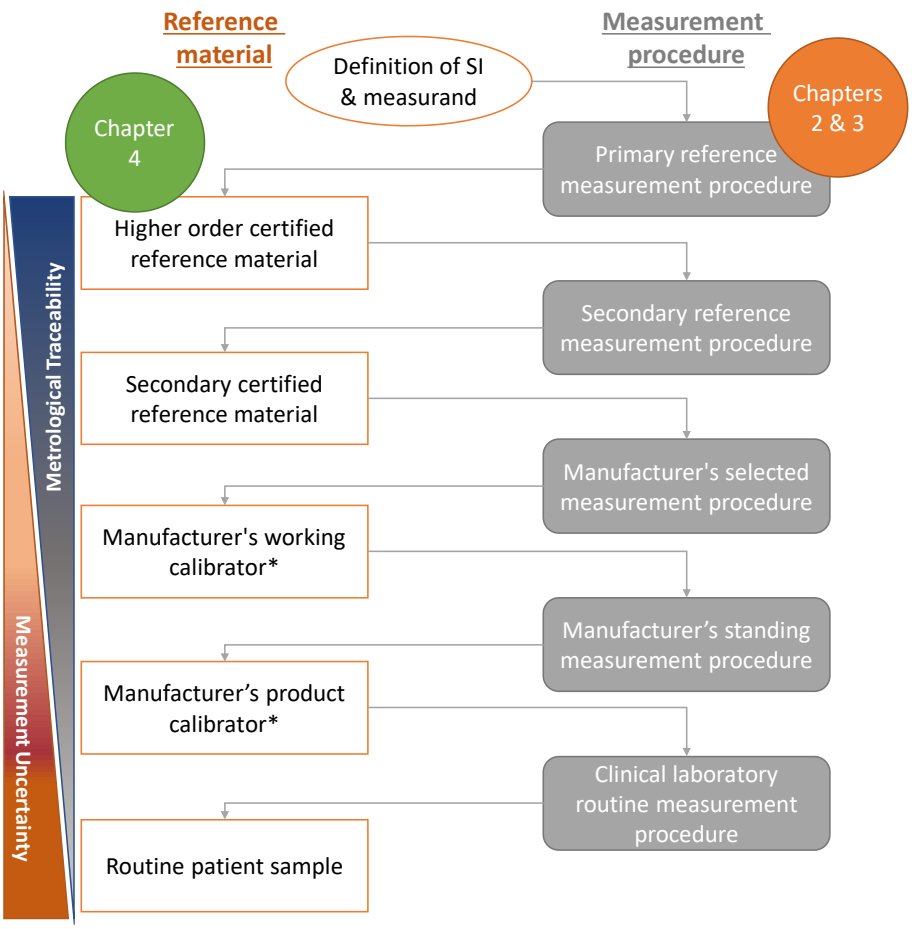
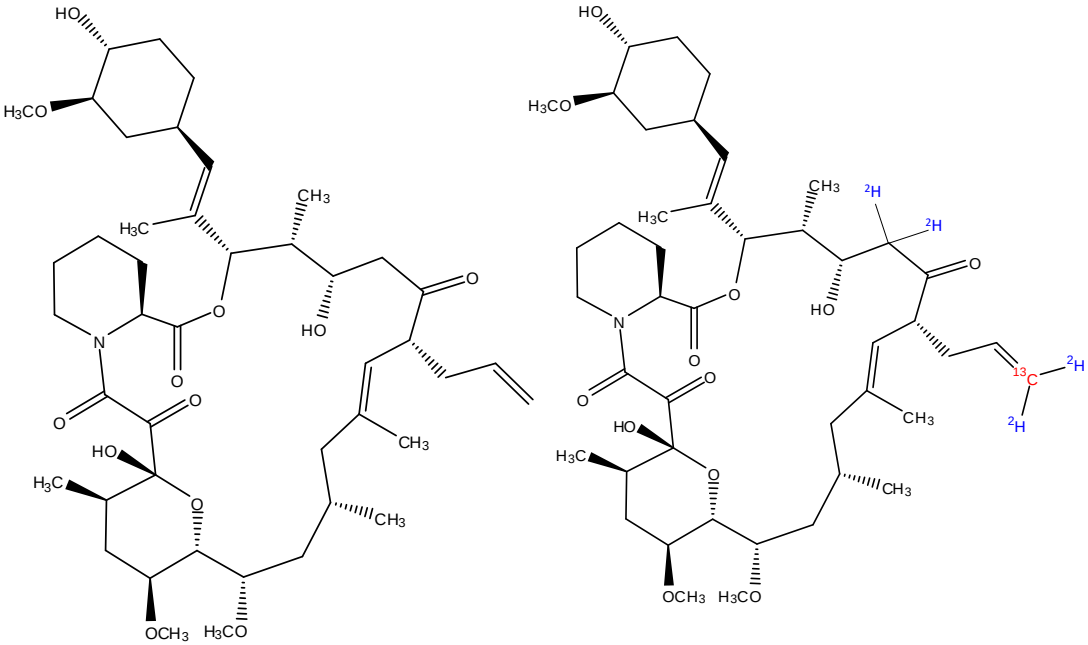


Figure 4.1 Situating this PhD thesis work on the metrological traceability chain diagram. The diagram demonstrating the link between the top SI units to the reported patient result at the end of the chain by a series of reference value assignments using calibrated measurement procedures. Adapted from (1)

This chapter investigates the sources of variability when using the same LC-MS with double exact-matched isotope dilution mass spectrometry (DEM-IDMS) technique for the analysis of tacrolimus, a molecule with a different chemical nature from that of metanephrines (METs). Contrary to METs, tacrolimus is a relatively large exogenous non-polar highly lipophilic molecule that is analysed in whole blood samples, see Figure 4.2 for structure and molecule information. Therefore, tacrolimus was selected as the second case study application to investigate sources variability in higher-order reference measurements in clinical applications.

Furthermore, assessing the variability in the process of characterising a higher-order cCRM encompassed thorough experiments to understand the material and identify potential sources of variability (see Experimental design 4.2.8). Many of these experiments were not performed as part of the plasma METs method development investigation. Should the candidate higher-order reference measurement procedure of plasma METs described in Chapter 2 be used for the characterisation of a plasma METs higher-order cCRM, then these experiments would be performed and the sources of variability investigated.

In this study, an investigation to understand the sources of variability and measurement uncertainty in reference measurement of nonpolar larger size small molecules in higher-order cCRM in patient blood was performed. The cCRM used to conduct this work was of the immunosuppressant tacrolimus which is an exogenous highly nonpolar compound (structure in Figure 4.2) that is bound strongly to erythrocytes (2–4). Hence, tacrolimus is analysed in whole blood samples rather than in plasma or serum (2–4). The analysis of whole blood samples introduced different challenges in the analysis process from those of plasma METs due to the complexity of whole blood that contains cells and other components.



Tacrolimus
$C_{44}H_{69}NO_{12}$

MW	804.02 g/mol
LogP	3.19
pK_a	9.96 (acidic)
pK_a	-2.9 (basic)

$^{13}C^{2}H_4$-Tacrolimus
$C_{43}{}^{13}CH_{67}D_4NO_{12}$

MW 809.02 g/mol
Isotopically-labelled
internal standard

Figure 4.2 Tacrolimus chemical structure of tacrolimus (left) and isotopically-labelled $^{13}C^{2}H_4$-tacrolimus structure (right). Tacrolimus formula, LogP and pK_a values were obtained from the Metabolic Innovation Centre Toxic Exposome Database (5). The structure of the isotopically-labelled tacrolimus was from the certificate of analysis of standards used in this research (see 4.2.2).

Therapeutic drug monitoring of tacrolimus was selected as the clinical application for this research, not only due to the chemical and analytical reasons mentioned above but also due to its clinical significance. This clinical test was selected also to build on previous efforts to standardise it (6,7). Section 1.12 in the first chapter covered an overview of the efforts made in the standardisation of tacrolimus in whole blood analysis however there seems still to be a spread in the results reported in external quality assurance schemes (EQAS) and proficiency testing schemes (PTS).

Here, an existing higher-order reference method (6) was used to produce a higher-order certified reference material of tacrolimus in pooled patient blood. This will enable the investigation of sources of variability that could be associated to the type of matrix where the compound was incurred in the blood through the patient administered with the medication and metabolising it and so on. This resembles a normal patient sample to a certain degree as it contained the metabolites of tacrolimus and differs from the existing higher-order CRM available which is tacrolimus spiked in human blood from healthy individuals.

Currently, tacrolimus is analysed in hospital laboratories using immunoassay and LC-MS based methods, immunoassay based methods have been reported to suffer from interferences and cross reactivity with endogenous molecules as well as tacrolimus metabolites (8–13). Therefore, producing this pooled patient blood material could provide an important tool for the hospital laboratories and immunoassay analytical kit manufacturers to assess their methods.

Higher-order RMPs measure mass fractions of mass per mass (ng/g) instead of the more commonly reported mass concentration of mass per volume (ng/mL). Higher-order RMPs include gravimetric preparation of the samples and produce measurements that would be traceable to the SI unit through a chain of calibration that is mainly gravimetrically bound (14). Hence, to facilitate the use of the material by end users (e.g., secondary standard manufacturers or hospital laboratories), the mass concentration of the material needed to be reported on the certificate. In this chapter, in addition to the quantification of tacrolimus in the material using the higher-order reference

measurement procedure by LC-MS, the measurement of the density of the material was performed to enable conversion from mass fraction to mass concentration.

Tacrolimus is an immunosuppressant drug (ISD) classified as a calcineurin inhibitor and is used mainly in solid organ graft transplant recipients for the prevention of allograft rejection (6). Like other immunosuppressants, it is crucial to individualize each patient's drug regimen to achieve successful organ transplantation and function (see section 1.3.2). Tacrolimus is subject to a high degree of inter-individual pharmacokinetics and hence to achieve the same blood concentration in patients, variable doses are used (2,3,3). Sub-optimal blood drug concentrations can cause either organ rejection or a range of adverse toxicity as a result of supratherapeutic drug concentrations such as kidney failure, see Table 4.1. Hence, therapeutic drug monitoring (TDM) of tacrolimus is essential as the oral dose does not indicate drug efficacy (4,15). TDM of tacrolimus is typically performed by measuring blood concentrations as stated in the National Institute for Health and Care Excellence (NICE) guidelines which are the evidence based recommendation for health and patient care that are followed in England (16).

Table 4.1 Overview of tacrolimus TDM to improve graft survival while reducing risk of drug toxicity. Adapted from (17).

Tacrolimus trough concentration	Clinical recommendation
< 5 ng/mL	Subtherapeutic level: consider dose increase to reduce risk of a rejection episode
5-10 ng/mL	Therapeutic index: maintain same dose
> 15 ng/mL	Supratherapeutic level: consider dose reduction to reduce risk of toxicity

4.1.1 Hypothesis and objectives

In line with the aim of this thesis expressed in Chapter 1 (section 1.13.2.1), the hypothesis for this chapter was that; “all the major sources of measurement uncertainty could be reliably determined for the measurements of clinically relevant small molecules (using tacrolimus as the model compound) at the higher end of the molecular weight range”.

In order to test this hypothesis, the following objectives were set:

- Identify the sources of measurement uncertainty in the characterisation of a cCRM.
- Assign a traceable reference value to the cCRM with a low measurement uncertainty.
- Conduct the analytical work required for the certification of a higher-order cCRM to be submitted to the JCTLM-database including

characterisation study, homogeneity study, stability study and density study.

4.2 Experimental

4.2.1 Candidate reference material production

The candidate reference material was produced by Analytical Services International (St. George's Hospital, London, UK) by pooling anonymised patient blood samples followed by mixing thoroughly prior to bottling it into vials. Approximately one Litre of EDTA blood samples from tacrolimus administered patients was pooled by pouring each of the samples into a one litre plastic bottle. The pool was mixed for four hours prior to aliquoting and approximately 1.20 ± 0.05 g of the blood was added to 2 mL tubes. The cRM vials were 2 mL Ribbed Skirted Tubes (E1420-23200) with standard screw caps (E1480-0104) that were purchased from Sarstedt (Nümbrecht, Germany). The vials were then stored at -80 °C.

4.2.2 Reagents and standards

Higher-order traceable pure tacrolimus powder ERM-AC022a (LGC standards, Teddington, UK) was used to prepare standards. The certified purity of the material is 97.65 ±0.68 % *m/m* at the 95 % confidence interval. Stock standards and subsequent dilutions were gravimetrically prepared in 100 % Promochem Optigrade® acetonitrile (LGC standards, Teddington, UK). Screw-capped amber glass conical flasks were used for standard preparation (University of Southampton glassblower, Southampton, UK). Although flasks were not silanised, they were tested to check if tacrolimus was sticking to the glass and it was not. Care was taken to avoid unnecessary exposure to light

by storing standards that were left to equilibrate to room temperature wrapped in foil and in a cupboard.

Stable isotopically-labelled tacrolimus, $^{13}C^{2}H_{4}$-tacrolimus, was synthesised by Alsachim (Strasbourg, France) with isotopic enrichment 99 % ^{13}C, 98 % ^{2}H at a purity of 95 %, this is referred to as the SILIS hereafter. The stock standard of the SILIS was prepared in 100 % Optigrade® acetonitrile, by dissolving 10.00 mg in 100.0 g of acetonitrile. The subsequent dilutions were prepared in 70 % Optigrade® methanol in water (LGC standards, Teddington, UK). The mass fraction of the SILIS was calculated by comparison of peak area response of isotopically-labelled tacrolimus to the peak area response of tacrolimus standard.

The SILIS in DEM-IDMS method was used to obtain a peak areas ratio of the compound to its SILIS (R'_B for sample blend and R'_{BC} for calibration blend). In the DEM-IDMS equation the mass fraction of the SILIS cancels out, only the mass of the SILIS addition and the peak areas ratio were used. Hence, the mass fraction of the internal standard was assigned by its response to the compound. Stock standards, intermediate dilutions and working dilutions were prepared gravimetrically and stored in a freezer below -15 ºC.

Tacrolimus in whole human blood ERM DA 110 (LGC standards, Teddington, UK) was analysed as a quality control sample. The certified value of the CRM was 7.41 ±0.25 ng/g at 95 % confidence interval (CI). This higher-order CRM is a spiked material where a tacrolimus standard in solvent was added to human blood from healthy individuals who had not been administered with tacrolimus. The matrix-matched calibration blends were prepared using blank

blood (free of tacrolimus) that was obtained from Analytical Services International (St. George's Hospital, London, UK). The blank blood was checked by LC-MS to ensure the SRM channels were blank i.e., free of tacrolimus.

The blood cells, including erythrocytes to which tacrolimus binds due to its lipophilicity, were lysed with a zinc sulphate solution (0.04 M in 1:4 methanol:water v/v). This was prepared using zinc sulphate heptahydrate (purity ≥ 99.0 %) from Sigma Aldrich (Dorset, UK).

When the tacrolimus standards preparation validation study failed (described later in 4.2.8.1 and 4.3.2) indicating instability of standards in solvent, an evaluation of the different types of acetonitrile was performed. For acetonitrile purity evaluation, the following sources of acetonitrile were obtained: AcroSeal™ extra dry over molecular sieve 99.99 % and ACROS Organic™ extra pure acetonitrile 99+ % (Fisher Scientific, New Hampshire, US), Anhydrous acetonitrile 99.8 % and Ultrapure acetonitrile (Sigma Aldrich, Missouri, US) and Biosolve ULC-MS acetonitrile, which was used to prepare the solvent standards thereafter (Greyhound, Birkenhead, UK).

The density of the cCRM whole blood was measured. For the whole blood density study two density QC materials were used: Distilled water density standard and Dimethyl phthalate density standard were purchased from H&D Fitzgerald (Saint Asaph, UK). These two standards were used as QCs and were measured to validate and check the quality of the density study method which was performed in an environment where temperature was controlled and measured, and the air pressure were measured.

4.2.3 Instrumentation

4.2.3.1 LC-MS/MS

Two instruments were used to perform the analytical studies due to lack of availability on a single instrument. The same LC-MS method was used on both instruments after optimising the MS conditions for each. The MS conditions optimisation was performed using standards in solvent infusions and flow injections. The optimisation was performed manually for each condition. As mentioned earlier, all the performance of all MS instruments used in this work was evaluated using system suitability measures and were cleaned regularly, see 2.2.2.3. A Thermo Scientific TSQ Vantage Tandem Mass Spectrometer with TLX-1 liquid chromatography system (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used in addition to an Agilent 6490 Tandem Mass Spectrometer with binary pump 1290 liquid chromatography system (Agilent, Santa Clara, California, USA). The LC column used on both LC-MS systems was a Hypersil Gold C_{18}, 50 x 3 mm, 5 µm particle size (Thermo Fisher Scientific, Waltham, Massachusetts, USA). A gradient of water: methanol was used that started with isocratic 40:60 water:methanol for 1 min followed by a linear ramp to 90 % methanol over 1 min, the 90 % methanol was held for 3 minutes followed by a step to 100 % methanol for 3 minutes to wash the column before going back to equilibrate at the starting conditions for 10 minutes. The flow rate was 0.4 mL/min throughout the gradient program. The column was heated using a Thermo Scientific Hot Pocket to 50 ºC when using the Thermo LC due to the lack of a column oven unit in the system. The Agilent system had a built in column oven that was used. The injection volume used for the sample and calibration blends was 10 µL. A diverter valve was used;

the eluent from the liquid chromatography column was diverted to waste from 0-2 minutes and from 6.2 minutes until the end of the run. The mass spectrometer was operated in positive ionisation electrospray mode acquiring SRM transitions for quantification and confirmation, see Table 4.2. The optimised variables of each of the MS instruments that were used are in Table 4.3 and Table 4.4. The SRM transitions and their optimum collision energies required were optimised manually by infusion and flow injections of the standards in solvent and by manually adjusting the collision energies until arriving at the optimum signal of the precursor and product ions.

Table 4.2 Quantitative and qualitative SRM transitions

Instrument	Molecules	SRM transitions	Collision energy (V)
Thermo Vantage	Tacrolimus	*m/z* 826.2 → *m/z* 616.1	33*
		m/z 826.2 → *m/z* 415.1	48
	$^{13}C^{2}H_{4}$tacrolimus	*m/z* 831.2 → *m/z* 621.1	33*
		m/z 831.2 → *m/z* 420.1	48
Agilent 6490	Tacrolimus	*m/z* 826.2 → *m/z* 616.1	40*
		m/z 826.2 → *m/z* 415.1	50
	$^{13}C^{2}H_{4}$tacrolimus	*m/z* 831.2 → *m/z* 621.1	40*
		m/z 831.2 → *m/z* 420.1	50
*Quantitative SRM transition			

Table 4.3 Thermo Vantage MS conditions

Probe Position	D, 2mm	Aux Gas Pressure	30 (AU)
ESI Spray Voltage	2800 V	Capillary Temperature	380 ºC
Vaporizer Temperature	450 ºC	S-Lens RF Amplitude	200 V
Sheath Gas Pressure	50 (AU)	Declustering Voltage	0 V
Ion Sweep Gas Pressure	0 (AU)	Collision Gas Pressure	1.8 mTorr
Q1 resolution (*m/z*)	0.2	Q3 resolution (*m/z*)	0.7
Scan width *m/z*	0.01	Scan time (ms)	0.1

Table 4.4 Agilent 6490 MS conditions

Delta EMV (+)	500	Sheath Gas Flow	11 /min
Delta EMV (-)	0	Capillary	3500 V
Gas Temperature	290 °C	Nozzle Voltage	1500 V
Gas Flow	18 L/min	High Pressure RF (+ & -)	150
Sheath Gas Temp	250 °C	Low Pressure RF (+ & -)	60
Dwell time (ms)	50	Cell accelerator voltage (V)	5

4.2.3.2 Other instruments

High accuracy and precision analytical balances; Mettler Toledo XP205 and Mettler Toledo micro balance XP6 were used for all gravimetric preparation. The U ionizer anti-static device by Mettler Toledo (Mettler Toledo, Leicester, UK), Eppendorf electronic pipette E3x single channel (Eppendorf, Hamburg, Germany) were used for most of the preparation work. In the density study the temperature of the environment was measured using a high precision thermometer that measured temperature to three decimal places and was calibrated to ISO 17025 by a UKAS accredited calibration laboratory: F250 MK II Precision Thermometer (Automatic systems laboratories, Leighton Buzzard, UK).

4.2.4 Sample preparation

4.2.4.1 Preparation of standard solutions

The preparation of standards in solvent was performed gravimetrically. The higher-order traceable solid tacrolimus was taken out of the freezer a few hours before the preparation to equilibrate to room temperature. The standard was then mixed thoroughly before an aliquot of 29.01 mg of the powder was accurately weighed in a clean pre-weighed glass cup. Same gravimetric preparation procedure described in 2.2.3.1 and in 3.2.4 was used.

The mass fraction of each of the standards in solvent was accurately obtained by the gravimetric preparation procedure. The measurement uncertainty estimates of the mass fractions of the standards included the variation in the weighing process, the balance uncertainty and, most importantly, the purity value of the higher-order traceable solid standard. All standards in solvent were stored at -20 °C and were taken out of the freezer at least two hours before use to ensure equilibration to room temperature.

Before the use of any standard in solvent, it was weighed and the mass was compared to the mass of the same standard from after the previous use to check if any significant evaporation had taken place. If a standard had more than 0.50 % loss in mass during the freezer storage and temperature equilibration, it would not be used and a new standard would be prepared.

4.2.4.2 Blends preparation

The cCRM vial was vortex mixed for 30 seconds prior to any blood aliquot followed by positioning the pipette tip a third way into blood in the vial, to ensure homogenous sample and representative aliquot of the vial. Four types

of blends in addition to the quality control blanks (see 4.2.7) were prepared; sample blends which were prepared from the sample blood and SILIS and the calibration blends which were made of blank blood, tacrolimus standard and SILIS. Moreover, QC sample blends and QC calibration blends were prepared.

To each sample blend and QC sample blend, a calibration blend was prepared that was matching the sample blend in mass fraction i.e., both SB and CB would be at 6.43 ng/g. Initial matching was between the tacrolimus and the SILIS to achieve a ratio of tacrolimus: $^{13}C^{2}H_{4}$tacrolimus of R'_B = 1.00:1.00 with accepted tolerance on the ratio of ±10 %. Then the calibration blend was prepared at a mass fraction that is equal to the sample blend to achieve a ratio of ratios R'_B / R'_{BC} = 1.00:1.00, where R'_B is the ratio of tacrolimus: $^{13}C^{2}H_{4}$tacrolimus of the sample blend and the R'_{BC} is the ratio of tacrolimus: $^{13}C^{2}H_{4}$tacrolimus in the calibration blend. This was done by knowing an estimate of the mass fraction beforehand by the production of the material for the candidate reference material and from the certificate for the QC material. Achieving ratios of 1:1 was usually achieved after a couple of iterations to match the SILIS and the mass fraction of sample blends.

The blank blood vial, the QC blood vial and the candidate reference material blood vials were taken out of the -80 °C freezer and stored in the fridge at 4 °C overnight to allow defrosting. Then on the day of sample preparation, the vials were taken out of the fridge to equilibrate to room temperature for 1.75 to 3.00 hours before preparation depending on the availability of the balances for use. However, a minimum of 1.75 hours was required to ensure temperature

equilibration to room temperature prior to weighing which is crucial for the weighing accuracy on the highly sensitive analytical balances. Each sample vial was vortex mixed twice for 30 seconds each time prior taking an aliquot of out of the sample to prepare the sample blends.

Each addition to the blend was prepared using a fresh positive displacement pipette tips (tip with built-in piston) that were obtained from Fisher Scientific (Loughborough, UK). The additions were however measured gravimetrically where the blend vial was weighed before and after the addition using an analytical highly accurate balance which had a measurement uncertainty of ± 41 µg. Each weighing step was repeated at least three times, if a trend in the measured masses occurred, whether loss or increase in mass, the weighing continued and if the trend persisted, it was investigated and resolved before continuing.

All blends were prepared gravimetrically in 4 mL screw-capped silanised amber glass vials. For sample blends, the pipette was set to aliquot the equivalent of 0.3 g of whole blood and of isotopically-labelled internal standard. The exact mass of each addition was measured by the gravimetric preparation procedure. Similarly, for the QC an accurate amount of aliquot of whole blood sample was taken and an accurate amount of SILIS added. For quality control blends (QC) the pipettes were set to aliquot the equivalent of 0.3 g of QC whole blood was taken and 0.35 g of SILIS added by measuring the accurate mass.

Likewise, for calibration blends an accurate measurement of each addition was achieved by the gravimetric preparation described above. The pipette

was set up to a volume equivalent to 0.30 g of blank whole blood, 0.08 g of tacrolimus standard in solvent and 0.30 g of SILIS was added (all values listed here are the closest nominal number, the exact masses taken are weighed to 5 decimal places). Finally, the calibration blends, used for quantifying the QC material, were prepared using the same method where about 0.30 g of blank whole blood was taken, 0.09 g of tacrolimus standard in solvent and about 0.35 g of SILIS was added to match the mass fraction of ERM DA 110. The sample blend preparation order was randomised.

Every single addition to every blend was weighed at least three times to five decimal places on the balance and the mean of two of the masses was used in the measurement equation. The numbers mentioned above such as 0.30 g or 0.08 g are only indicative of the nearest nominal mass.

4.2.4.3 Cell lysis and solid phase extraction

The extraction and solid phase extraction (SPE) method was adapted from the procedure currently used at the Royal Brompton and Harefield Trust Hospital for routine analysis of tacrolimus in patient samples. To release the tacrolimus, red blood cells were first lysed using a zinc sulphate extraction solution. The solution was prepared on the day of blends preparation by dissolving 0.838 g of zinc sulphate heptahydrate (99.5 % purity, Acros Organics, from Fisher Scientific, Loughborough, UK) in 15 mL of high purity (>18 MΩ.cm) Elga water. To this stock, 60 mL of methanol was added to generate the extraction solution used for lysis of red blood cells and tacrolimus extraction.

After the gravimetric preparation was complete, 0.7 mL of the extraction solution was added to each blend. Each vial was then vortex mixed twice for 30 seconds. The blends were left to equilibrate in a refrigerator overnight. The following day, the samples were vortex mixed for 30 seconds and then centrifuged for 10 min at 3300 g at 4 ºC. The supernatant was transferred by decanting the vial to a polypropylene centrifuge tube leaving behind the sedimented layer of cellular matter created by the centrifugation. Highly pure water by the Elga system (1.25 mL) was added to each sample extract to reduce the percentage of organic solvent to below 40 % prior to SPE.

SPE was conducted using a 3M Empore™ C_{18} SPE 96 well plate with 10 mg sorbent mass (Sigma Aldrich, Dorset, UK). After initial pre-conditioning steps with methanol and then water, the extract was loaded onto the SPE plate under vacuum, washed with 0.5 mL of water:methanol (6:4 v/v) solution three times, followed by elution with 250 µL methanol into a 96-well collection plate, which was transferred to the LC-MS/MS autosampler for analysis.

4.2.5 DEM-IDMS sequence running order

In studies that used DEM-IDMS, each sample/QC blend (SB) was injected five times on the LC-MS/MS, each time bracketed by a calibration blend (CB); [CB-SB-CB] x5 for each sample and each QC. The overall order of samples in the sequence was: blank blends (single injection each), QC blend 1, the sample blends, QC blend 2, blank blends (single injection each). Within this the order of samples in the analytical batch was randomised to that of the preparation order to rule out any trends in data due to the preparation order. As an example, the sample that was prepared first in the order of sample preparation (SB1) was placed at a random place in the order of the samples analyse (e.g. seventh sample to be analysed) so the SB1 five brackets were analysed after the seventh prepared sample (SB7), i.e. [CB7-SB7-CB7] x5

would be followed by [CB1-SB1-CB1] x5 in the analytical batch. To summarise, each sample blend with its corresponding calibration blend and each QC blend and its corresponding calibration blends were analysed five times and each of the blanks was analysed twice.

4.2.6 Measurement equation and uncertainty calculation

The calculated amount of tacrolimus in each of the sample blends was determined using the simplified double IDMS equation (18,19). The same DEM-IDMS equations and calculations as those described in Chapter 2 in section 2.2.4 were used for the tacrolimus measurements.

4.2.6.1 Coverage factor components

The coverage factor (k) used for the work described in this chapter was approximated to 2 (for details of the coverage factor see 1.90) since there were sufficient degrees of freedom to make this valid. The numbers of blends were combined for each study because the candidate CRM samples were all originally sub-aliquoted after thorough mixing from a common liquid pool of blood that was homogenous and showed no freeze/thaw issues. For example, in the characterisation and homogeneity study where 10 cCRM vials were analysed in triplicate individually prepared blends on three different weeks and each blend was analysed five times by LC-MS with their bracketing CBs, the mean measurement of these five replicate measurements was reported as the mass fraction for that blend, then the three mean mass fractions from the three different weeks was reported as that sample vial's mass fraction. The intermediate mean of the LC-MS analysis of each blend was used which made n = 30 and df = 29 for the characterisation and homogeneity study alone. For the final candidate reference value, the measurements of all blends analysed

across the five studies were all used which gave a total of n = 108 and df = 107. The appropriate coverage factor for two-sided 95% confidence intervals on a normal distribution at 100 degrees of freedom would be 1.9853 (20). This was simplified to k = 2, which was used for the expanded uncertainty calculations.

4.2.7 Quality control

4.2.7.1 Analysis of control material

With each batch, one vial of the higher-order CRM tacrolimus in whole human blood ERM DA110a was analysed in duplicate (two independent measurements from two individually prepared blends with their corresponding CBs). One blend was analysed 5 times [CB-QC-CB] prior to the sample analysis in the sequence and the second after the samples were analysed, at the end of the sequence before the analysis of the blanks. QC measurements were compared to the certified reference value and if the measurement agreed with the reference value then the batch would pass the QC check.

4.2.7.2 Comparison of quantification & confirmation SRMs

For all assays two SRMs were acquired, the data acquired was processed using Xcalibur™ software for peak integrations. The peak areas were then put into Microsoft® Excel spreadsheets to calculate the ratios of tacrolimus to isotopically-labelled tacrolimus and the ratio of ratios of the sample blends to the calibration blends. The application of the measurement equations for mass fraction calculation were also performed using Microsoft® Excel. The SRM with the greatest signal-to-noise ratio was used for quantification, which was *m/z* 826.2 → *m/z* 616.1 for tacrolimus and *m/z* 831.2 → *m/z* 621.1 for

$^{13}C^{2}H_{4}$tacrolimus. The confirmation SRM transition data was processed and used to confirm the data produced by the quantification SRM where the mass fractions would agree within 10 % of each other.

4.2.7.3 Blanks

The following blanks were analysed at the beginning and the end of each sequence:

Table 4.5 Types of blank samples analysed.

Blank blend	Label	Blood	tacrolimus standard	$^{13}C^{2}H_{4}$tacrolimus standard
Blank sample	B-S	~0.3 g sample blood	Not added	Not added
Blank QC sample	B-QC	~0.3 g QC blood	Not added	Not added
Blank blood	B-B	~0.3 g blank blood	Not added	Not added
Blank blood with tacrolimus	B+N	~0.3 g blank blood	~0.8 g added	Not added
Blank blood with isotopically-labelled tacrolimus	B+L	~0.3 g blank blood	Not added	~0.3 g Added

4.2.8 Experimental design

4.2.8.1 Assessment of the preparation of the standards in acetonitrile

The tacrolimus standards in solvent that were prepared gravimetrically needed to be validated to ensure no source of error was introduced to the measurement by the standards preparation process. Additionally, validating standards in solvent enabled assessing the potential variability introduced to the measurement from using different standards. This study was performed by having two different analysts prepare stock standards gravimetrically from

two different higher-order traceable solid tacrolimus standards vials on two different days in addition to subsequent dilution standards.

The two independent working level dilution standards prepared by the two analysts at the NML (Dima AlMekdad and Dr Camilla Liscio) were then analysed using the DEM-IDMS method mentioned earlier. One standard was considered to be the sample and the other was considered to be the calibration standard. Then sample blends and calibration blends of the standards were prepared in solvent only (no matrix). The LC-MS measurement of the sample standard was expected to be equal to the value assigned from the gravimetric preparation of the standard. If the measurement agreed to the value within ±5 % that indicated the standards preparation agreed and the standards validated were ready to be used for the analysis of blood samples.

4.2.8.2 Characterisation & homogeneity study

To assign a reference measurement to the pooled patient blood candidate reference material and assess its homogeneity, ten cCRM vials were analysed. The analysis was performed over three weeks, each week a single aliquot (blend) of each cCRM vial was analysed by LC-MS five times. The reference value assigned to each vial was the mean of the triplicate blend analysis over three weeks and its measurement uncertainty encompassed the measurement uncertainties of each of the independent values. The mean of the 10 vials measurements was used to assign the reference measurement of the entire candidate reference material that was to be assigned on the certificate of the cCRM.

4.2.8.3 Storage temperature stability

To assess stability of the material, a short term expedited stability study was performed. The samples had been stored at -80, -20, 18 and 37 °C for 24 hours, 2, 4 and 6 months. Two vials were stored per temperature and duration conditions. The complete analyses were performed over four batches over four different weeks. The analyses were performed using DEM-IDMS. The preparation and sequence run orders were randomised for each batch. The measurements of the different samples were compared with the reference value that was assigned in the characterisation study, using single factor analysis of variance (ANOVA). Any value that was statistically-significantly different (two-sided) from the reference value was deemed unstable using t-test and ANOVA.

4.2.8.4 Within vial homogeneity study

Because of the complexity and heterogeneity of whole blood samples, in the sample preparation process each aliquoting of blood from the cCRM vial to the sample blend vial was preceded by 30 seconds of vortex mixing twice followed by positioning the pipette tip a third way into the blood in the cCRM vial. Additionally, the aliquoting of a single vial to be analysed in triplicate or duplicate in any of the studies was done randomly and in some studies on different weeks. Hence, to evaluate the homogeneity within the cCRM vials using the sample preparation method was needed. This was evaluated by analysing five cCRM vials by DEM-IDMS in triplicate without randomisation of preparation order. These sample blends were then randomised in the LC-MS sequence. The data obtained was investigated for trends in measurement i.e.,

if the first aliquoted blend had lower or higher values than the second and then the third aliquot.

4.2.8.5 Usage stability study: freeze thaw cycles and opened vial stability

To assess the impact of the usage of the material by the end user on potential variability in measurement and material stability, a simplified usage stability study was performed. The study investigated the impact of freezing and thawing the material 5 times and opening the vial, taking an aliquot and then storing it in the fridge for 2 weeks at <7 °C. Duplicate vials underwent each set of conditions and an additional vial that did not undergo any treatment was used as a reference to compare the treated vials against. All vials were all analysed in triplicate using DEM-IDMS. Results were compared against the untreated reference vial that was only stored at –80 °C and did not undergo freeze/thaw since the production of the material.

4.2.8.6 Blood density study

Five vials of the material and the two density QC materials were equilibrated to room temperature prior to performing the study. Accurate measurement of the room temperature was performed using a calibrated thermometer probe that measures to three decimal places with certified measurement uncertainty. Additionally, measurements of the air pressure in the area of the study (Teddington, UK) were obtained via the National Physical Laboratory online barograph (21) which is conveniently located in Teddington on the adjacent site to the building.

The pipette used for the density study was also the Eppendorf Multipette electronic pipette that uses positive displacement pipette tips that were preassembled with pistons. The measurements consisted of calibrating a 0.5 mL pipette tip with highly pure water from the Elga system by adding 0.5 mL to pre-weighed vials 10 times, the %RSD of the 10 masses of the water addition was 0.04 %. The same tip was then used to make additions of 0.5 mL of the candidate reference material whole blood into empty pre-weighed glass vials which were also equilibrated to room temperature and the mass of the volume added was then weighed. The materials density measurement were done in the following order:

1- Fitzgerald distilled water density standard
2- The cCRM blood material
3- Fitzgerald dimethylphthalate standard.

4.3 Results & Discussion

4.3.1 Labelled internal standard SRM transitions optimisation

Isotopically-labelled tacrolimus was custom made by Alsachim as at the time there was no commercially available SILIS of tacrolimus. Upon receipt of the standards at the method development stage, mass spectra of both tacrolimus and isotopically-labelled tacrolimus were generated by MS direct infusion and loop injections (or flow injections) with mobile phase containing ammonium acetate without column. The spectra of isotopically-labelled tacrolimus immediately showed an unconventional isotope distribution of the SILIS (see Figure 4.3). There was overlap between tacrolimus isotope distribution and $^{13}C^{2}H_{4}$tacrolimus isotope distribution. Hence, the selection of the precursor

and product ions of the internal standard were selected based on the elimination of this overlap rather than the masses with the highest signal *per se*. Eliminating the isotope overlap between the molecule and its internal standard (both at the precursor and product ions) is crucial to the specificity and accuracy of the method and to eliminate bias in measurement.

This purity issue with the SILIS of tacrolimus did not deem it unfit for use, because the use of the selected mass for the SRM transitions helped overcome the contribution of the other isotopes such as the overlapping ions (see Figure 4.3), and checks were put in place to monitor this such as analysing blank blends to detect for any signal from contributing isotopes.

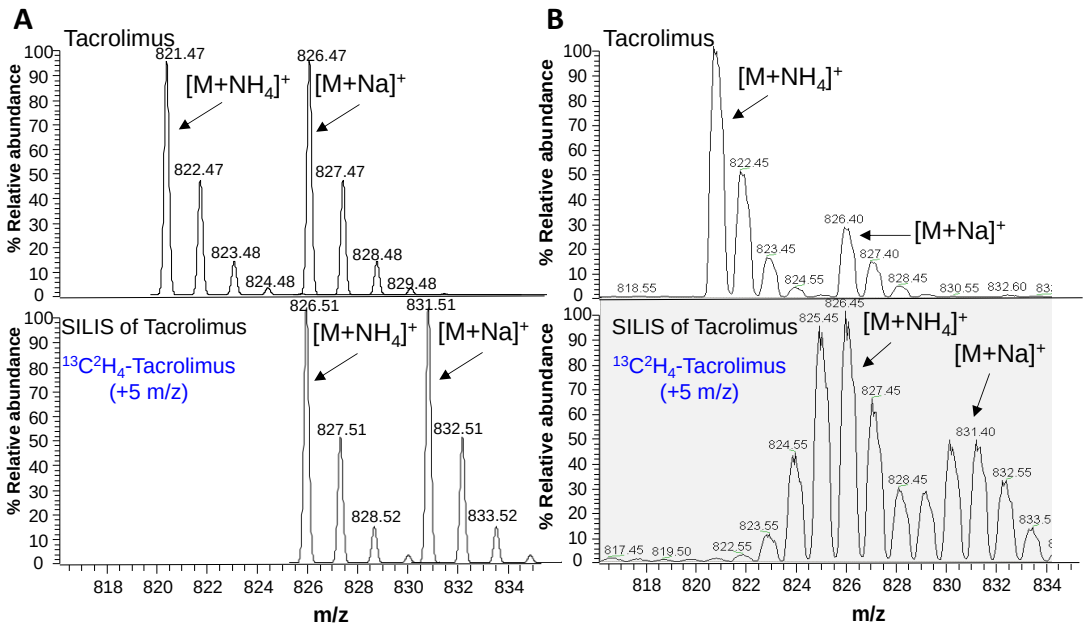


Figure 4.3 Mass spectra of tacrolimus and the isotopically-labelled tacrolimus using the Thermo Vantage TS-Q: A) software-generated spectra of tacrolimus and isotopically-labelled tacrolimus with one ^{13}C and four $^{2}H_4$, showing that the two isotopologues do not overlap; B) MS generated spectra of tacrolimus and the isotopically-labelled tacrolimus purchased from Alsachim obtained by injecting the solvent standards with mobile phase containing ammonium acetate. The same peaks were observed when using LC-MS injections.

The contribution of the SILIS to tacrolimus SRM transitions was assessed by analysing a 'blank + labelled' sample which is made up of the blank blood that

was used in the preparation of the calibration blends with the addition of SILIS only. Assessing the contribution in tacrolimus SRM transition was important because any contribution in the tacrolimus peak from the SILIS would bias the ratio of the natural to isotopically-labelled tacrolimus subsequently causing bias in measurement. Ideally, there would be no contribution or minimal contribution to improve the specificity of the method and reduce variability in measurement.

The peaks detected in the tacrolimus SRM channels were measured against the peaks of the SILIS. The isotopically-labelled tacrolimus contribution into the tacrolimus SRM transitions was <0.7 % and <0.9 % for the quantification and confirmation SRMs, respectively. This very small contribution is acceptable of the DEM-IDMS. In a single IDMS method (see 1.8.2) such contribution would result in bias of the measurement but this low contribution was negligible and was considered suitable for the method because it would be the same in both the sample and the calibration blends as in this higher-order reference method the gravimetric preparation provides the exact masses of internal standard added to the blends which would account for the bias in the ratios that would not affect the ratio of ratios of the sample blend to the calibration blends bracket which was the component of the measurement in the measurement equation (see Equation 2.1)

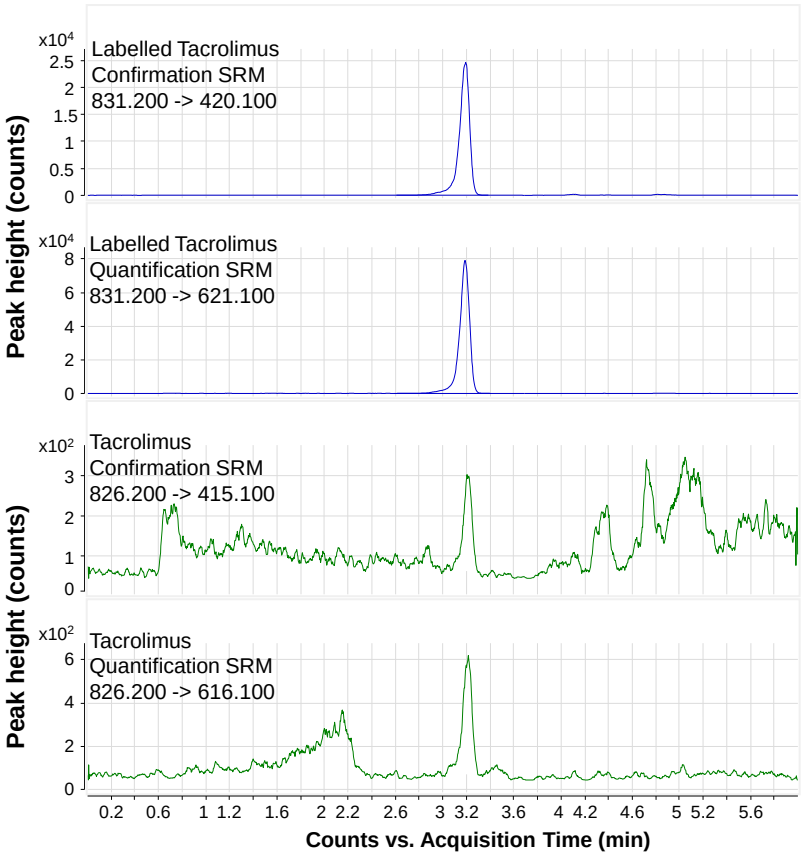


Figure 4.4 Chromatogram of blank blood with the addition of the tacrolimus SILIS at 7 ng/g, showing the small contribution of <1 % peak in the compound SRM transitions to that of the peaks in the SILIS transitions.

4.3.2 Standards in solvent validation study

The tacrolimus standard in solvent was the cornerstone of the measurement and was the key that enables traceability of measurement. It was the standard that was used to prepare the calibration blends based on which the sample blends were measured. Hence, validation of the standards in solvent used for calibration blend preparation was crucial. Standards validation included preparing two or more stock solvent standards in addition two subsequent dilution standards down to the working mass fraction level, on two different days by two different analysts. These standards were then measured against each other by LC-MS. The measured mass fraction of the standard should agree with the gravimetric preparation mass fraction.

Tacrolimus in blood cCRM analysis studies were performed at two different time points with about a year gap in between. At the first set of analyses, two sets of standards in solvent were prepared and validated by DEM-IDMS analysis (see Figure 4.5.A). The reference value to check the measured mass fractions against was the mass fraction obtained by the gravimetric preparation of the standards in solvent (see 4.2.4.1 for gravimetric preparation of standards in solvent).

In the comparison study, one standard was used as the sample to prepare the sample blends and the other was used as a standard to prepare the calibration blends. The LC-MS measured mass fraction of the standard in solvent agreed with the reference value with % difference of the measured vs the gravimetric mass fractions of <1 % and expanded measurement uncertainty estimate of < 2 % at 95 % confidence intervals (CI) using coverage factor, k=2. Hence, this set of standards was used for the characterisation and homogeneity study and storage temperature study. However, upon resuming the analytical work after a pause for a few months, standards in solvent comparison studies failed (see example in Figure 4.5.B), this will be discussed later on.

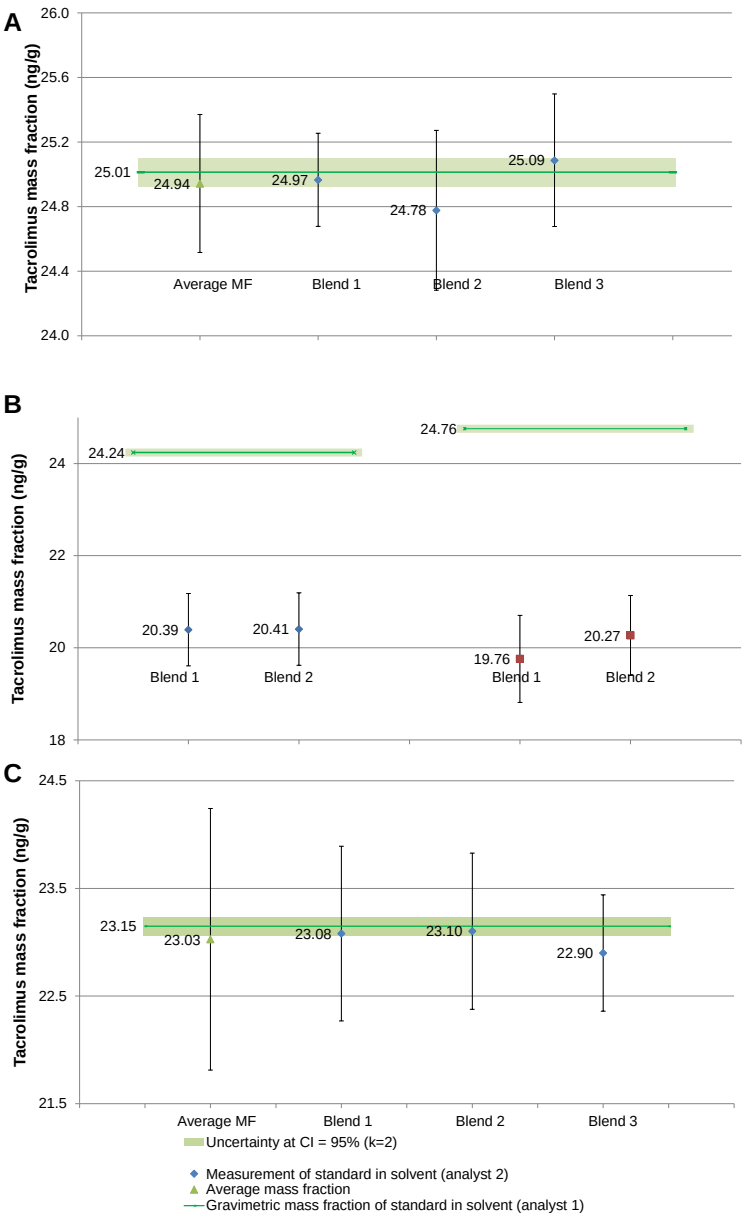


Figure 4.5 Standards in solvent validation studies: A) A successful standards in solvent comparison of the two standards were used for the characterisation and homogeneity study and the storage temperature study; B) Failed standards in solvent comparison, these standards were not used for any analytical work; C) A successful standards in solvent comparison of the standards were used for within vial homogeneity study, usage stability study and commutability study.

4.3.3 Characterisation and homogeneity study

The characterisation and homogeneity study aimed to assign a reference value with its corresponding measurement uncertainty estimate to the tacrolimus in pooled patient blood candidate reference material and assess the overall homogeneity of the material. This was performed by analysing ten different vials by DEM-IDMS, the analysis was performed on three different weeks where each week the vials were analysed in singlets. This resulted in three independent measurements each vial, the mean of these three measurements was the assigned value for each vial. The results of the study (see Figure 4.6) proved the material to be homogenous and the reference value and corresponding expanded measurement uncertainty estimate that was assigned to the material was 6.53 ± 0.19 ng/g (CI 95 %, k=2).

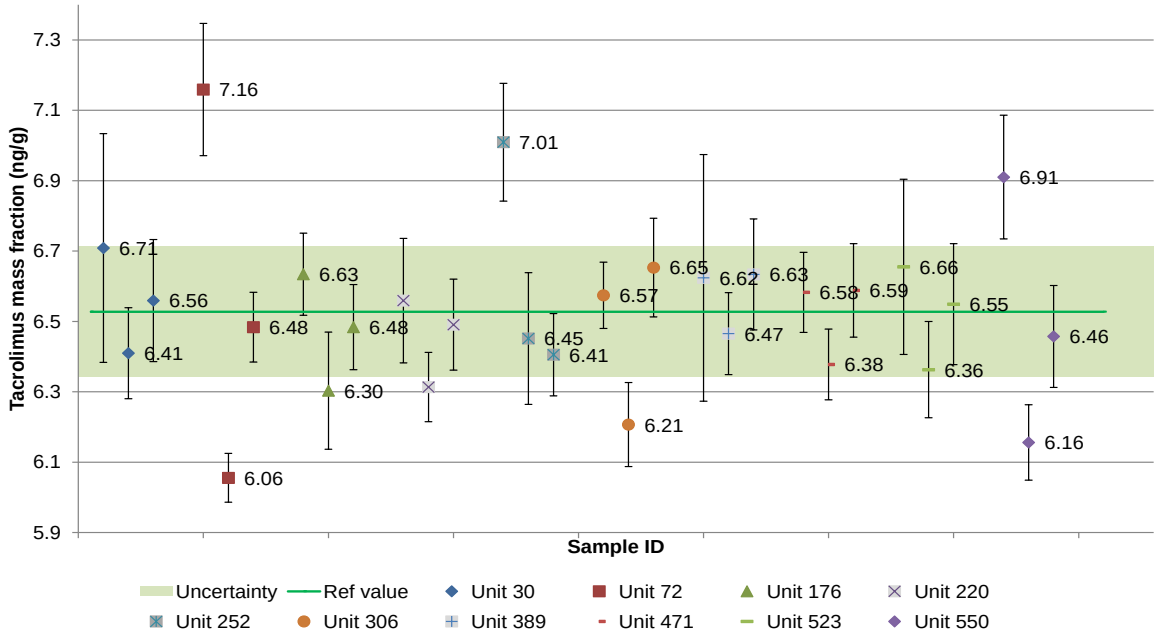


Figure 4.6 Reference value assignment for tacrolimus cCRM showing most of the replicate blends measured overlapping with the overall mean measurement and its expanded uncetainty. Each point is a measurement the mean of five replicate LC-MS injections of a single aliquot of one vial. Each vial was analysed three times on three different days. The solid green line is the reference measurement that was calculated using the mean of all measurements and the measurement uncertainty assigned as shaded area around the reference value is assigned by combining the measurement uncertainties of each of the measurements and multiplied by coverage factor k=2 to expand the uncertainty. Error bars = expanded uncertainty (95 % CI). The measurements that do not fall in the green shaded area were legitimate measurements that were taken into account in calculating the measurement uncertainty and fall within a normal distribution of results. The uncertainties of each of the measurements varied from day to day as expected however all were within the acceptable overall range of uncertainties for the method.

To add another layer of selectivity to any quantitative LC-MS method, it is best practice to acquire more than one SRM transition for the compound and internal standard. It is recommended that the ion ratios of the compound abundance is calculated between the two SRM transitions (22,23). In this study, not only were the ion ratios calculated but the peak areas acquired in the confirmation SRM transition were used for measurement calculations. Figure 4.7 shows the agreement between the measurements obtained by quantification and confirmation SRM transitions which agreed within 0.5 % confirming the specificity of the method which was well within the acceptable criteria defined by the NML IDMS flexible standard operation procedure (SOP).

The data plotted in the figures is indicates the assigned cCRM vial code. This code is based on the order of which the materials were bottled i.e., unit 55 was the 55th vial to be bottled from the main pool of blood. In each study the selection of the vials was randomised and the data was examined for any trend in the bottling order, no trend was observed. Additionally, with mass fraction plots such as the ones above and the following ones throughout this chapter, each measurement is compared against the Y-axis and the solid green line with the shaded area on the chart only. The X-axis is an ordinal not a scalar variable for the points to be plotted.

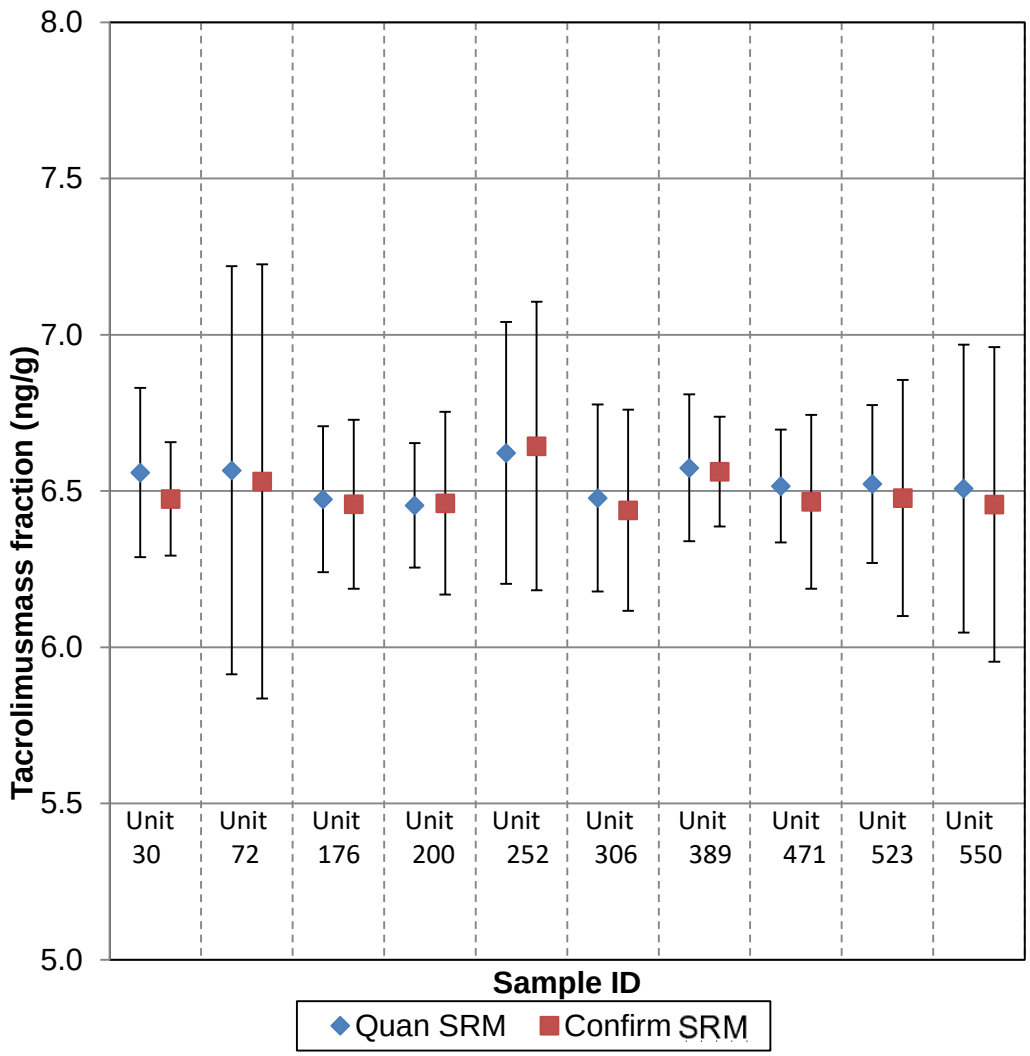


Figure 4.7 Comparison of quantitative and confirmation SRM transitions data showing agreement between measurements obtained using the two different SRM transitions proving the high selectivity of the method. Each data point is the mean of the three measurements of one vial. Data of all ten vials that were used to assign the reference value to the material are shown. Error bars = Expanded uncertainty (95 % CI). The uncertainties of each of the measurements varied as it would be expected however all were within the acceptable overall range of uncertainties for the method. All of these uncertainties were incorporated in the final combined measurement uncertainty estimate of the material.

4.3.4 Storage temperature stability study

A storage temperature study was performed to assess the stability of the material at different temperatures (-80, -20, 18, 37 °C) over different durations of time of 24 hours, 60, 120 and 180 days. Two vials underwent each of the storage conditions. Only vials stored at 37 °C for 24 h could be analysed as

the contents of the vials stored for longer times had solidified. Twenty-six vials of the total of 32 vials were analysed on two sample sets and each set was analysed twice in singlicate independently on two different batches on two different weeks, the mean of the two measurements was used to assign a value to each vial. The results of the study shown in Figure 4.8 demonstrate the stability of the material at the different storage condition as all measurements with their uncertainties overlapped with the reference value. Additionally, an analysis of variance (ANOVA) was performed to evaluate the significant difference of the temperature of storage on stability of the material overall, regardless of duration of storage at each temperature. This was a single factor ANOVA and no significant difference was observed between the three different temperatures used for storage: $F(45, 2) = 0.05$, $p = 0.95$. This indicated the material was stable at – 80 °C, -20 °C and 18 °C up to six months of storage.

Moreover the 52 measurements of the 26 samples had a relative standard deviation of 1.5 %. This was in exceptionally good agreement considering that the variables it encompasses were numerous and include; a) different vials of cCRM, b) analysis of each blend of the duplicate of each vial was performed on different days including sample preparation, c) sample vials were stored at different temperatures and d) different bottling order in production of material. Such good agreement in measurements did not only indicate the stability of the material but additionally the robustness of the method as is expected of a higher-order RMP.

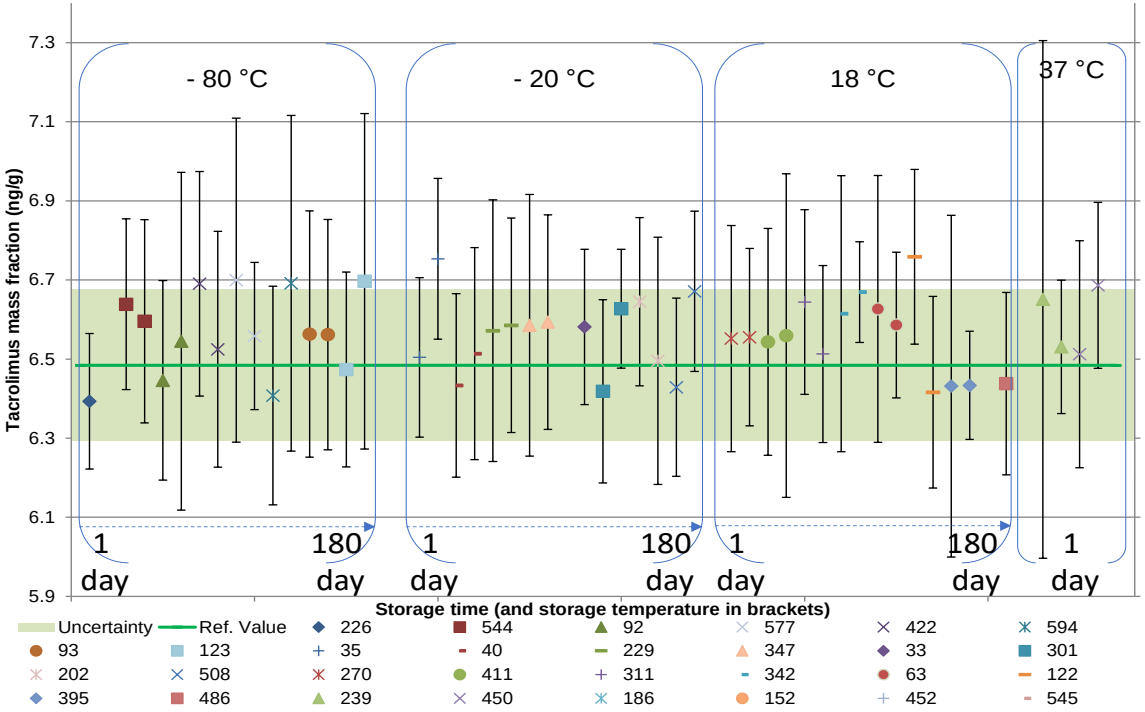


Figure 4.8 Storage temperature stability study showing the cCRM to be stable up to six months at storage temperature – 80 °C up to 18 °C. Each data point is the mean of five replicate LC-MS injections of a single aliquot of one vial. The data points are grouped according to storage temperature and are in order from left to right in increasing storage time from 1 day, 60 days, 120 days to 180 days. The reference value assigned through the characterisation study is shown as a solid green line with its expanded measurement uncertainty (k = 2) as the shaded area. Vials stored at 37 °C for more than 1 day all solidified and were not analysable. Error bars = Expanded measurement uncertainty (95 % CI). The uncertainties of each of the measurements varied as it would be expected however all were within the acceptable overall range of uncertainties for the method. All mean mass fractions fell within the reference measurement range, only a few means (e.g., a replicate of each of vials 123, 122) were just above the range however their corresponding uncertainties overlapped with the reference measurement range (green shade).

4.3.5 Standards in solvent investigation

After the completion of the characterisation and homogeneity study and the stability study, the work on the material was paused for over a year due to competing demand on resource at the NML (e.g. availability of instruments). A new set of tacrolimus standards were prepared from the higher-order solid powder tacrolimus CRM before starting the work again. The preparation of new standards was evaluated by a standards in solvent comparison study like the one mentioned earlier (see 3.2.9.1.2). However, the standards comparison study failed where the measured mass fraction of the standard was 10 % smaller the gravimetrically prepared mass fraction. Several new sets of standards were prepared and several standards validation studies were performed to investigate the issue but also failed. The several studies performed investigated many possible reasons for the discrepancy including but not limited to; the type of glass vessel being used for the standards preparation, light conditions, different pure tacrolimus standards and different mass fraction levels. Figure 4.5.B is an example of one of the several failed standards comparison studies, where the LC-MS measured mass fractions and their expanded measurement uncertainties did not encompass the reference measurement obtained by the gravimetric preparation of the standards. The percentage differences between the measured mass fractions and the reference mass fractions obtained by gravimetric preparation were 15.8 % and 19.2 % for standards shown in Figure 4.5.B.

The work focused on understanding what the cause of the degradation was to eliminate it rather than what was the degradation products were. Stock standards that were at mg/g level were compared by LC-UV to assess if the

degradation or an error in preparation had occurred at the stock standards level because the stability of tacrolimus standards in acetonitrile was evaluated in the past. The results of the LC-UV showed that different stock standards that were prepared on different days by different analysts from different higher-order solid powder tacrolimus CRM all agreed when analysed fresh on the day of preparation and days and weeks later. Then higher mass fraction (0.5 µg/g) standards in solvents were also analysed by LC-MS and the measured mass fractions agreed the gravimetric mass fractions. Hence, it was concluded that the degradation could be mass fraction dependent as the degradation was more pronounced at lower ng/g mass fraction which is the working dilution level (25 ng/g). Further work was performed to understand what was causing the degradation at the low level mass fractions.

In the process of assessing the stock standards, an analysis by nuclear magnetic resonance (NMR) was performed to compare the stock solvent standards. As part of the NMR evaluation, an aliquot of the acetonitrile that was used to prepare the standards was analysed as a blank. NMR showed that the acetonitrile contained water, methanol and other unknown impurities. An investigation of the acetonitrile commenced that included different acetonitrile types/suppliers and purities where the different solvents of acetonitrile, which were analysed by NMR (see 4.2.1). Figure 4.8 includes eight overlaid NMR spectra, the most impure acetonitrile was the one that was being used for the standards comparison studies that kept failing. That acetonitrile not only contained water (≤3 %) but and methanol in addition to other unknown impurities of ≤1 %. Methanol was not used to prepare any standards in solvent in this work because previous work at the National

Measurement Laboratory (NML) at LGC investigated the stability of tacrolimus in methanol and found that tacrolimus interacts with methanol causing a ring opening by oxidation (24).

Moreover, the impact of the other unknown impurities on the degradation of tacrolimus is unknown and could be of significance. Further work would be required to understand the degradation pathways and what were the triggers of the degradation. Upon comparing the different types of acetonitrile, Biosolve ULC-MS® grade acetonitrile was found to the be the purest with no NMR detectable impurities, this acetonitrile was then used to prepare new standards. A standard validation study was performed, and the standards agreed with percentage difference of the measured vs the gravimetric mass fractions of < 0.5 % and expanded measurement uncertainty of < 2 % at 95 % confidence intervals using coverage factor k=2, see Figure 4.5.C. These standards were subsequently used for the within vial homogeneity study, the usage stability study and within vial homogeneity study.

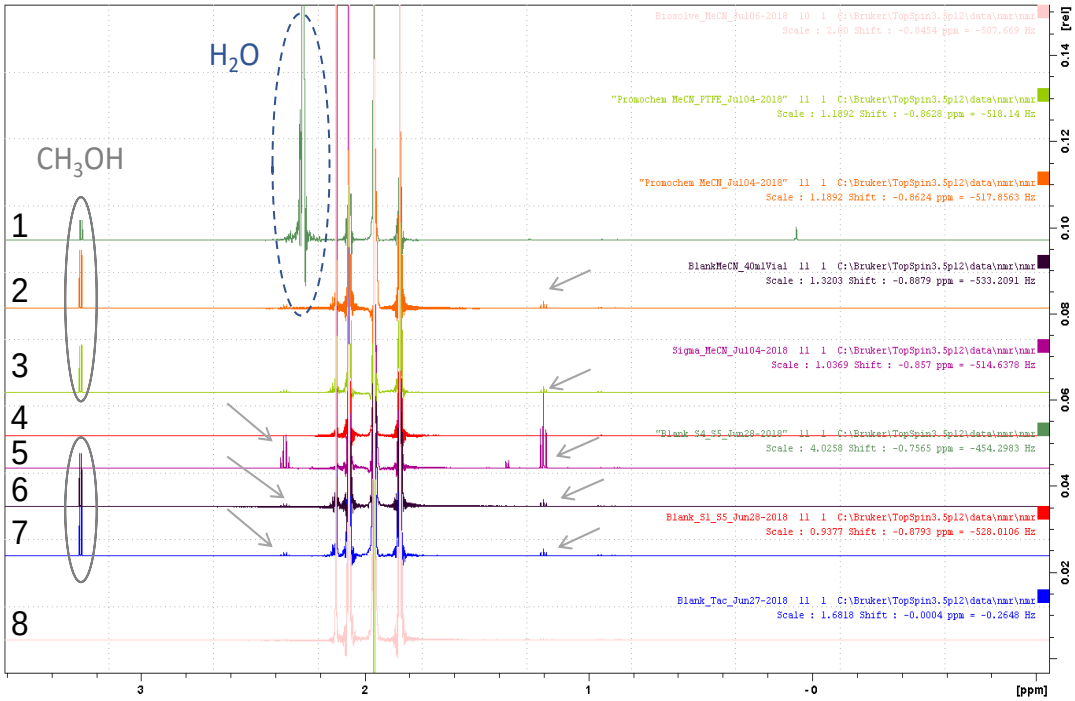


Figure 4.9 NMR spectra of eight different bottles of acetonitrile; arrows pointing at unknown impurities. 1) Promochem Optigrade acetonitrile used for the preparation of one of the sets of the standards in solvent which failed the validation test; spectra shows presence of water, methanol and other impurities 2) Promochem Optigrade acetonitrile different batch number, shows presence of methanol and other impurities 3) Promochem optigrade acetonitrile of the same batch as 1 but a freshly opened bottle shows presence of methanol and other impurities 4) Fisher Scientific AcroSeal™ extra dry over molecular sieve 99.99 % showing trace amounts of unknown impurities 5) Sigma Aldrich Anhydrous acetonitrile 99.8 % showing small amounts of unknown impurities 6) Sigma Aldrich Ultrapure acetonitrile showing small amounts of unknown impurities in addition to methanol 7) Fisher Scientific Acros extra pure acetonitrile 99+ % showing the presence of methanol and other unknown impurities 8) Biosolve ULC-MS acetonitrile, the purest acetonitrile with clear spectrum with no detectable impurity, this was subsequently used to prepare solvent standards for analytical measurements.

4.3.6 Within vial homogeneity study

The characterisation and homogeneity study (see 4.3.3) included triplicate independent measurements of 10 cCRM vials. Among these thirty measurements, six measurements and their respective uncertainties did not overlap with the mean measurement and its combined measurement uncertainty (see Figure 4.10). This could be explained as the measurements falling within 5 % of the normal distribution of the results that falls outside the population of data that is covered by the expanded measurement uncertainty.

Nevertheless, these six measurements could indicate an issue with the homogeneity within the cCRM vials. Tacrolimus is a highly lipophilic molecule and is extensively bound to red blood cells rather than being suspended in plasma (3). It could be that the aliquoting process by the pipette out of the vial was not representative of the sample. Despite the cell lysis step and vortex mixing for 30 seconds twice prior to aliquot, it could be that the aliquot taken for analysis had less or more amount tacrolimus than another aliquot. Therefore, the aliquoting process could be affecting the amount of tacrolimus taken in the preparation procedure. Evaluating within vial homogeneity under the sample preparation method was hence important to determine any variability in the measurement that could be caused by it.

The within vial homogeneity study was performed by analysing five vials in triplicate on the same day using the same exact sample preparation and aliquoting method as was used in the characterisation and homogeneity study. As per the method (see 4.2.4) each vial was vortex mixed for 30 seconds twice prior to taking an aliquot out to prepare each sample blend. Each sample blend was prepared with a fresh pipette tip and the aliquoting of blood to prepare the

three sample blends was performed consecutively without randomising the order of preparation while the LC-MS run order was randomised. The data of the study (see Figure 4.10) did not replicate the phenomenon that was seen in the characterisation and homogeneity study where random measurements did not overlap with the reference measurement, all measurements in this study agreed with the reference value.

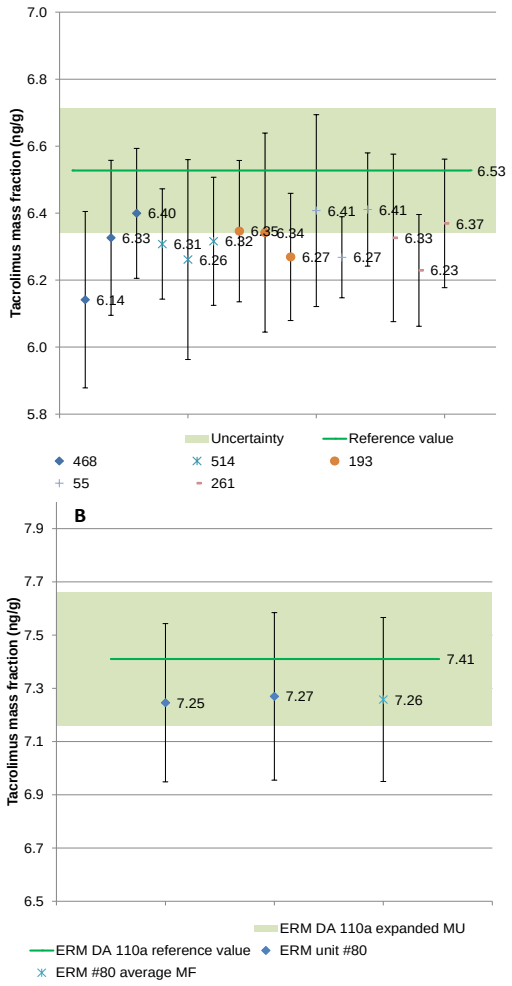


Figure 4.10 Within vial homogeneity study results; A) the measurements of the three independent measurements of each tacrolimus candidate reference material. All measurements overlap with the reference value. B) the measurements of the duplicate independent aliquots of the QC material vial; ERM DA 110a, the higher-order CRM of tacrolimus spiked in human whole blood. The result of the mean two measurements is in very good agreement with the certified reference value. Error bars = Expanded uncertainty (95 % CI), (k=2).

The order of preparation did not show any trends in measurements except for the one vial, this could be random distribution of results, data of more vials was required to assess that. However, because no trend was observed within vials in any of the analytical studies performed on the material, it was safe to conclude that the material was homogeneous within vial using the preparation method conditions. A crucial step in the preparation was a minimum of 30 seconds vortex-mixing the sample vial twice prior to sampling. This ruled out within vial homogeneity as a source of variability in the measurement. It also indicated that the six measurements that fell outside the measurement uncertainty in the characterisation and homogeneity study were most likely normally distributed data points, i.e., random variation, that were among the 5 % of the population.

When measuring compounds that bind strongly with the matrix components, the homogeneity of samples and sampling technique could introduce a variability to the measurement. Therefore, aliquoting the blood from the sample vial process included a 30 second vortex mixing step twice, followed immediately by taking the aliquot by positioning the pipette tip a third way into the blood in the vial. This process proved to be enough to take a representative sample. However, within vial homogeneity could be a source of variability if less or no proper mixing of sample vials was performed. Therefore, the assessment of within vial homogeneity is recommended when measuring highly lipophilic molecules in complex biological matrices as whole blood.

4.3.7 Usage stability study: freeze thaw cycles and opened vial stability

Tacrolimus binding to red blood cells indicated the need to rule out that freeze and thaw cycles would not introduce variability in the data. During freeze/thaw cycles, red blood cells would be likely to rupture. However, this would not be necessarily equal every time which could result in variability in measurement where different samples would have different results. Hence, a study to evaluate the impact of freeze/thaw cycles as a source of variability in the measurement was performed. This is important for the end user as although in the RMP the blood is lysed anyway, but the efficiency of blood lysis could differ in routine method in hospital laboratories resulting in different levels of variability from different freeze/thaw cycles.

Moreover, based on the storage stability study (see Figure 4.8) that indicated the tacrolimus cCRM could be stored at 18 °C. The end users of the material could potentially store the cCRM vial in the fridge to use it again later after opening it and closing it. The user of the cCRM would most likely be a secondary reference standards manufacturer's laboratory or possibly a hospital laboratory. Hence, it was important to evaluate if such treatment of cCRM vials would introduce variability to the measurement. For the above reasons, this evaluation study was performed and was called a usage stability study.

Measurements made in this evaluation study (see Figure 4.11) demonstrated the material to be stable under the tested conditions because upon comparison with the reference sample vial which did not undergo any treatment, the results agree. The vials that were treated by five freeze/thaw

cycles and opening vials and storing them in the fridge at <7 °C for two weeks all agreed with the reference measurement that was assigned in the characterization study.

Both data of this study and the data of the within vial homogeneity study (Figure 4.10) show an overall negative bias to the reference value. These studies were among the studies that were performed using the new tacrolimus standard after identifying the issue with the acetonitrile purity. This could be a factor as to why this negative bias was observed, furthermore these studies were performed over 2 years after the material was produced and there was no data about the stability of the material over that period of time, although the literature (25) suggests tacrolimus is stable in whole blood but further stability testing would be required to assess the material stability. Nevertheless, in both studies' measurements of the QC material, ERM DA 110a tacrolimus spiked in blank whole blood, agreed with the certified reference value with percentage difference in measurement of 2.1 % and 1.4 % for the within vial homogeneity study and usage stability study, respectively (see Figure 4.10.B and Figure 4.11.B). Moreover, the measurement uncertainties of all QC measurements overlapped with that of the certified reference value of the QC material indicating study data reliability.

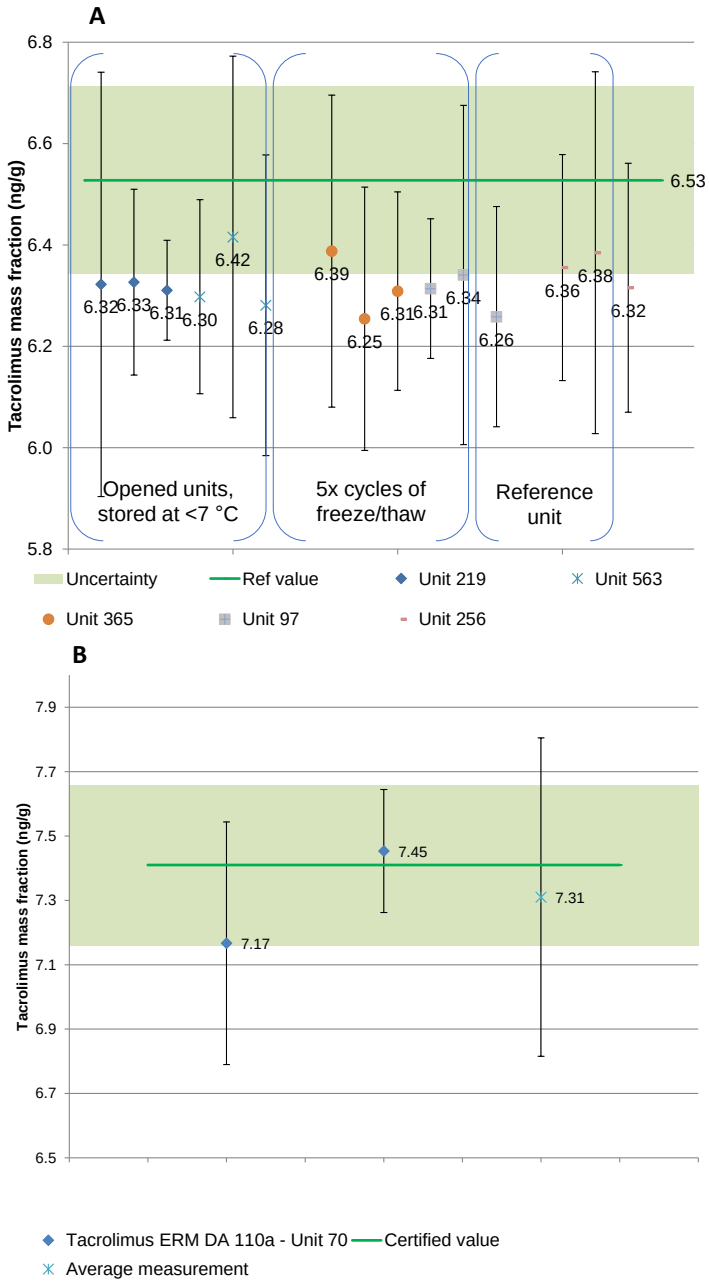


Figure 4.11 Usage stability study results: A) the five vials analysed in triplicate independent measurements, each tacrolimus candidate reference material. All measurements overlap with the reference value; B) the measurements of the duplicate independent measurements of the QC material vial (ERM DA 110a), the higher-order certified reference material of tacrolimus spiked in human whole blood. The result of the mean two measurements is in agreement with the certified reference value. Error bars = Expanded measurement uncertainty (95 % CI).

4.3.8 Blood density study

The measured density of the candidate reference material whole blood was 1.0503 ± 0.0019 g/mL at 22.83 °C. The density measurement calculation was achieved by dividing the mean mass of blood weighed from the five replicate units by the volume of the pipette tip that was measured by the pipette calibration. The uncertainty of the measurement was the combined measurement uncertainty of the variability of the mass measurements of the five replicate, the balance uncertainty and the pipette calibration uncertainty.

The accuracy of the measured density was derived from the two QC materials that were measured bracketing the blood measurements (see Figure 4.12). The measured density of the Distilled water density standard, that was measured immediately before the blood, was 1.00 ±0.001 g/mL. The measured density was only -0.02 % different from the certified reference value and its measurement uncertainty encompassed the certified value. Similarly, the measured density of the Dimethylphthalate density standard, which was of higher density than blood and was measured immediately after the blood, was 1.1903 ±0.0043 g/mL. The measured density was 0.15 % different from the certified reference value and its measurement uncertainty encompassed the certified the value. The percentage expanded measurement uncertainties were 0.1 % and 0.4 % for the Distilled water density standard and the Dimethylphthalate density standard, respectively. The density measurements were of high accuracy and precision because the values obtained for QC measurements were <0.5 %.

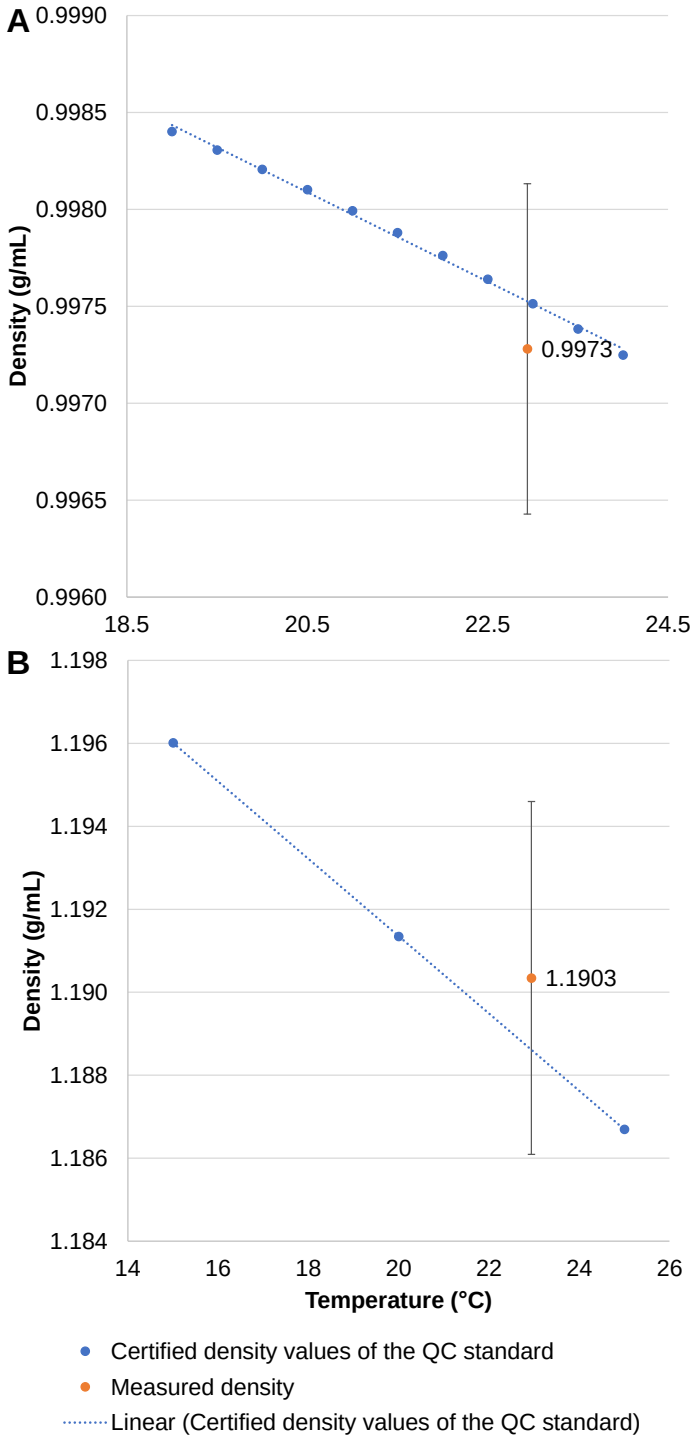


Figure 4.12 Density study QC materials results; two QC materials were measured before the whole blood measurements and after. A) Measured density of Fitzgerald Distilled water QC material 0.09973 ±0.0009 g/mL against the densities provided in the certificate. B) Measured density of Fitzgerald Dimethylphthalate density standard 1.1903 ± 0.0043 g/mL against the densities provided in the certificate. The expanded measurement uncertainty of the density was 0.4 %. Error bars = Expanded measurement uncertainty (95 % CI), (k=2).

The Dimethylphthalate density standard was of higher measurement uncertainty due to the high viscosity of the material which made accurate pipetting more difficult. Hence, this introduced more variability to the measurement than the water QC. However, this variability still resulted in a small measurement uncertainty estimate of 0.4 % at 95 % confidence interval. The percentage expanded measurement uncertainties for the cCRM blood and the Distilled water were 0.18 % and 0.15 %, respectively. Such small uncertainties reflect the accuracy in the pipetting and the accuracy and precision of the balance used.

To report the density of the blood material at different temperatures, a coefficient of the thermal expansion of blood would need to be used as per Hinghofer-Szalkay *et al.* (26). The accuracy of the volume taken by the pipetting process was based on the calibration of the pipette by water using the XP205 balance in a controlled and measured temperature laboratory as well as monitored air pressure. However, the viscosity of water is less than that of blood, which was not accounted for in the process because the viscosity of blood is dependent on different conditions including haematocrit and temperature (27). For the purposes of reporting the mass fraction of the material as mass concentration the obtained density was sufficient and any variability in the density measurement would be incorporated in the measurement uncertainty budget of the final uncertainty assigned to the material.

Using the approach, described in the paragraph above, to measure the density of biological materials may be applicable to other matrices and compounds such as plasma and METs. Although the density study process is

lengthy and laborious, it ensures a more accurate density rather than relying on the average human blood density value. This is because the density of blood is affected by many factors, for example the density of high density lipoproteins impact the overall density of plasma (28). Blood platelets count and other cellular make up of blood could impact the density of the blood as well (26). Other factors that could impact density include temperature, air pressure. Additionally, certain pathological conditions impact the blood density (29). From a quantitative measurement perspective, whenever a mass fraction (i.e., m/m e.g., ng/g) would be reported as mass concentration (i.e., m/v e.g., ng/mL) an evaluation of the density would be important to incorporate the variability introduced in the conversion to the final expanded measurement uncertainty.

4.3.9 The candidate reference value and measurement uncertainty

The above reported studies make up most of the analytical studies required to characterise and certify a higher-order reference material. The remaining important piece of analytical work is the evaluation of the commutability of the material (see 1.2.2 for the commutability definition). This involves other laboratories analysing the material using different techniques and methods. The commutability study is not part of this thesis because of a confidentiality of data agreement between the participating laboratories and the NML. Hence, in this section an estimated measurement uncertainty of the cCRM has been calculated. Upon the completion of the commutability study the NML Statistics Team would assign the final measurement and its uncertainty. This will then be the reference value which appears on the cCRM a certificate of analysis. This certificate would be part of the application submitted to the

JCTLM database to enlist the cCRM as a higher-order certified reference material.

For the assignment of the estimated measurement uncertainty to the material a combined uncertainty was calculated, see Table 4.6. The uncertainty assignment approach applied for the material was a top-down approach where the measurements from the different analytical studies that were performed and their standard uncertainties were combined using Equation 3.1.. Applying a top-down approach to the uncertainty assignment for the material was selected although the Guide to the Expression of Uncertainty in Measurement (GUM) recommends a bottom-up approach. This is because the GUM uses a bottom-up approach, which would provide an idea of the level of uncertainty associated with a single measurement rather than a method and/or a material (30,31). While the bottom-up approach characterises a single measurement performed on a single day, the top-down approach considers matrix-associated errors and the actual day-to-day variability that is observed in a laboratory. However, the way the measurement uncertainty assignment to the cCRM combined both bottom-up and top-down approaches where: a) every single measurement acquired by the LC-DEM-IDMS was assigned a measurement uncertainty using the bottom-up approach; and b) the final assignment of the measurement uncertainty included both the uncertainty of the method and of the material covering most of the spectrum of sources of variability possible using the top-down approach.

Using the top-down approach, the variances of each of the mean measurements achieved in the studies described earlier were combined to

estimate the estimated expanded uncertainty. The estimated reference value of the cCRM was calculated to be 6.43 ± 0.22 ng/g (95 % CI, k = 2), see

Table 4.6. The mean measurement for each of the studies was the mean of measurement of all the blends of the material's vials that were analysed in each study. The only vials that were not included in the mean measurement of the stability study were the vials that were stored at temperatures above 18 °C. The use of the number of vials provides a more comprehensive estimation as it encompasses the sources of variability including the matrix, the bottling of the material, the day-to-day variability.

Table 4.6 The assignment of the estimated reference value and its measurement uncertainty

Section in text	Description	Average measurement (ng/g)	u (ng/g)	u^2 $(ng/g)^2$
4.3.3	Characterisation and homogeneity study	6.53	0.053	0.00276
4.3.4	Storage temperature stability study	6.56	0.047	0.00217
4.3.6	Within unit homogeneity study	6.31	0.073	0.00532
4.3.7	Usage stability study	6.32	0.046	0.00212
Calculated values	Indicative reference value		6.43	
	Combined uncertainty		0.11	
	Expanded uncertainty (95 % CI, k=2)		0.22	

4.4 Conclusions

An investigation of sources of variability in the process of characterising a higher-order candidate reference material was performed. It was found that a comprehensive characterisation and homogeneity study was required to be able to assess the variability in the material and in the measurement procedures used. It was demonstrated that for the measurement of tacrolimus, a large lipophilic small molecule, an array of experiments was required to ensure no additional sources of variability in the measurement were introduced. Sampling the whole blood vials proved to be an aspect to be assessed to ensure a representative sample is aliquoted from the blood vial. This was not the case for the METs in plasma which makes the initial expectation that these two sets of compounds (i.e. small polar and large lipophilic) behave differently and could introduce different sources of variability as well as common sources across the compounds classes, such as stability. Tacrolimus in patient blood exhibited stability under different storage temperatures and different sample handling conditions. This could indicate that large molecules with high lipophilicity could be stable and have less variability in measurement due to sample handling and storage.

Another aspect of discrepancy in sources of variability between the two classes of molecules is that lipophilic larger small molecules with complex structures proved to require a thorough investigation of standards in solvents. Tacrolimus stability in standards in solvent proved to be an issue of concern requiring the use of ultra-pure acetonitrile for the preparation of standards. This required extensive work to know the source of the issue and eliminate it to be able to perform the measurement.

Finally, a double exact-matched isotope dilution LC-MS based reference method was applied to characterise a candidate higher-order reference material of tacrolimus in pooled patient blood and a reference value of 6.43 ± 0.22 ng/g was assigned and homogeneity, stability of the material were assessed.

Therefore, the hypothesis of this chapter holds true as the objectives were achieved and all the major sources of measurement uncertainty for the measurements of clinically relevant small molecules at the upper end of the molecular weight range were determined and the measurement uncertainty budget was calculated.

4.5 Novelty and value

Currently, the commercially available reference materials and standards (primary, secondary and higher-order) are all composed of tacrolimus spiked into blank human blood. This work shows that traceable characterisation of such a material containing a large, non-polar molecule is possible. Particular challenges that were overcome include the preparation and validation of standards in solvent which constitute the cornerstone of the traceable measurement. This work represents the first higher-order candidate reference material of tacrolimus in pooled patient blood intended for submission to the JCTLM database. The material provides an accurate tool to assess method accuracy as it is a material made of patient blood consequently it contains tacrolimus metabolites. This is especially important for the evaluation of immunoassay-based methods because they are more prone to cross-reactivity with the metabolites resulting in positive biased measurements. Such biased measurements would result in the clinician incorrectly adjusting

the tacrolimus dosage to a lower dose which could risk an organ rejection episode. Moreover, it is of great value for hospital laboratories, secondary reference standards and materials producer and in-vitro diagnostic (IVD) kit manufacturers as it provides a traceability and a tool to assign measurement uncertainty of their method and materials.

4.6 Future work

By the time the official certification of the material is all done and the release of the material to the market, it would be over 18 months, therefore a quick check of the material stability would be helpful to evaluate the long-term stability of the material. The commutability of the material needs to be validated by collaborating with other laboratories using different analytical methods both LC-MS based methods and immunoassay kits. Further work is required to better understand the degradation of tacrolimus in standards in solvent and its contributory factors. Subsequently, additional analytical measurements of the material are required after the standards in solvent have been validated to assess the source of the negative bias observed in the later studies.

4.7 References

1. Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardization and harmonization of clinical laboratory tests123. Am J Clin Nutr. 2016 Sep;104(Suppl 3):907S-912S.
2. Koop DR, Bleyle LA, Munar M, Cherala G, Al-Uzri A. Analysis of tacrolimus and creatinine from a single dried blood spot using liquid chromatography tandem mass spectrometry. J Chromatogr B. 2013 May 1;926:54–61.
3. Venkataramanan R, Swaminathan A, Prasad T, Jain A, Zuckerman S, Warty V, et al. Clinical pharmacokinetics of tacrolimus. Clin Pharmacokinet. 1995 Dec;29(6):404–30.
4. Wallemacq P, Armstrong VW, Brunet M, Haufroid V, Holt DW, Johnston A, et al. Opportunities to optimize tacrolimus therapy in solid organ transplantation: report of the European consensus conference. Ther Drug Monit. 2009 Apr;31(2):139–52.

5. T3DB: Tacrolimus [Internet]. [cited 2019 Nov 25]. Available from: http://www.t3db.ca/toxins/T3D4732
6. Annesley TM, McKeown DA, Holt DW, Mussell C, Champarnaud E, Harter L, et al. Standardization of LC-MS for Therapeutic Drug Monitoring of Tacrolimus. Clin Chem. 2013 Jul 31;clinchem.2013.209114.
7. Levine DM, Maine GT, Armbruster DA, Mussell C, Buchholz C, O'Connor G, et al. The Need for Standardization of Tacrolimus Assays. Clin Chem. 2011 Jan 12;57(12):1739–47.
8. Barau C, Frangie C, Goujard C, Tribut O, Parant F, Taburet A-M, et al. Falsely elevated whole blood tacrolimus concentrations due to interference in an affinity column-mediated immunoassay method on Xpand dimension. Ther Drug Monit. 2009 Apr;31(2):267–8.
9. Amann S, Parker TS, Levine DM. Evaluation of 2 Immunoassays for Monitoring Low Blood Levels of Tacrolimus. Ther Drug Monit. 2009 Apr;31(2):273.
10. Martín BB, Marquet P, Ferrer JM, Puig BC, Bennasar AB, Prieto MR, et al. Rheumatoid Factor Interference in a Tacrolimus Immunoassay. Ther Drug Monit. 2009 Dec;31(6):743.
11. Moscato D, Nonnato A, Adamo R, Vancheri M, Caropreso A. Therapeutic monitoring of Tacrolimus: Aberrant results by an immunoassay with automated pretreatment. Clin Chim Acta. 2010 Jan 4;411(1):77–80.
12. Hermida J, Tutor JC. Falsely Increased Blood Tacrolimus Concentrations Using the ACMIA Assay Due to Circulating Endogenous Antibodies in a Liver Transplant Recipient: A Tentative Approach to Obtaining Reliable Results. Ther Drug Monit. 2009 Apr;31(2):269.
13. Rostaing Lionel, Cointault Olivier, Marquet Pierre, Josse Anne-Gaelle, Lavit Michel, Saint-Marcoux Franck, et al. Falsely elevated whole-blood tacrolimus concentrations in a kidney-transplant patient: potential hazards. Transpl Int. 2009 Sep 15;23(2):227–30.
14. Greaves RF. A Guide to Harmonisation and Standardisation of Measurands Determined by Liquid Chromatography – Tandem Mass Spectrometry in Routine Clinical Biochemistry. Clin Biochem Rev. 2012 Nov;33(4):123–32.
15. Baker R, Mark P, Patel R, Stevens K. Clinical Practice Guideline Post-Operative Care in the Kidney Transplant Recipient [Internet]. [cited 2019 Jan 30]. Available from: https://bmcnephrol.biomedcentral.com/track/pdf/10.1186/s12882-017-0553-2.pdf
16. Immunosuppressive therapy for kidney transplant in adults | Guidance and guidelines | NICE [Internet]. [cited 2019 Jan 30]. Available from: https://www.nice.org.uk/guidance/ta481
17. McMaster P, Mirza DF, Ismail T, Vennarecci G, Patapis P, Mayer AD. Therapeutic drug monitoring of tacrolimus in clinical transplantation. Ther Drug Monit. 1995 Dec;17(6):602–5.
18. Bedson P. Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS) [Internet]. 2002 [cited 2018 Jun 25]. Available from: http://pubs.rsc.org/en/Content/eBook/978-0-85404-418-4 - pages 1-34.
19. Alonso J, Gonzalez P. Isotope Dilution Mass Spectrometry 2013 [cited 2019 Nov 26]. Available from: https://pubs.rsc.org/en/content/ebook/978-1-84973-333-5 - pages 10-150
20. Appendix A: Statistical Tables. In: Using Statistical Methods for Water Quality Management [Internet]. John Wiley & Sons, Ltd; 2005 [cited 2020 Dec 2]. p. 305–13. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471733199.app1
21. On-line barograph [Internet]. [cited 2019 Jan 10]. Available from: http://resource.npl.co.uk/pressure/pressure.html

22. Watson JT, Sparkman OD. Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation, Fourth Edition [Internet]. 4th ed. John Wiley and Sons; 2008. Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470516898 - pages 53-228 and 639-688
23. French D. Chapter Five - Advances in Clinical Mass Spectrometry. In: Makowski GS, editor. Advances in Clinical Chemistry [Internet]. Elsevier; 2017 [cited 2019 Nov 29]. pages. 98-153. Available from: http://www.sciencedirect.com/science/article/pii/S0065242316300828
24. Bharma A, Mussell C. Proposed Structure of the Methanolysis Conversion Product of Tacrolimus (FK-506) and Determination of the Rate of Reaction. London, UK: Forensic Science & Drug Monitoring, King's College London, 150 Stamford Road, SE1 9NH; 2011 Jul p. 15.
25. Freeman DJ, Stawecki M, Howson B. Stability of FK 506 in whole blood samples. Ther Drug Monit. 1995 Jun;17(3):266–7.
26. Hinghofer-Szalkay H, Greenleaf JE. Continuous monitoring of blood volume changes in humans. J Appl Physiol Bethesda Md 1985. 1987 Sep;63(3):1003–7.
27. Rand PW, Lacombe E, Hunt HE, Austin WH. Viscosity of normal human blood under normothermic and hypothermic conditions. J Appl Physiol. 1964 Jan 1;19(1):117–22.
28. Bowry VW, Stanley KK, Stocker R. High density lipoprotein is the major carrier of lipid hydroperoxides in human blood plasma from fasting donors. Proc Natl Acad Sci. 1992 Nov 1;89(21):10316–20.
29. Kenner T. The measurement of blood density and its meaning. Basic Res Cardiol. 1989 Apr;84(2):111–24.
30. Martinello F, Snoj N, Skitek M, Jerin A. The top-down approach to measurement uncertainty: which formula should we use in laboratory medicine? Biochem Medica [Internet]. 2020 Jun 15 [cited 2020 Dec 7];30(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7138004/
31. Lee JH, Choi J-H, Youn JS, Cha YJ, Song W, Park AJ. Comparison between bottom-up and top-down approaches in the estimation of measurement uncertainty. Clin Chem Lab Med. 2015 Jun;53(7):1025–32.

5 Sources of variability in higher-order measurement of large non-polar molecules using a novel micro-sampling device

5.1 Introduction

This chapter focuses on the sampling stage of the analytical process. In particular, the work described here focused on assessing a new micro-sampling device and assigning measurement uncertainty to this novel technique possibly to improve its reliability and measure its precision. It demonstrated the importance of using RMPs and CRMs for the evaluation of new technologies and how these can aid the understanding and assign measurement uncertainty of novel techniques.

Micro-sampling is a term often used in conjunction with techniques that are less invasive and by which biological samples can be collected in much lower volumes than standard sampling (< 50 µL) for quantitative analysis (1–3). One of the common less invasive sampling approaches is a finger prick rather than a venous blood draw from the cubital vein in the forearm of the patient. In recent years, there has been an increased interest in micro-sampling which has been partially enabled by the advancement in analytical techniques with enhanced sensitivity, selectivity and robustness (4). Micro-sampling of blood offers numerous advantages over conventional venous collection such as being a less invasive approach for patients. Due to their small size and ease of storage and handling, dried blood microsamples have been considered to be more stable and less hazardous than venous blood sample (5). Consequently there is a reduced need for complex logistics such as

refrigeration during storage and transportation (5–7). On the other hand, despite the added complexity in the collection and storage of traditional blood samples, the question of comparability of venous blood and capillary venous blood remains an important consideration.

A number of micro-sampling techniques such as capillary micro-sampling (CMS) and dried blood spots (DBS) have been developed and are being increasingly reported in literature (8). More recently, the volumetric absorptive micro-sampling (VAMS) or Mitra® device was developed as a mean of collecting dried blood samples to avoid the frequently reported issues related to variable patient haematocrit which can lead to quantitative bias associated with DBS (7). Moreover, with DBS, the automated systems are set to make a standard punch in the centre of the circle drawn on the collection paper although the actual blood sample may not be centred and symmetrical inside the circle on the paper. Hence, the DBS punch would sometimes be in the centre of the sample blood or to the side of it. The amount of sample blood taken in the punch is not defined which introduces variability to the measurement. VAMS on the other hand provide accurately predefined sample volume which is extracted as a whole in the analysis (7,8).

DBS has been widely reported as a micro-sampling technique (2,4). It is a form of dried matrix collection where blood is collected from a finger prick as a spot onto a filter paper, which is dried and then analytes extracted and analysed (1,2,6,7). However, its use has been limited due to several reasons; the most widely reported being the effect of haematocrit on the blood spot collected (9). Haematocrit is defined by the National Cancer Institute as *"The amount of whole blood that is made up of red blood cells. It depends on the*

number and size of red blood cells" (10). Haematocrit affects the viscosity and spread of blood and may lead to non-representative sample collection at different haematocrit values with subsequently significant assay bias depending on the target analyte and the difference in haematocrit between blood in calibrants and in samples (6,9,11–13).

The VAMS device shown in Figure 5.1 consists of a ribbed plastic handle with an absorbent tip attached, which wicks up a fixed amount (either 10 µL or 20 µL formats currently available) of blood by capillary action (7,11). It has been reported to absorb the defined volume of blood with a coefficient of variation of less than 10 % (7,11,13).

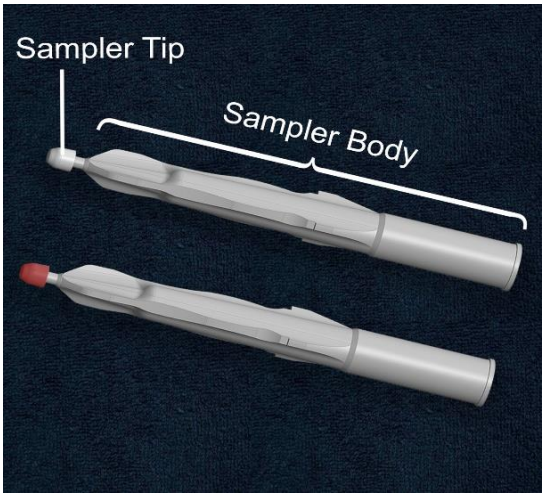


Figure 5.1 Computer generated graphic of unused Mitra®/VAMS device (top) with sampler body and sampler tip components indicated. Used device (bottom) showing tip after sampling blood and left to dry. Figure drawn with permission from Neoteryx (14).

A benefit of micro-sampling is the ability to obtain samples from peripheral sites in the body such as the finger or heel of the foot, which would be more convenient to be used at home or for remote drug-testing (6). Patients who undergo therapeutic drug monitoring (TDM) may particularly benefit from this

as it reduces the need for travel or the waiting periods associated with hospital visits (5). Patients with reduced blood availability such as geriatrics and young children may also benefit from micro-sampling (6).

TDM of tacrolimus is typically performed by measuring trough blood concentrations as stated in the National Institute for Health and Care Excellence (NICE) guidelines (15). The UK Renal Association (16) recommends tacrolimus is analysed through venous sampling as frequently as three times weekly in the first few months post-transplantation (5,17,18). This process may become tedious, causing poor patient compliance to hospital visits and resulting in inappropriate profiling of blood concentrations (5). Additionally, the collection of a venous blood sample requires a phlebotomist, a cost that could be avoided by using at-home micro-sampling. Given the major impact of any failure in performing TDM reliably, incorporation of robust micro-sampling procedures could provide a more convenient method for the TDM of tacrolimus, and potentially increase compliance and/or reduce the indirect associated economic costs such as interim dialysis in tacrolimus supratherapeutic concentration induced kidney failure (5,19,20).

The aim of this work was to demonstrate that the VAMS devices could be analytically suitable to perform tacrolimus TDM, prior to being used in clinical studies with patients and to evaluate sources of variability in the measurement using these devices. The VAMS approach was selected over the dried blood spots for the evaluation for two reasons: a) it overcomes the variability observed in dried blood spots due to the differences in haematocrit among patients and; b) from a measurement perspective, the VAMS could provide an

accurate measurement of the amount of blood taken which could be a source of variability in measurement.

The existing DEM-IDMS methodology which uses venous blood sample aliquots of 300 µL (described in Chapter 4 in 4.2.3.1 and 4.2.4) to assign reference values to inter-laboratory comparisons studies (21) was adapted and optimised to be used with 20 µL VAMS devices for this preliminary evaluation. This study was the first to demonstrate that VAMS could be a suitable technique to use for TDM of immunosuppressant drugs.

Through the adaptation of an existing higher-order RMP to evaluate the Mitra®/VAMS samplers, a better understanding of the sources of variability in the measurement using these samplers was attained, with a view to improving micro-sampling based clinical analysis generally in terms of accuracy, precision and reliability of using such small sample size. The relevance of the use of tacrolimus as an example large nonpolar molecule was based on the need for a compound to be a) measured in whole blood samples and b) lipophilic and bound to erythrocytes which may or may not be an issue using the Mitra®/VAMS devices. Understanding the sources of variability in the measurements made using Mitra®/VAMS devices potentially applicable to any other larger size small molecules that would be analysed in whole blood samples.

The complexity of whole blood as the matrix for analysis that was investigated in Chapter 4 built a base for this work. Moreover, the better understanding of measurement uncertainty components and sources of variability in the measurement of larger small molecules in whole blood enabled investigating

sources of variability specific to the use of small sample volumes of blood i.e., 20 µL vs 300 µL.

5.1.1 Hypothesis and objectives

In line with the aim of this thesis expressed in Chapter 1 (section 1.13.2.1), the hypothesis for this chapter is that "the VAMS devices could provide traceable measurements with measurement uncertainties that were fit for routine analysis (<15%) when using the adapted RMP from Chapter 4".

In order to test this hypothesis, the following objectives were set:

- Evaluate the accuracy and precision of the mass of sample blood taken by the VAMS devices.
- Assess the use of IDMS linear calibration versus DEM-IDMS calibration using VAMS.
- Analyse reference materials and patient pooled samples that were assigned reference measurements using the VAMS samplers.

5.2 Experimental

5.2.1 Reagents and standards

The same reagents and standards used for the work described in Chapter 4 (section 4.2.2) were used to conduct this work.

5.2.2 Blood samples and procedures

For the preparation of matrix matched calibration blends, blank pooled human blood supplied by Analytical Services (ASI: St. George's Hospital, London,

UK) was used. All blends were prepared gravimetrically in 2 mL screw capped clear silanised glass vials from Agilent (California, USA).

Sample blends were prepared using five different samples. The samples that were used to conduct this work were a) pooled tacrolimus patient whole blood candidate reference material (cCRM) (6.53 ± 0.19 ng/g at the 95% CI), characterised in Chapter 4 and b) four pooled anonymised patient blood samples (P-02, P-04, P-06 and P-09). These samples were kindly provided by Waters™ Corp who commissioned ASI to produce these samples which were also analysed by the participating laboratories for the global inter-laboratory comparison study used to demonstrate the reproducibility of the Waters MassTrak™ kit for TDM of tacrolimus (21). These samples were assigned traceable reference measurements by LGC as part of the same study, see 0. All blood materials used were produced at ASI by pooling patient blood collected in EDTA preserved collection tubes. The use of the blood samples used in this study were approved by the LGC Bioethics Committee.

5.2.3 Instrumentation

The LC-MS used is described in section 4.2.3.1. The MS instrument used was the Thermo Scientific TSQ Vantage Tandem Mass Spectrometer with a TLX-1 liquid chromatography system (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used.

5.2.3.1 Analytical sequence

For linear calibration analysis, each sample blend was injected in triplicate, in 3 sets (randomised order in each set) which were bracketed by a set of calibrators (in total each calibrator was injected 4 times). For DEM-IDMS

calibration, each sample blend was analysed five times, with each sample blend injection being bracketed by its associated calibration blend (analysed five times). Quantification was performed using the peak area ratio of the product ion for tacrolimus (*m/z* 826.5) to its SILIS (*m/z* 831.5). The average of the five measurements was then used as the measured value, the relative standard deviation of the 5 LC-MS measurements was a component of the measurement uncertainty budget.

5.2.3.2 Other instruments

Other instrumentation used in this work are found in 4.2.3.2, in addition to the automated SPE Extrahera™ system from Biotage (Uppsala, Sweden).

5.2.4 Sample preparation

5.2.4.1 Preparation of VAMS dried blood samples

Mitra®/VAMS devices, 20 µL format, were supplied by Neoteryx (Torrance, USA). Empty, capped, 2 mL silanised glass vials and the unused blank VAMS tips (removed by pulling from the device) were pre-weighed. After vortex mixing the blood sample tube for 30 seconds, a 30 µL aliquot of blood was transferred with a pipette on to a polystyrene weighing boat surface to simulate a blood drop on the skin. The Mitra®/VAMS sampler was placed at a 45° angle to the blood surface (as instructed by manufacturer) during the wicking stage. Directly after the wicking process, the sampler tips were ejected into the pre-weighed 2 mL clear silanised glass vials by pushing the tip with a disposable stick with minimal contact with the tip. This was followed by a weighing step to obtain the mass of blood wicked up by the tip. The tips were

dried overnight under ambient laboratory conditions in free circulation laboratory air (20 °C, 69 % relative humidity, monitored but not controlled).

In the evaluation described in this chapter, two approaches have been used for calibration: a double exact matched IDMS (DEM-IDMS) procedure and conventional linear regression isotope dilution mass spectrometry (IDMS).

Three sample blends and corresponding calibration blends were prepared in triplicate from 3 different cCRM vials and analysed using the DEM-IDMS procedure. The DEM-IDMS procedure was described in Chapter 4 (see 4.2.3.1 and 4.2.4). For linear calibration, a six-point concentration line was constructed covering the concentration range of 0-20 ng/g. The sample blends were prepared in triplicate using a new Mitra®/VAMS device for each calibrant.

Isotopically-labelled tacrolimus standard solution was added to all blends except blank blends; B-B, B+N and B-S (described later) at a mass fraction of 6.53 ng/g and 7 ng/g for DEM-IDMS analysis and linear calibration IDMS, respectively. Tacrolimus standard solution was added to blank blood to prepare the blank blend (B+N) and the calibration blends. Calibrants were spiked with tacrolimus standards at 6.53 ng/g for the DEM-IDMS analysis. Calibration blends for the linear calibration IDMS, tacrolimus standard was spiked at 0, 4, 8, 12, 16 & 20 ng/g. Calibration and sample blends were prepared gravimetrically. The weights recorded during the sample blends preparation and which were used in the subsequent calculations of mass fractions were;

- the empty vial
- the blank VAMS tip

- the weight of the vial plus tip with blood
- the weight of vials plus tip and blood plus water
- the weight of vial after addition of internal standard
- For calibration blends, the weight of the tacrolimus standard added was also recorded.

5.2.4.2 Extraction of blood from VAMS tips

The first step in the extraction of the dried tips was to immerse them in 850 µL water for 15-20 minutes prior to spiking blood/water mix with internal standard. After blends preparation, 500 µL of 0.04 M zinc sulphate methanolic solution (1:4 methanol:water v/v), that had been freshly prepared on the day of analysis, was added to each vial for lysis of erythrocytes, protein precipitation and extraction of tacrolimus. Each vial was vortex mixed for 30 seconds. Sample and calibration blends were then mixed for 30 minutes using a Stuart mini SSM1 orbital shaker (Staffordshire, UK), set to 300 rpm, prior to storing the extracts overnight to equilibrate in a fridge (≤9 ˚C)

5.2.4.3 Solid phase extraction

The same SPE method was used as described in 4.2.4.3. However, an automated SPE Extrahera™ system was used to perform the SPE that applies positive pressure to the SPE plate instead of manual operation under negative pressure as described in Chapter 4. The conditions of the automated SPE system were optimised manually and are described in Table 5.1.

Table 5.1 Extrahera™ conditions

SPE step	Pressure (bar)	Duration (s)
Conditioning step 100 µL methanol	3	45
Conditioning step 300 µL water	3	200
Loading step	3	200
Wash steps 3x 500 µL 4:6 (v/v) methanol:water	5	200
Elution step 250 µL methanol	3	200

5.2.5 Measurement equation & uncertainty calculation

See section (2.2.4).

5.2.6 Quality control

5.2.6.1 Analysis of control material

With each batch, one vial of the higher-order CRM tacrolimus in whole human blood ERM DA110a was analysed in duplicate: two individual measurements from two individual aliquots. One replicate was injected at the start of the sequence prior to the analysis of the samples and the second at the end of each sequence.

5.2.6.2 Blanks

A single blend of each of the blank blends described below was prepared and analysed twice by LC-MS, at the beginning and the end of each LC-MS sequence to check for interferences in the SRM transitions:

- B-S: One candidate reference material sample without the addition of the SILIS.

- B+N: Blank blood with the addition of tacrolimus only.
- B+L: Blank blood with the addition of isotopically-labelled tacrolimus only.
- B-B: Blank blood without any addition.

5.3 Results & Discussion

5.3.1 Blood sampling and dry blood extraction

The method used for the analysis was adapted from the reference measurement procedure used at LGC for assigning reference values to tacrolimus in whole blood CRMs and inter-laboratory comparison samples (21,22). At the start the methodology was used without significant modification. However, the VAMS samplers gave both significant positive and negative bias initially ranging from -20 % to +20 %. The negative bias was postulated to be as a result of precipitation of protein and cellular matter on the exterior of the tip preventing complete extraction of tacrolimus from within the tip, see Figure 5.2. Since tacrolimus is highly bound to red blood cells, and the SILIS was added to the blend solution rather than loaded onto the tip itself (prior to wicking the blood sample), complete extraction of the compound is critical to achieving accurate measurement (21,22).

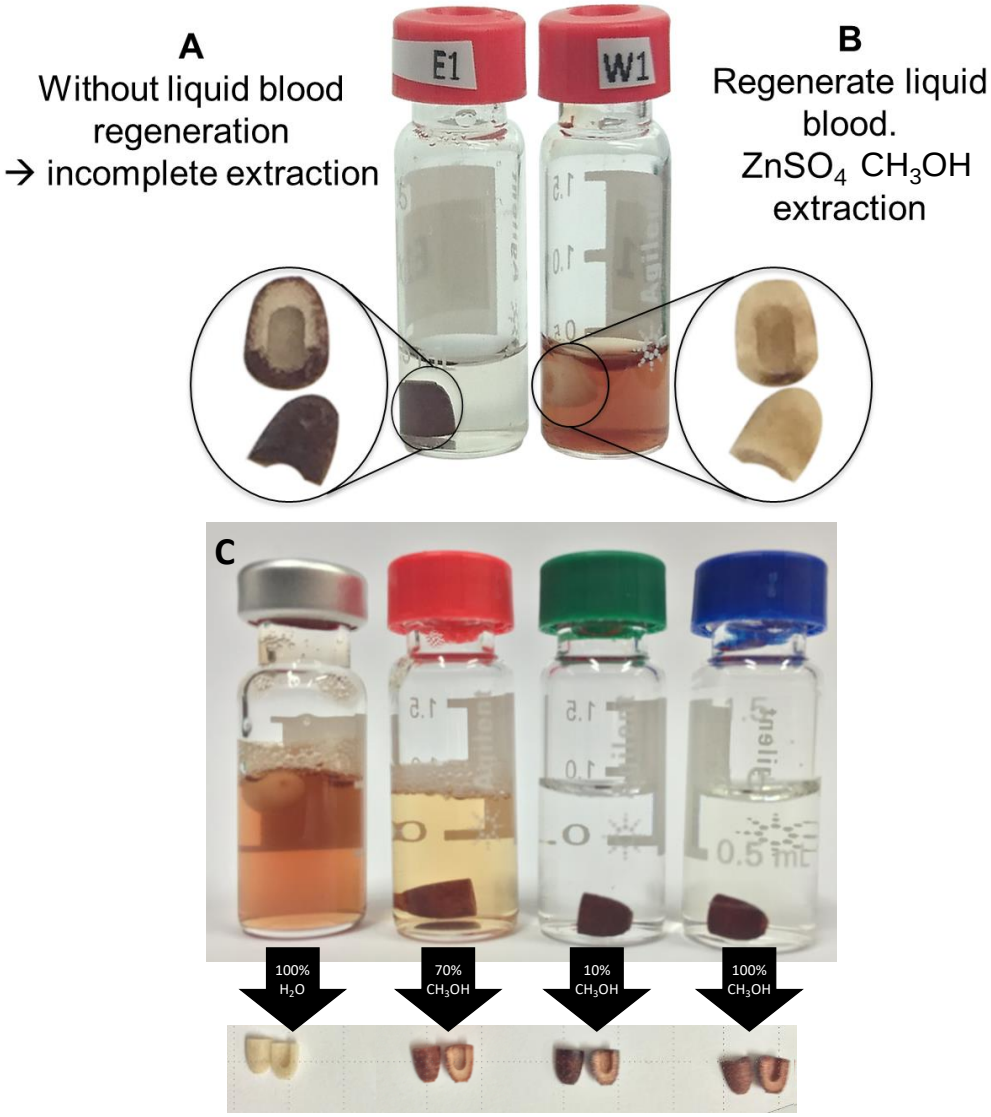


Figure 5.2 Extraction of dried blood from the Mitra®/VAMS devices. The tips of the Mitra®/VAMS devices were cut lengthwise after analysis for visual inspection. A) Extraction using methanolic solution of zinc sulphate (85:15 methanol:water). Significant cellular matter was observed on each tip post analysis. B) Extraction using water to generate a blood/water mix prior to addition of the methanolic solution of zinc sulphate. C) Visual comparison of the extraction by a range of extraction solutions.

The first step to developing the method was to assess the extraction of the dried blood off the absorptive tip. This was a crucial step because the extraction from the tip could be a major source of variability in the amount of tacrolimus released. The material which the absorptive tip is not disclosed by the manufacturer hence a variety of solvents was needed to assess the extraction.

A brief evaluation of various solvents with varying organic strength was performed using 100 % methanol, 70:30 methanol:water, 10:90 methanol:water and 100 % water. Water resulted in the highest release of blood material from the dry tips without protein precipitation based on visual observation of the tip and colour of the solvent. Water with extensive mixing was therefore further investigated for the extraction of sample blends, see Figure 5.2. This indicated that the material that the wicking tip was made of is not hydrophobic and is highly wettable. Such material is crucial to enable the absorption of blood and is inline with what is required for lipophilic compounds as well because the blood components (e.g., red blood cells) would act as a carrier of these compounds as is the case of tacrolimus. This enables the device to act in a very similar way to intravenous blood sampling process where the blood is trapped at such small volume in the wicking tip rather than the vacutainers as with venous blood sampling.

Blood sampling from a hydrophobic flat surface rather than by directly dipping/wicking from the originally supplied sample vessel was a critical change made to the original method. This mimicked as close as possible, the typical way the VAMS samplers would be used for home sampling from blood droplet generated from a finger prick. It also prevented the wicking of

excessive blood by the sampler tips when dipped directly into the blood tube which was reported in the past to cause significant assay bias (8,9).

Some MS parameters were also further optimised to increase assay sensitivity which was required due to the reduced initial sample volume being analysed. Historical data at the NML and data obtained in Chapter 4 had shown that selectivity was not an issue with tacrolimus. Therefore, to increase signal response, only one SRM transition was monitored for both tacrolimus and the stable SILIS for this study because it allows more scanning time for the main SRM transition hence better characterisation of the peak and better signal. The injection volume was also increased from 10 µL in the existing method to 20 µL. As a result, the sensitivity increased where the peak areas were doubled by the injection volume increase and no distortion in peak shape was observed.

After the cRM vials were analysed with the optimised method and results in agreement with the assigned reference mass fraction were obtained, patient samples were analysed by linear calibration IDMS – in a similar way to how it was envisaged they would be analysed in a routine clinical laboratory.

5.3.2 Blank blends analysis

The following blank blends were analysed:

- B-B: Blank blood without any addition.
- B-S: One candidate reference material sample without the addition of the SILIS.
- B+N: Blank blood with the addition of tacrolimus standard only.

- B+L: Blank blood with the addition of isotopically-labelled tacrolimus standard only.

An example chromatogram of most of the analysed blank blends is shown in Figure 5.3. No chromatographic peak was observed in the SRM transitions of tacrolimus and the SILIS when analysing the control blank blends in both the DEM-IDMS and linear calibration IDMS experiments. This indicated several things including: (a) no carry-over in the chromatographic method; (b) no interference from the blank blood used for matrix matching in calibration blends preparation; (c) no contribution of the SILIS precursor mass into the quantification SRM of tacrolimus; (d) no contribution of the tacrolimus precursor mass into the quantification SRM of the SILIS and (e) no interference from the sample cCRM into the SILIS SRM. All of these validated the method's specificity, see Figure 5.3.

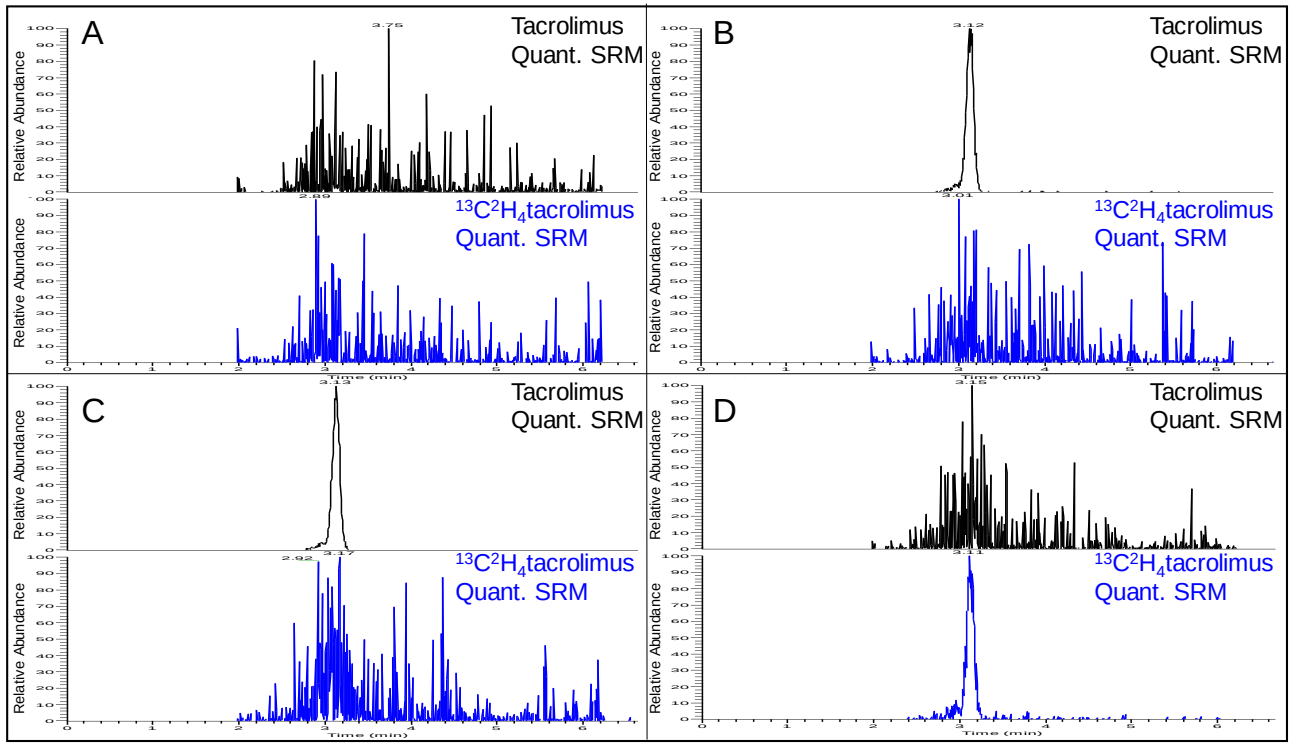


Figure 5.3 Blank blends analysis: A) B-B: Blank blood without any addition; B) B-S: One candidate reference material sample without the addition of the SILIS; C) B+N: Blank blood with the addition of tacrolimus only; D) B+L: Blank blood with the addition of isotopically-labelled tacrolimus only.

5.3.3 cCRM analysis by two calibration approaches

Figure 5.4 summarises the data obtained from analysing tacrolimus in patient blood cCRM by DEM-IDMS (Figure 5.4.A) and by linear calibration IDMS (Figure 5.4.B). Excellent precision was achieved in both methods; in the case of DEM-IDMS calibration %RSD of 1.9% was achieved between all Mitra®/VAMS devices (3 cRM vials analysed in triplicate), as for linear calibration IDMS %RSD of 2.7% was achieved for all Mitra®/VAMS devices used (single analysis of 3 cRM vials).

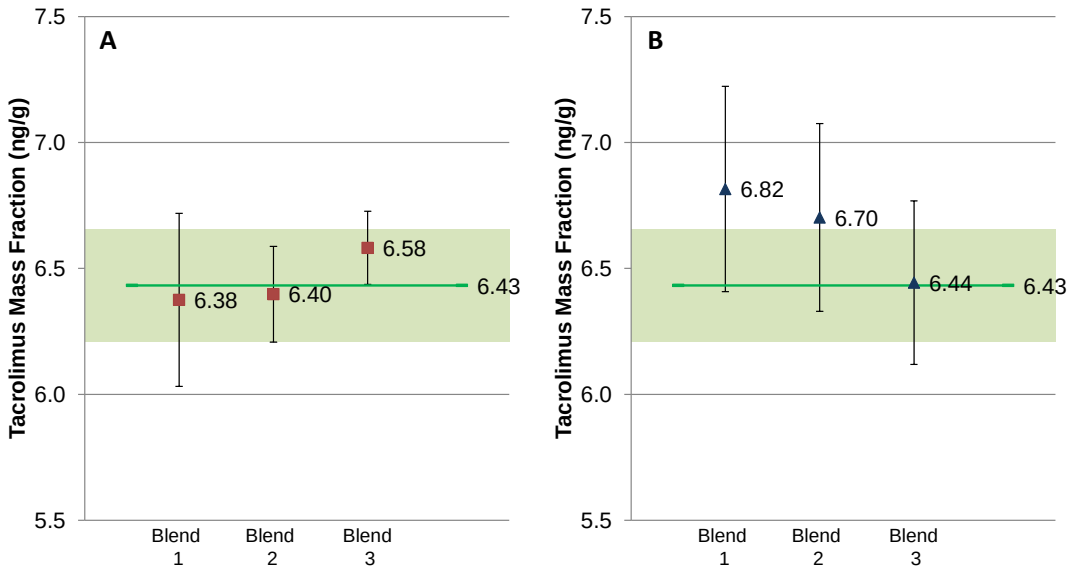


Figure 5.4 Analysis of candidate reference material of tacrolimus in patient blood. Each data point is the average of 5 LC-MS replicate analysis of a single blend derived from an individual Mitra®/VAMS device; A) DEM-IDMS analysis of cCRM vial in triplicate. B) Linear calibration IDMS analysis of one cRM vial analysed in triplicate. Error bars = Expanded measurement uncertainty at the 95% CI. Green shaded area = expanded measurement uncertainty (at 95% CI) of cCRM reference measurement.

DEM-IDMS procedure would normally be reserved for high accuracy low measurement uncertainty reference value assignment purposes. Typically, exact matching would be achieved if the calibration blend mass fraction matches the sample blend mass fraction within ±5 %. In some cases, multiple iterations may be required before this would be achieved. DEM-IDMS would use tight bracketing, alternating calibration blend and sample blends five times, whereby a calibration blend would be injected immediately before and after a sample blend significantly reducing instrumental drift effects. Consequently, this tight bracketing would have the potential to deliver high accuracy, high precision and low measurement uncertainty data. Due to the substantial increase in time and cost associated with the DEM-IDMS procedure it is not practicable for use in routine high throughput analysis (23,24). Therefore, for this work the two calibration models were assessed (linear calibration and DEM-IDMS) and the measurement uncertainty obtained by both was compared.

As expected, DEM-IDMS results obtained for the analysis of the cCRM vials were more accurate with 0.3-3 % difference from reference value. Whereas the linear calibration measurements were on average 6.6 % different from the reference value. It was thought that this degree of accuracy would be fit for purpose for routine TDM of tacrolimus as such small differences would not have an impact on the clinical decision of the TDM.

5.3.4 Analysis of patient blood pools samples

There was no defined accepted measurement uncertainty for the analysis of tacrolimus in patient blood in routine hospital laboratories from a clinical point of view. Hence, the threshold in difference in measurement between a method using microsamples and another using larger volumes of blood was not set. The higher-order CRM of tacrolimus spiked in human blood ERM DA 110 had an expanded measurement uncertainty of 3.3 % whereas the reference mass fraction of the cCRM of tacrolimus in pooled patient blood (described in chapter 4) was assigned with an initial expanded measurement uncertainty estimate of 2.9 %. As a rule of thumb, if a higher-order reference measurement procedure has an expanded measurement uncertainty of ~3 %, it would not be expected generally that application in a routine hospital laboratory setting would measure with a smaller uncertainty than 3 %. As a result, the acceptable threshold of 5 % difference in the measurements achieved by the two techniques; using 300 µL and 20 µL of sample blood was selected. The analysis of patient pooled blood samples (P-02, P-04, P-06 and P-09) gave data with average percentage difference ≤4.7 % from the reference values with low measurement uncertainty estimates (see Table 5.2).

Table 5.2 Comparison of Mitra®/VAMS devices (20 µL sample blood) linear calibration IDMS and LGC (300 µL sample blood) DEM-IDMS analysis of pooled patient blood materials

Panel	VAMS Linear Calibration Analysis (ng/g) [a]	LGC DEM-IDMS Analysis (ng/g) [b] (21)	Difference % [d]
P-02	04.34 ± 0.44	04.32 ± 0.11	0.5
P-04	07.64 ± 0.54	07.30 ± 0.18	4.7
P-06	11.53 ± 0.59	11.35 ± 0.29	1.6
P-09	19.45 ± 2.23	19.02 ± 0. 37	2.7
LGC cRM	06.65 ± 0.53	06.43 ± 0.22 [c]	3.4

[a] Values in table are expressed as the average of triplicate analysis ± expanded uncertainty at 95 % CI (k=2)
[b] These values have been reported previously through the MassTrak global inter-laboratory comparison study (21). Expanded uncertainty is expressed at the 95 % CI (k=2)

[c] Indicative reference value of the candidate tacrolimus in patient blood cCRM under production at LGC (see Table 4.6).
[d] The difference expressed as a percentage between the values of the VAMS linear calibration analysis and the reference LGC DEM-IDMS method.

The unweighted calibration line prepared using Mitra®/VAMS devices to perform the analysis, produced a straight line with a Pearson's correlation coefficient (r) value of 0.9999 and a coefficient of determination (R^2) of 0.9998. The sample measurements obtained using Mitra®/VAMS devices are shown in Figure 5.5. The %RSD for triplicate measurements (individual Mitra®/VAMS devices) ranged from 2-6 %. The differences between the average measured sample mass fractions were 0.5-4.7 % from the reference values, demonstrating the Mitra®/VAMS devices accuracy and precision over a range of mass fractions suggesting the device would be suitable for diagnostic purposes.

The criteria to assign the estimated limit of detection (LOD) and the estimated lower limit of quantification (LLOQ) was based on a signal-to-noise ratio peak to peak ($S:N_{p/p}$). The $S:N_{p/p}$ of 3:1 and 10:1 were accepted for the LOD and LLOQ, respectively (25–27). The method had LOD and LLOQ of 0.1 ng/g and 0.4 ng/g, respectively. The lowest calibration standard which was at 4 ng/g had a $S:N_{p/p}$ of 100:1, this is shown in Figure 5.6. This is well within the scope of the clinical testing for tacrolimus in whole blood as the mass fraction range for detection is the therapeutic window of 5-20 ng/g (28). As expected, the measurement uncertainty reported by the linear regression calibration was greater than the measurement uncertainty reported by DEM-IDMS measurements (see Table 5.2) because the tight bracketing with calibration blends in DEM-IDMS reduces the variability introduced by instrumental drift.

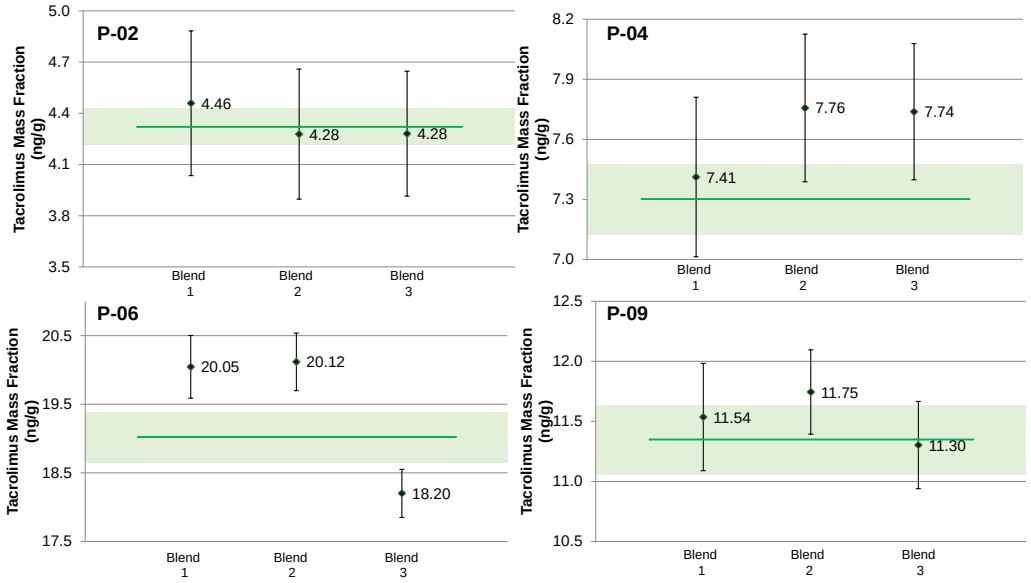


Figure 5.5 Samples analysis: Each data point is the average of 5 LC-MS replicate analysis of a single blend derived from an individual Mitra®/VAMS device. Each pooled patient blood sample was analysed in triplicate; three blends using individual Mitra®/VAMS devices. Reference values previously assigned to each of the samples shown as a solid green line and the measurement uncertainty for each reference value is indicated by the light green shaded region of the graph (21). Error bars = the expanded measurement uncertainty at the 95 % CI.

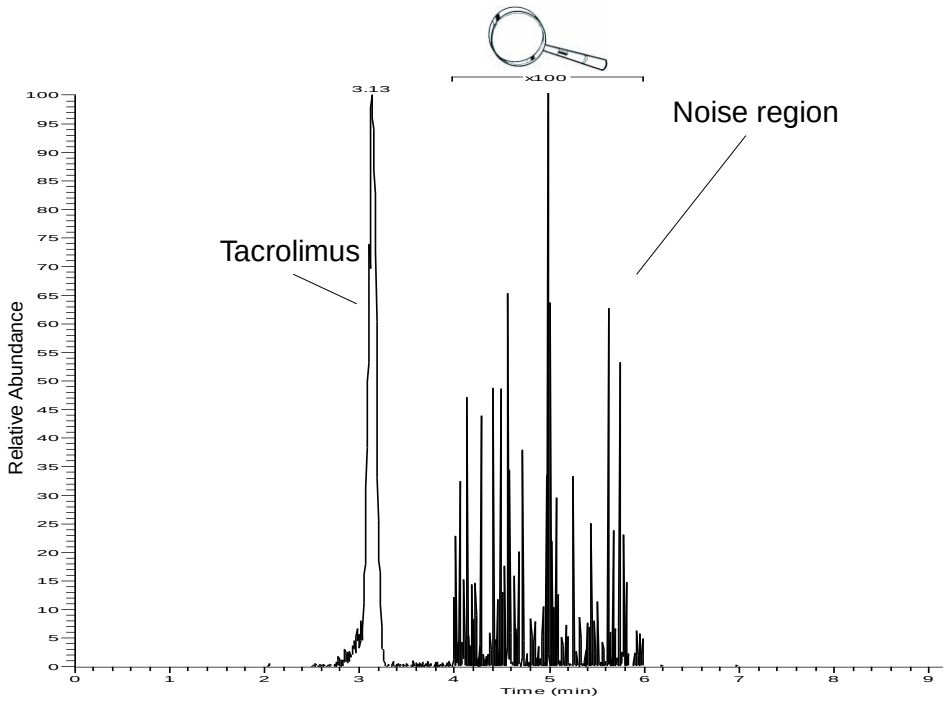


Figure 5.6 Chromatographic peak of tacrolimus matrix matched calibration blend at 4 ng/g with noise region 4-6 min magnified 100 times indicating a signal-to-noise peak to peak of $S:N_{p/p}$ = 100:1. The magnifying class indicating the noise region selected and zoomed in x100 times.

5.3.5 Sample amount

Because the preparation of blends was done gravimetrically, the exact mass of blood wicked up by Mitra®/VAMS devices was recorded. The masses were then converted to volumes using the measured density of each of the blood samples. Figure 5.7 shows the volume of blood wicked by 25 Mitra®/VAMS devices. The average volume of blood wicked up was 22.06 µL with a %RSD of 3.3 %. This demonstrated not only the high reproducibility of the volume of blood wicked but also good accuracy as the average volume of blood was within 3 % agreement with the manufacturers stated volume of 21.4 µL (manufacturer's reported %RSD is 3.44 %).

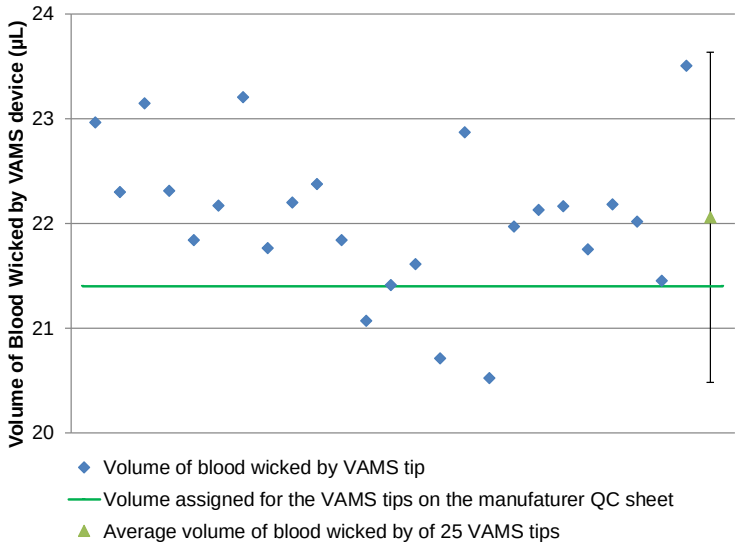


Figure 5.7 Volume of blood wicked up by the Mitra®/VAMS devices during this study. Tested batch of Mitra®/VAMS devices QC certificate stated a calculated average blood wicking volume of 21.4 µL of blood on average, shown in the chart as a solid green line. The data point to the right is the average volume of blood wicked up. Error bars on this value are equivalent of 2 x standard deviation, n = 25.

When using the adapted method from the higher-order RMP, gravimetric preparation was performed. This would not be the case for the hospital laboratory where they would use the manufacturer's reported volume for each batch of tips. The variability of 3.3 % as reported in this work or 3.4 % as reported by the manufacturer on the volume of blood taken would then need to be incorporated in any measurement uncertainty estimates calculated if the hospital laboratory were to report the measurement uncertainty. This variability did not impact on this study as the exact masses of blood taken were used for the calculations of the mass fractions. In this higher-order RMP, gravimetric preparation minimised any variability in the measurement of the sample amount taken and provided a small measurement uncertainty associated with the sample amount because of the use of high accuracy analytical balance.

In this work the Mettler XP6 was used to weigh the Mitra®/VAMS tips and all the additions made in the blend preparation including the vial, the blood, the isotopically-labelled IS as indicated in 5.2.4.1. The measurement uncertainty of the balance was ± 3 µg which was used to assign the measurement uncertainty component of the mass of sample blood taken. As an example, the average mass of blood taken which was 23.09 mg, would then have the percentage uncertainty 0.1 % of the mass of blood taken. The gravimetric preparation minimised the contribution to the measurement uncertainty. On the other hand, if a hospital laboratory would report the measurement uncertainty to their result, they would need to use the 3.4 % %RSD as the uncertainty component for the sample amount taken. This would increase the percentage measurement uncertainty component of the amount of sample from 0.1 % to 3.4 % which would be about 25 times increase in this component alone.

Performing gravimetric preparation was not only laborious and time consuming, but additionally required expensive high accuracy balances, resources like these may not necessarily be readily available in a hospital laboratory. This demonstrated the importance of the use of higher-order RMPs in a national measurement laboratory setting where such laborious work would support hospital laboratories; to understand the sources of variability in their measurement and incorporate these into their measurement uncertainty budget.

5.3.6 General discussion

The manufacturer of the Mitra® device recommends a drying time of blood on samplers to be approximately three hours. Previous studies using Mitra®/VAMS devices to measure a range of compounds used drying times from 1 h to overnight (7–9,13). Therefore, due to the limited resources no evaluation of the optimum drying times was performed and the safer option of a longer drying time, all Mitra®/VAMS devices used in this study. All devices were dried overnight, for about 18 hours. An overnight drying time also resembles the final real-life scenario if a patient were to provide a sample at home then dry it and send it by post, it would not reach the hospital laboratory on the same day. A standardised drying time of 18 hours would be suitable for the procedure to remove any potential variability that could be caused by varying drying times.

Biases associated with the use of Mitra®/VAMS device have been reported (7,8,13). These biases were not associated with the drying times but rather with the variability in the volume of blood taken, the biases were < 15 % and were considered acceptable (7,8). In this work the highest percentage difference from the reference values was 4.7 % reflecting the accuracy of the measurements achieved using the adapted of higher-order RMP to evaluate the devices. This was another example of the difference between the use of higher-order RMPs and routine methods. The determining factor of evaluating a method and its performance would be the purpose of the method. For example, 15 % bias reported in some of the applications by Denniff and Spooner (7) was considered acceptable as it was what would be fit for the purpose of their application. Similarly, with higher-order reference

measurement work the biases or the % differences obtained were considered acceptable. This was because the % differences of the reference values to those obtained by Mitra®/VAMS ranged from 0.5-4.7 % only, (see Table 5.2), while the starting sample amount used for the analysis was 15 times smaller (20 µL vs. 300 µL of blood).

The analyses performed in this study used venous whole blood samples in contrast to the intended patient use scenario which would be capillary blood directly sampled from the fingertip for example. A study (29) comparing measured tacrolimus concentrations from both capillary and venous samples taken from transplant patients showed that tacrolimus levels in a capillary were lower than those measured in venous whole blood. However, with a small mean difference of 1 ng/mL across the therapeutic index concentration range measured (2-16 ng/g) which was unlikely to have clinical significance Additionally, the median difference for concentrations below 10 ng/mL was 0.5 ng/mL. Dickerson *et al.* reported in their study that for routine clinical purposes tacrolimus concentration levels measured in capillary and venous blood can be used interchangeably (29).

Mitra®/VAMS devices were reported to overcome the two main issues that DBS suffer from, these include; a) the variability introduced by sub-punching the paper of the DBS and b) the haematocrit effect and the impact of the difference of the blood used for calibrants and samples. The haematocrit issue was found to be critical when sampling blood microsamples from neonatal and paediatric patients (7,9,12,13). Similarly, Lawson *et al.* recently reported that differences in DBS factors such as blood spot size and punch location in addition to haematocrit differences markedly impacted the measured

concentration in the UK screening programme (30). The use of micro-sampling for the TDM of paediatric transplantation would provide the clinician with a better approach for estimating the starting dose and ongoing dose control especially in the immediate period following surgery. Moreover, among the paediatric population, Mitra®/VAMS could provide a better patient experience and compliance especially when sampling at home could be performed by parents.

A proof-of-concept manuscript was submitted for publication after the completion of this work however another paper was published just at the time of submission that evaluated Mitra®/VAMS for the analysis of tacrolimus. This paper by Kita and Mano from the University of Tsukuba in Japan, developed an analytical method for the analysis of tacrolimus using 10 µL Mitra®/VAMS devices (31). While the study described in this chapter was investigating the applicability of the use of Mitra®/VAMS devices for the analysis of tacrolimus in whole blood from an analytical aspect, the Kita and Mano study aimed at evaluating the impact of haematocrit on the measurement as the main aim of their study. Kita and Mano demonstrated that the measurement of tacrolimus was independent of haematocrit which was the core added value that the Mitra®/VAMS devices could provide over DBS.

The Kita and Mano study also provided a thorough level of method validation which was not performed in this study. The aim was to provide a proof of principle that the Mitra®/VAMS devices could analytically provide the accuracy required. Furthermore, such level of validation was not crucial because this work was performed using higher-order CRM and cCRM in addition to patient

pooled blood samples that were assigned traceable higher-order reference measurement.

Furthermore, the blood samples analysed by Kita and Mano were not patient samples but ‘naïve’ blood that came from healthy individuals that was fortified with tacrolimus standards. The authors investigated the binding of the standard with the red blood cells to mimic patient samples where they used blank blood and fortified it with tacrolimus standard. With their approach, the issue of patient blood interferences such as metabolites remained unassessed. In this work, however, real patient blood was used (cCRM and the four pooled blood samples) where the patients had been administered tacrolimus was used. Moreover, while the method in this thesis used stable isotopically-labelled tacrolimus as an internal standard the Kita and Mano study used ascomycin, an analogue compound to tacrolimus. The authors assessed the accuracy of the method using ascomycin by inter-batch assay accuracy and precision and accepted range of ± 20 % which was a much larger tolerance range than the reported percentage differences found in this study which were 0.5-4.7 %, (see Table 5.2). The higher accuracy in the measurements in this study were due to the use of the adopted higher-order RMP which included the use of IDMS and gravimetric preparation that improved the performance of the method.

An important issue that Kita and Mano investigated briefly which would require further investigation and was not addressed in this chapter was the stability of tacrolimus on the Mitra®/VAMS devices in storage. The data showed degradation of tacrolimus on the tip of the device at room temperature however tacrolimus was stable on the devices when stored in the freezer. This

fundamentally questions the applicability of using the Mitra®/VAMS devices for home-sampling if tacrolimus was not found to be stable on the Mitra®/VAMS devices at room temperature, it could be that the shipping of the samples might be complicated with shipping with ice packs which would have cost implications. However, the devices' applicability to clinical studies and pharmacokinetics studies remains strong. Kita and Mano's work was performed using 'naïve' blood rather than patient samples hence, further work was required to assess the stability of tacrolimus in patient blood samples on Mitra®/VAMS devices which was not performed as part of this study due to limited resources.

As for method performance, LOD and LLOQ reported in the method in this chapter were lower than those reported by the Kita and Mano study which reported and LLOQ of 1 ng/mL (31) whereas the LLOQ for this method is 0.4 ng/g. For the ease of comparability, the LLOQ reported by Kita and Mano was converted to ng/g by multiplying by the average density of blood of 1.055 g/mL (32) the LLOQ would be 1.055 ng/g which would be 2.6 times higher than the LLOQ reported in this method. This lower LLOQ was achieved because of the use of the adapted higher-order RMP which uses 300 g of blood as a sample size instead of 20 g. This demonstrated an example of the differences in the measurements achieved by a higher-order reference methods and reference materials vs routine methods.

Finally, if the manufacturer is interested in assigning measurement uncertainty to the Mitra®/VAMS devices, extensive work would be required. The uncertainty of these devices would need to reflect not only the variability in the blood amount taken additionally it should encompass all other aspect of the

analysis including the variability of day-to-day, user-to-user, extraction method, compound specifics. The work described here provides a promising starting point to build upon in terms of assigning a relatively low uncertainty on the devices.

5.4 Conclusions

In conclusion, tacrolimus in pooled patient blood cCRM plus previously reference value assigned inter-laboratory comparison study patient whole blood samples were used to demonstrate the potential for Mitra®/VAMS devices to be used for TDM of the immunosuppressant drug tacrolimus. Generation of a blood/water mix before using extraction solvents was critical to removing the significant negative bias first observed during evaluation. Between Mitra®/VAMS devices reproducibility was excellent for the batch assessed (both blood volume wicked up and measured amount of tacrolimus). Additionally, better understanding of the sources of variability in the measurement was obtained by the comparison the measurement uncertainty of volumetric and gravimetric preparation of samples and the use of SILIS.

This work added to the body of literature on the use of Mitra®/VAMS devices for the measurement of tacrolimus. The work described here compared to the Kita and Mano work (31) has provided the first method to use isotopically-labelled tacrolimus internal standard method for the Mitra®/VAMS devices and it is the first to assess the devices using real patient samples.

The next step required in the evaluation would be a large-scale study comparing data from venous blood collected in EDTA tubes to finger prick collected blood by Mitra®/VAMS devices.

Ultimately, improved TDM of immunosuppressants through standardisation and implementation of home-sampling would improve patient welfare from both a patient convenience and an improved TDM perspective. Improved TDM could reduce organ rejection episodes due to avoidance of subtherapeutic levels of ISDs and prevent adverse effects due to supratherapeutic levels of ISD including nephrotoxicity and virus induced malignancy.

Therefore, the hypothesis of this chapter holds true in part. The VAMS provided traceable measurements with measurement uncertainties that were fit for routine analysis. However, the blood that was used for this study was venous blood, a capillary blood evaluation is still required and is part of the future work.

5.5 Novelty and value

At the time of performing this study, there was no evaluation of the use of Mitra®/VAMS devices for the analysis of tacrolimus in whole blood. This study demonstrated that devices could be a suitable technique to use for TDM of tacrolimus, confirming the findings in the literature. This was to potentially provide a home sampling technique that would save the immunosuppressed patients hospital visits for regular blood tests.

In this study the use of adapted higher-order RMP for the evaluation of the novel sampling devices allowed better understanding of the sources of variability and uncertainty associated with the volumetric preparation of samples rather than the higher-order gravimetric preparation.

5.6 Future work

Building on the highly accurate measurements provided in this chapter with the recent findings in the literature, a few main areas remain to be investigated. Most importantly, would be the stability of tacrolimus in patient blood on the Mitra®/VAMS devices at different temperatures. Stability at room temperature would be key to assess the applicability of the Mitra®/VAMS devices for home sampling for organ transplant patients. Moreover, the performance and validation of the adapted higher-order RMP would help in comparison with the previously published routine methods. Another area that could add value to the research performed on the use of the Mitra®/VAMS devices is to investigate would be pre-loading the SILIS on the tips prior to sampling. This could provide an opportunity to improve the measurements as the internal standard would be undergoing the same conditions the compound would undergo from the moment of sampling hence compensating for any losses in the process.

5.7 References

1. Oliveira RV, Henion J, Wickremsinhe ER. Automated direct extraction and analysis of dried blood spots employing on-line SPE high-resolution accurate mass bioanalysis. Bioanalysis. 2014 Aug;6(15):2027–41.
2. Dillen L, Loomans T, Van de Perre G, Versweyveld D, Wuyts K, de Zwart L. Blood micro-sampling using capillaries for drug-exposure determination in early preclinical studies: a beneficial strategy to reduce blood sample volumes. Bioanalysis. 2014 Feb;6(3):293–306.
3. Nilsson LB, Ahnoff M, Jonsson O. Capillary micro-sampling in the regulatory environment: validation and use of bioanalytical capillary micro-sampling methods. Bioanalysis. 2013 Mar;5(6):731–8.
4. Zhang J, Lok D, Gray J, Grossman S, Jones M. Development of a novel noncapillary plasma micro-sampling device for ultra-low volume of blood collection. Bioanalysis. 2016 May;8(9):871–80.
5. Cheung CY, Van Der Heijden J, Hoogtanders K, Christiaans M, Liu YL, Chan YH, et al. Dried blood spot measurement: application in tacrolimus monitoring using limited sampling strategy and abbreviated AUC estimation. Transpl Int. 2008 Feb 1;21(2):140–5.

6. Rincón JP, Meesters RJW. Evaluation of peripheral blood micro-sampling techniques in combination with liquid chromatography-high resolution mass spectrometry for the determination of drug pharmacokinetics in clinical studies. Drug Test Anal. 2014 Jun;6(6):568–77.
7. Denniff P, Spooner N. Volumetric Absorptive Micro-sampling: A Dried Sample Collection Technique for Quantitative Bioanalysis. Anal Chem. 2014 Aug 19;86(16):8489–95.
8. Denniff P, Parry S, Dopson W, Spooner N. Quantitative bioanalysis of paracetamol in rats using volumetric absorptive micro-sampling (VAMS). J Pharm Biomed Anal. 2015 Apr 10;108:61–9.
9. De Kesel PMM, Lambert WE, Stove CP. Does volumetric absorptive micro-sampling eliminate the hematocrit bias for caffeine and paraxanthine in dried blood samples? A comparative study. Anal Chim Acta. 2015 Jun 30;881:65–73.
10. NCI Dictionary of Cancer Terms [Internet]. National Cancer Institute. 2011 [cited 2019 Dec 4]. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms
11. Spooner N, Denniff P, Michielsen L, De Vries R, Ji QC, Arnold ME, et al. A device for dried blood micro-sampling in quantitative bioanalysis: overcoming the issues associated blood hematocrit. Bioanalysis. 2014 Dec 16;7(6):653–9.
12. Kip AE, Kiers KC, Rosing H, Schellens JHM, Beijnen JH, Dorlo TPC. Volumetric absorptive micro-sampling (VAMS) as an alternative to conventional dried blood spots in the quantification of miltefosine in dried blood samples. J Pharm Biomed Anal. 2017 Feb 20;135:160–6.
13. Mano Y, Kita K, Kusano K. Hematocrit-independent recovery is a key for bioanalysis using volumetric absorptive micro-sampling devices, Mitra™. Bioanalysis. 2015 Aug 1;7(15):1821–9.
14. Cordova C. Permission to use Mitra image. 2016.
15. Immunosuppressive therapy for kidney transplant in adults | Guidance and guidelines | NICE [Internet]. [cited 2019 Jan 30]. Available from: https://www.nice.org.uk/guidance/ta481
16. Baker R, Mark P, Patel R, Stevens K. Clinical Practice Guideline Post-Operative Care in the Kidney Transplant Recipient [Internet]. [cited 2019 Jan 30]. Available from: https://renal.org/wp-content/uploads/2017/06/final-post-operative-care-guideline.pdf
17. Wallemacq P, Armstrong VW, Brunet M, Haufroid V, Holt DW, Johnston A, et al. Opportunities to optimize tacrolimus therapy in solid organ transplantation: report of the European consensus conference. Ther Drug Monit. 2009 Apr;31(2):139–52.
18. Hoogtanders K, van der Heijden J, Christiaans M, van de Plas A, van Hooff J, Stolk L. Dried blood spot measurement of tacrolimus is promising for patient monitoring. Transplantation. 2007 Jan 27;83(2):237–8.
19. Koop DR, Bleyle LA, Munar M, Cherala G, Al-Uzri A. Analysis of tacrolimus and creatinine from a single dried blood spot using liquid chromatography tandem mass spectrometry. J Chromatogr B. 2013 May 1;926:54–61.
20. NICE public health guidance 43: Hepatitis B and C: ways to promote and offer testing to people at increased risk of infection (2012) [Internet]. [cited 2016 Nov 15]. Available from: https://www.nice.org.uk/guidance/ta85/chapter/tacrolimus, %20[Accessed %2022/08/2016].#tacrolimus
21. Annesley TM, McKeown DA, Holt DW, Mussell C, Champarnaud E, Harter L, et al. Standardization of LC-MS for Therapeutic Drug Monitoring of Tacrolimus. Clin Chem. 2013 Jul 31;clinchem.2013.209114.
22. Levine DM, Maine GT, Armbruster DA, Mussell C, Buchholz C, O'Connor G, et al. The Need for Standardization of Tacrolimus Assays. Clin Chem. 2011 Jan 12;57(12):1739–47.

23. Barbagallo RP, Boley N, Holcombe G, Merson S, Mussell C, Pritchard C, et al. Production and certification of four frozen human serum certified reference materials containing creatinine and electrolytes. Ann Clin Biochem. 2008 Jan 3;45(2):160–6.
24. Stokes P, O'Connor G. Development of a liquid chromatography–mass spectrometry method for the high-accuracy determination of creatinine in serum. J Chromatogr B. 2003 Aug 25;794(1):125–36.
25. Evard H, Kruve A, Leito I. Tutorial on estimating the limit of detection using LC-MS analysis, part I: Theoretical review. Anal Chim Acta. 2016 Oct 26;942:23–39.
26. Evard H, Kruve A, Leito I. Tutorial on estimating the limit of detection using LC-MS analysis, part II: Practical aspects. Anal Chim Acta. 2016 Oct 26;942:40–9.
27. Duggan JX. Quantification below the LLOQ in regulated LC–MS/MS assays: a review of bioanalytical considerations and cautions. Bioanalysis. 2019 Apr 1;11(8):797–814.
28. MULTICENTER RANDOMIZED TRIAL COMPARING TACROLIMUS (FK506) AN... : Transplantation [Internet]. [cited 2019 Mar 28]. Available from: https://journals.lww.com/transplantjournal/Fulltext/1997/08150/MULTICENTER_RANDOMIZED_TRIAL_COMPARING_TACROLIMUS.12.aspx
29. Dickerson JA, Sinkey M, Jacot K, Stack J, Sadilkova K, Law YM, et al. Tacrolimus and sirolimus in capillary dried blood spots allows for remote monitoring. Pediatr Transplant. 2015 Feb 1;19(1):101–6.
30. Lawson A, Bernstone L, Hall S. Newborn screening blood spot analysis in the UK: influence of spot size, punch location and haematocrit. J Med Screen. 2016 Mar 1;23(1):7–16.
31. Kita K, Mano Y. Application of volumetric absorptive micro-sampling device for quantification of tacrolimus in human blood as a model drug of high blood cell partition. J Pharm Biomed Anal. 2017 Sep 5;143(Supplement C):168–75.
32. Kenner T. The measurement of blood density and its meaning. Basic Res Cardiol. 1989 Apr;84(2):111–24.

6 Conclusions

The aim of this PhD research was to investigate and better understand the sources of variability in the process of higher-order reference measurement of small molecules in biological samples for clinical diagnosis and patient monitoring. This was successfully achieved by developing and applying a candidate higher-order reference measurement procedure (RMP).

For the first time, this work presents the method development of a candidate higher-order reference measurement procedure for the low molecular weight, polar metanephrines in plasma (Chapters 2 & 3). Additionally, the candidate higher-order reference material (CRM) of tacrolimus in pooled patient blood that was characterised (Chapter 4) is the first material of its kind. This material, produced from pooled patient blood, is the first to have been assigned a higher-order reference value. Tacrolimus in the cCRM was incurred in the blood through patients' administration rather than fortification of healthy individual blood pools with reference materials. That meant the material did not only have tacrolimus but also its metabolites because it was produced from patients blood that have metabolised the drug in their livers. This added value of the new material would enable better evaluation of methods in hospital laboratories, especially immunoassay-based methods that reported cross reactivity with metabolites. This work is the first to use standardisation tools of higher-order RMPs and CRMs to evaluate a novel micro-sampling device (Chapter 5).

All of the above accomplished objectives provided a better understanding of the sources of variability in the measurement of such small molecules in

human biological matrices and practical ways to mitigate these were subsequently found. Importantly, the relative contribution of each mitigation approach was characterised extensively. As a result, there are several conclusions that are outlined below, all loosely framed around the DEM-IDMS measurement uncertainty budget equation, Figure 6.1.

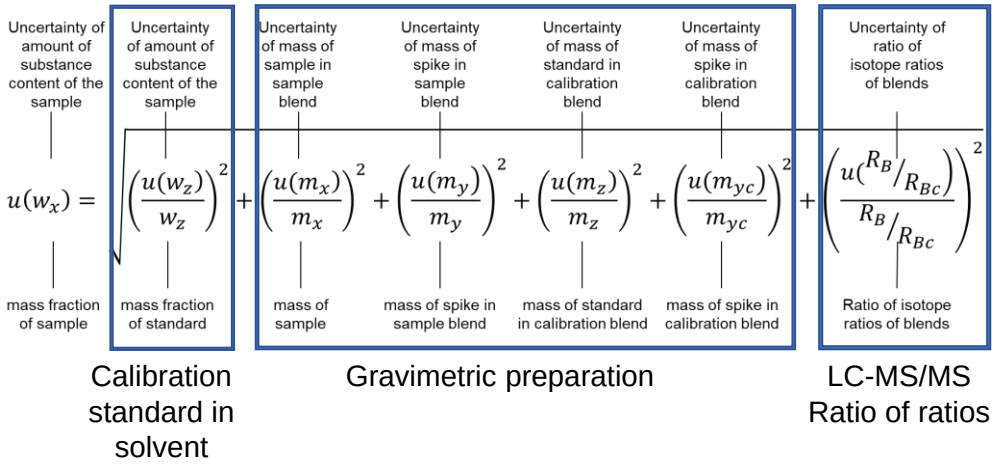


Figure 6.1 The equation for the total relative measurement uncertainty of the amount of substance in a sample measured by DEM-IDMS (1–3).

A. The preparation and, more importantly, **control** of standards prepared in pure solvents required more careful consideration than expected. Inaccurate preparation of the standards in solvent fundamentally impacted the accuracy and traceability of the measurement. The inaccuracy arising from the use of such standards came from different sources including: a) the preparation of standards themselves, for example human variability when standards are prepared by different analysts (Figure 3.3 and Figure 4.5); and b) when preparing low-level concentration standards, impurities in the solvent become very significant due to the degradation of the standard. Unexpectedly, the purity of the solvent was a source of variability that lay largely outside the control of the analyst where solvents sold as

‘pure’ contained sufficient impurities to affect the results, these impurities being assigned by qNMR (Figure 4.9). Such sources of variability mandated the implementation of measures to control and ensure the integrity of the standards throughout the measurement process.

B. Achieving low measurement uncertainties is MS instrument dependent. Upon testing eight different MS instruments at four different laboratories using the same standards and LC methods the performance highly varied among the different instruments (Figure 2.10). Improving the instrument capability to a practical performance level was achieved ultimately using a micro-flow LC method which further capitalised on the MS performance itself. This was achieved by the reduction of the LC flow rate to 30 µL/min of 98 % aqueous mobile phase, this helped improve the spray formation and the ionisation efficiency.

C. Optimisation of the MS dwell time was critical to reduce the instrumental measurement variability by half and regardless of analyte properties or the MS instrument. Across all seven different MS instruments for both metanephrines and tacrolimus, the one MS condition that had the largest impact on reducing the measurement variability was the MS dwell time (Figure 2.14). A good example of this was shown when switching instruments during the analysis of tacrolimus. Low measurement uncertainty was only reproducible across different instrument once the dwell time was optimised (Tabl 4.6) and the chromatography method was easily transferrable.

D. The second important factor to reducing instrument variability across all MS instruments tested was the equilibration time of the MS instrument. All MS

instruments tested required about ten hours of operation prior to achieving reproducible low measurement uncertainty.

E. Lower expanded measurement uncertainty at higher mass fractions (e.g., ≤3.5 % at 4-20 ng/g) was achieved independent of the LC-MS instrument used (Table 4.6). The uncertainties increased at ultra-low mass fractions sub 200 pg/g (e.g., ≤10.5 % and ≤15.3 % for 70-100 pg/g and <25 pg/g, respectively) and were instrument dependent (Table 3.4).

F. Current technology can achieve an expanded measurement uncertainty of ±20 % when measuring small molecules in small amounts of plasma at ultra-low mass fractions. For example, in this work, 3MT (167.2 g/mol) was measured in 0.5 g of plasma at 1-3 pg/g ±20 % (Figure 2.18). Although this uncertainty may not be sufficiently small for a higher-order reference measurement, it demonstrates that such measurement is now achievable with current technologies, which was not possible with older instruments. Future generations of MS instruments are likely to lower measurement uncertainties at such ultra-low levels as technology advances.

G. Establishing measurement uncertainty that is fit for purpose in the clinical context at least for MET is an iterative process between analytical scientists and clinicians. This is because, the uncertainty of the measurement must be less than the biological variability being assessed. The analytical scientist may be able to provide a very small measurement uncertainty but at a cost, usually time to either improve a method's performance or acquire more data. Similarly, the number of patients that need to be assessed before any review is warranted needs to be decided based on properly defined clinical end points. Hence, assigning clinical decision cut-off points and ranges is best as a collaborative process between clinicians and

analytical scientists. For example, if a clinical decision point is set at 8 ng/g for a compound to be healthy and any measurement above or below would require treatment, then if the analytical method has an uncertainty of 10 % then any value from 7.2-8.8 should be expected to be healthy.

H. The sample preparation method was found not to be as critical as the LC-MS instrumentation used for the analysis to achieve the measurement uncertainty required for higher-order reference measurements. For both metanephrines and tacrolimus the same sample preparation method was used to analyse samples using different LC-MS instruments. The importance of the instrument was more pronounced with the metanephrines due to the much lower mass fractions. Further optimisation of the sample preparation method did not improve the performance as much as the MS instrument did (Figure 2.10 shows a comparison of four different triple quadrupole mass spectrometers).

I. Alternative sampling techniques (e.g., VAMS, assessed gravimetrically) could offer practical solutions for new measurement capabilities with acceptable measurement uncertainties (≤12 %) that could be used for clinical hospital laboratories (Table 5.2 and Figure 5.5). These would not only provide a better patient healthcare experience but also suffer less variability than dried blood spots collected on paper cards, where the variability is inherent from the haematocrit differences and the types of paper card used and the sample by the punch of the blood spot on the card.

J. For accurate and precise measurements of small molecules at the extreme upper end of the small molecule weight range in the presence of erythrocytes in the sample, thorough sample mixing twice of thirty seconds each time (with a short pause in between each 30 seconds mixing step)

prior to taking an aliquot for analysis was found to be crucial. This was not the case for polar small molecules where a 10 second single mixing step was sufficient. This was found to be important regardless of sample size where for example, for tacrolimus when using both 300 µg and 20 µg starting sample, extensive mixing was required (see Figure 5.2).

6.1 Contribution to the field

This research is an interdisciplinary subject of clinical chemistry and laboratory medicine (CCLM), metrology and analytical science. In each of these areas the findings of this research have achieved a new milestone.

This is the first time the study of these compounds has been made to this level of accuracy and precision, enabled by a new, deeper understanding of the sources of variability that have either been previously understudied or overlooked. The application of metrological concepts to chemistry and clinical chemistry is not new; however, the analysis of small organic molecules at the extreme ends of the molecular weight range and at such low levels is a first. In fact, the Joint Committee for Traceability in Clinical Chemistry and Laboratory Medicine (JCTLM) database that has RMPs listed through a thorough peer-review process of NMIs and clinical reference laboratories, includes a majority of RMPs that measure at mass fractions ranging from ng/g to mg/g. Out of the hundred and six JCTLM-database RMPs (analysing forty-one small clinical organic molecules), only three methods analyse molecules at mass fractions below 100 pg/g (4). Furthermore, these three analytes are all larger in molecular weight than metanephrines.

Additionally, the work conducted to assess the alternative sampling devices (Mitra®/VAMS assessed gravimetrically) is the first to demonstrate the capability of reference measurement tools of RMPs and CRMs to assess new technologies. Lastly, the tacrolimus material that was characterised in this work is a novel material that is produced from pooled patient blood instead of healthy individual blood spiked with the compound. This is by far a better tool that would enable hospital laboratories evaluate and improve their methods as it was produced by patient blood samples and contains the metabolites profile of tacrolimus and characterised using a higher-order RMP.

Finally, this work paves the way for developing higher-order RMPs of molecules with similar properties to those tested. The work provides the tools to standardise these clinical applications to improve the diagnosis and monitoring of patients and ultimately improve patient health care in the UK and globally.

6.2 Recommendations for future work

This research provides the grounds to further investigate and improve the abilities and impact of clinical chemistry metrology in several areas. Several recommendations for future research are suggested below.

6.2.1 Other compounds

This thesis focussed on two case examples of low molecular weight compounds at either end of the range, and of analyte polarity. Knowledge that was acquired through this research should now be transferred and applied to other compounds with similar properties to the METs and tacrolimus to check the generalisability of some of the observations made herein. For example,

these could include: a) catecholamines (adrenaline, noradrenaline and dopamine) that are small and polar molecules; and b) other cyclic immunosuppressant drugs with larger molecular weights of 800-1000 Da, e.g., sirolimus and everolimus. The conclusions of this work pave the way to developing and applying higher-order reference measurements for these compounds. This will help assess if the same principles apply to other compounds of similar chemical and biological properties to the compounds used to conduct this research.

6.2.2 Measurements at ultra-low mass fractions

This work builds the basis for future work for higher-order reference measurements of small molecules (< 200 Da) in biological samples at ultra-low mass fractions. Potential avenues for this work to continue is to assess the impact of increasing starting sample mass on potentially increasing the sensitivity and therefore reducing the uncertainty.

6.2.3 Calibrations standards prepared in ‘pure’ solvents

The findings of this work provided a better understanding and impact of the use of ‘pure’ solvents for the preparation of standards and brings to light a generally overlooked issue. Further work on this is urgently required to assess the impact of different solvents and their purity on the stability of analytes. The degradation of standards due to solvent impurities raises the question of the use of mixed standards and how their combination could affect their relative stability. Furthermore, given the issue was detected at lower mass fraction standards, it is important to elucidate whether this effect is compound dependent, or mass fraction dependent, or both and to what degree. More experiments investigating standards in solvent stability would fundamentally

improve the measurement which would ensure not only accuracy, but also traceability of measurements.

6.2.4 Clinical community engagement

The work conducted in this research was conducted at the National Measurement Laboratories (LGC, Teddington, UK) where traceable measurements with low measurement uncertainties lay in the core of the research conducted across many fields not only clinical chemistry, e.g., food chemistry, industrial paints, etc. The choice of the case study applications was made in consultation with the clinical community. An important area to progress this research is to communicate the findings or this research with hospital laboratories and reference standards manufacturers by publishing papers in peer-reviewed journal. Additionally, this work enables the collaboration with external quality assurance scheme (EQAS) providers to assign reference measurement to EQAS samples to potentially start a global inter-laboratory study comparison. This could be performed on both METs and tacrolimus, for the first time. It would also pave the way to starting standardisation initiatives for other compounds.

6.3 References

1. Bedson P. Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS) [Internet]. 2002 [cited 2018 Jun 25]. Available from: http://pubs.rsc.org/en/Content/eBook/978-0-85404-418-4 - pages 1-34.
2. Ellison S, Williams A. EURACHEM/CITAC Guide CG4 - Quantifying Uncertainty in Analytical Measurement [Internet]. Eurochem & CITAC - Co-Operation on International Traceability in Analytical Chemistry; 2012 [cited 2017 Nov 14]. Available from: https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf
3. Joint Committee for Guides in Metrology. JCGM 100:2008 - GUM 1995 with minor corrections. Evaluation of measurement data — Guide to the expression of uncertainty in measurement [Internet]. 1st ed. 2008. 134 p. Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
4. BIPM - About the BIPM [Internet]. [cited 2020 Oct 13]. Available from: https://www.bipm.org/en/about-us/

7 Appendices

7.1 Appendix 2 - Certificates of analysis of solid standards

7.1.1 Certificate of analysis (CoA) of Metanephrine standard from the supplier

Clearsynth
Inspiring Research

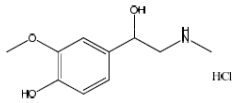
ASIA OFFICE:
Clearsynth Labs Ltd.
17, Lotus Business Park,
Off New Link Rd, Andheri (W)
Mumbai – 400053 India
Tel. +91.22.45049945

ASIA R&D CENTRE:
Clearsynth Research Centre
Plot No. 177, IDA,
Mallapur, Hyderabad - 500076
Telangana, India
T: +91.40.27155481

NORTH AMERICAN R&D CENTRE
Clearsynth Canada Inc.
2195 Speakman Drive,
Lab, Suite 1001, Mississauga,
Ontario L5K 1B4, Canada
Tel: +1.289.729.0060

CERTIFICATE OF ANALYSIS

Product Name	**rac Metanephrine Hydrochloride Salt**		
Chemical Name	4-(1-hydroxy-2-(methylamino)ethyl)-2-methoxyphenol hydrochloride		
Synonyms	Not Available		
CAS #	881-95-8		
Catalogue #	CS-T-57570	**Batch #**	CS-ME-AAA-0530-01
Molecular formula	$C_{10}H_{16}ClNO_3$	**Molecular weight**	233.69
Date of Analysis	23/05/2019	**Retest Date**	23/05/2024
Storage conditions	Keep in a tightly closed vial,store refrigerator		



Test	Specification	Result
Appearance	White to off white solid	White to off white solid
Solubility	Soluble in DMSO	In DMSO
1H NMR spectrum	Conform Structure	Confirms
Mass Spectrum	Conform molecular weight	Confirms
Purity by HPLC	Not less than 90%	98.38%

This material complies.

24/05/2019
Reviewed & Approved by
(QC Analytical)

24/05/2019
Reviewed & Approved by
(Quality Assurance)

The re-test data is assigned based on the literature available. The balances used are calibrated with weights traceable to the National standards NIST.

Quality Accreditations – ISO 9001:2015|ISO 14001:2015|OHSAS 18001:2007|ISO Guide 34:2009| ISO/IEC 17025:2005
info@clearsynth.com | www.clearsynth.com | Worldwide Helpline No:- +1.800.650.8239 | CIN :- U24100MH2010PLC201433

"The 'Catalyst' in making your research work"

Page 1 of 1

7.1.2 CoA of Normetanephrine standard from the supplier

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, Saint Louis, MO 63103, USA
Website: www.sigmaaldrich.com
Email USA: techserv@sial.com
Outside USA: eurtechserv@sial.com

Certificate of Analysis

Product Name:
DL-Normetanephrine hydrochloride - ≥98%

Product Number:	**N7127**
Batch Number:	**SHBD2227V**
Brand:	SIGMA
CAS Number:	1011-74-1
MDL Number:	MFCD00012882
Formula:	C9H13NO3 · HCl
Formula Weight:	219.67 g/mol
Quality Release Date:	09 MAY 2013
Date Retested:	08 JAN 2019
Recommended Retest Date:	JAN 2022

Test	Specification	Result
Appearance (Color)	White to Off White	Off-White
Appearance (Form)	Powder	Powder
Infrared Spectrum	Conforms to Structure	Conforms
Solubility (Turbidity)	Clear to Very Slightly Hazy	Very Slightly Hazy
Solubility (Color)	Colorless to Yellow	Faint Yellow
Solubility Concentration: 50 mg/mL in Water		
Purity (HPLC)	≥ 98 %	100 %
Recommended Retest Period 3 years	----------	----------

Michael Grady, Manager
Quality Control
Sheboygan Falls, WI US

Version Number: 3

Page 1 of 1

7.1.3 Certificate of the traceable purity analysis of Normetanephrine solid standard by the NML

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.docx

LGC

UKAS CALIBRATION 0423

^{1}H qNMR Assay of a Normetanephrine salt
Lot: SHBD2227V (OA/19/3114), using
SOP INS/B1-0416 (version 5)

22nd October 2019

LGC Report Number: OA_19_3114_qNMR assay

Prepared by:

Eli Achtar 22/10/2019 — Eli Achtar

Checked by:

C. Clarkson 22 Oct 2019 — Cailean Clarkson

Approved by:

C. Clarkson — Cailean Clarkson, Science Leader
Purity and Calibration

Date: 22 Oct 2019.

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.docx

Contents

Revision History

Date	Version	Description	Author
22nd October 2019	1	First version	Eli Achtar

Client
Dima Almekdad, Organic Analysis Team, LGC Teddington

Analyst
Eli Achtar, Purity and Calibration, Teddington, Lab 6.15

LGC data reference
Purity and Calibration, Lab 6.15, LNB Organic 0046, page 142
Date of sample receipt: 1st September 2019
Description of sample: White powder
Sample storage: Approximately 5°C
Date of analysis: 20th September, 4th and 17th October 2019

Sample information
A sample of Normetanephrine hydrochloride (Lot: SHBD2227V) was received for quantitative NMR (qNMR) analysis by SOP INS/B1-0416 issue 5 under UKAS calibration accreditation schedule 0423. Any opinion and interpretation within this report falls outside the scope of this accreditation. The sample was registered as OA/19/3114 and structure confirmation by NMR and assay by qNMR were performed on the sample. The reference numbers for the sample and corresponding NMR experiments are given in Table 1.

Table 1. Reference numbers for the sample and NMR experiments

Sample Reference	PCS Reference	NMR Experiment	NMR Dataset Reference
DL-Normetanephrine HCl	OA/19/3114	Identification	
Lot: SHBD2227V		^{1}H	OA_19_3114_Sep20-2019 14
		COSY	OA_19_3114_Sep20-2019 15
		HSQC	OA_19_3114_Sep20-2019 16
Supplier: Sigma Aldrich		HMBC	OA_19_3114_Sep20-2019 17
		qNMR Assay	
		Rep1	OA_19_3114_Rep 1_Oct04-2019 10
			OA_19_3114_Rep 1_Oct04-2019 11
			OA_19_3114_Rep 1_Oct04-2019 12
		Rep2	OA_19_3114_Rep 2_Oct04-2019 10
			OA_19_3114_Rep 2_Oct04-2019 11
			OA_19_3114_Rep 2_Oct04-2019 12
		Rep3	OA_19_3114_Rep 3_Oct04-2019 10
			OA_19_3114_Rep 3_Oct04-2019 11
			OA_19_3114_Rep 3_Oct04-2019 12
		Rep4	OA_19_3114_Rep 4_Oct17-2019 10
			OA_19_3114_Rep 4_Oct17-2019 11
			OA_19_3114_Rep 4_Oct17-2019 12
		Stability	
		^{1}H (T=0 hrs)	OA_19_3114_Rep 1_Oct04-2019 10
		^{1}H (T ~5 days)	OA_19_3114_Rep 1_Oct09-2019 10

Sample preparation – structure confirmation
Approximately 10 mg of sample and Benzoic acid was dissolved in 1 mL of dimethylsulfoxide-d_6 (DMSO-d_6, Sigma Aldrich) and 3 drops of D_2O were added. The solution was vortexed, sonicated and transferred to a 5 mm NMR tube for analysis.

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.docx

NMR analysis – structure confirmation
To enable structural characterisation of the sample, ^{1}H NMR spectra and correlation spectroscopy (COSY), heteronuclear single quantum coherence (HSQC) and heteronuclear multi bond correlation (HMBC) spectra were obtained using a Bruker Avance 600 MHz NMR spectrometer with a 5 mm broadband inverse probe.

For the ^{1}H NMR analysis, the following experimental parameters were used:

Number of scans: 128
Receiver delay: 1 second
Temperature: 298 K
Spectral width: 20.0 ppm

Results – structure confirmation
The ^{1}H spectrum assignments of Normetanephrine are given in Table 2, corresponding to the numbering system shown in Figure 1. The NMR spectra are shown in Appendix 1 to 4. The NMR data shows one main component in sample OA/19/3114 and is consistent with Normetanephrine.

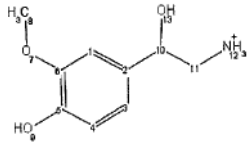


Figure 1. Structure of Normetanephrine

Table 2. Assignment of ^{1}H NMR signals of the main component

Chemical Shift /ppm	No of Hydrogens	Multiplicity	Assignments
6.92	1	doublet	H1
6.76	2	multiplet	H3, H4
4.67	1	double doublet	H10
3.75	3	singlet	H8
3.64	-	-	Water
2.95	1	double doublet	H11a
2.83	1	double doublet	H11b
2.50	-	quintet	DMSO-d_5

Method development – qNMR
The use of Benzoic acid (NIST, PS1) as an internal standard in DMSO-d_6 was found to be suitable with a few drops of D_2O. The longitudinal relaxation time (T1) of Normetanephrine with Benzoic acid in DMSO-d_6 was investigated and the longest T1 value for a signal used for quantitation was 2.89 sec. Sufficiently long delay times (D1) were used in the standard qNMR experiments to cover this. The solution was shown to be stable over days.

The analyte signal used for quantitation did not appear to overlap with any impurity signals and is representative of the Normetanephrine molecule.

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.docx

Sample preparation – qNMR
Approximately 10 mg of Normetanephrine and 10 mg of the Benzoic acid (NIST, PS1) internal standard were accurately weighed and 1 mL of DMSO-d_6 with 3 drops of D_2O was added. The resultant solution was vortexed and sonicated for one minute and transferred to a 5 mm NMR tube for analysis. Three independent solutions were prepared and qNMR experiments were acquired in triplicate for each solution on 04/10/2019. A further solution was prepared and qNMR experiments were acquired in triplicate for this solution on 17/10/2019. This was done to confirm previous results.

NMR analysis – qNMR
The ^{1}H qNMR analyses were performed on a Bruker Avance 600 MHz NMR. The following experimental parameters were used:

Number of scans:	16
Relaxation delay:	60 seconds
Spectral width:	20.0 ppm
Temperature:	298.0 K
FID processing software:	Topspin 3.5 pl 2
Baseline correction:	Manual, polynomial baseline correction
Signal integration:	Manual, including ^{13}C satellites
Analyte signal:	4.68 ppm
Standard signal:	7.55 ppm
Sino value, Analyte signal:	3214
Sino value, Standard signal:	18020

The assay results are summarised in Table 3 and an example NMR spectrum is shown in Appendix 5. All calculations can be found in:

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.xlsx

Results – qNMR

Table 3. Results of ^{1}H NMR assay for OA/19/3114

qNMR Replicate	qNMR Experiment	Experiment % Purity (m/m)	Replicate Mean % Purity (m/m)
1	10 11 12	82.38 82.36 82.32	82.35
2	10 11 12	82.10 82.08 82.07	82.08
3	10 11 12	81.74 81.77 81.79	81.77
4	10 11 12	82.14 82.16 82.18	82.16
		Mean% Purity (m/m)	**82.09**

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.docx

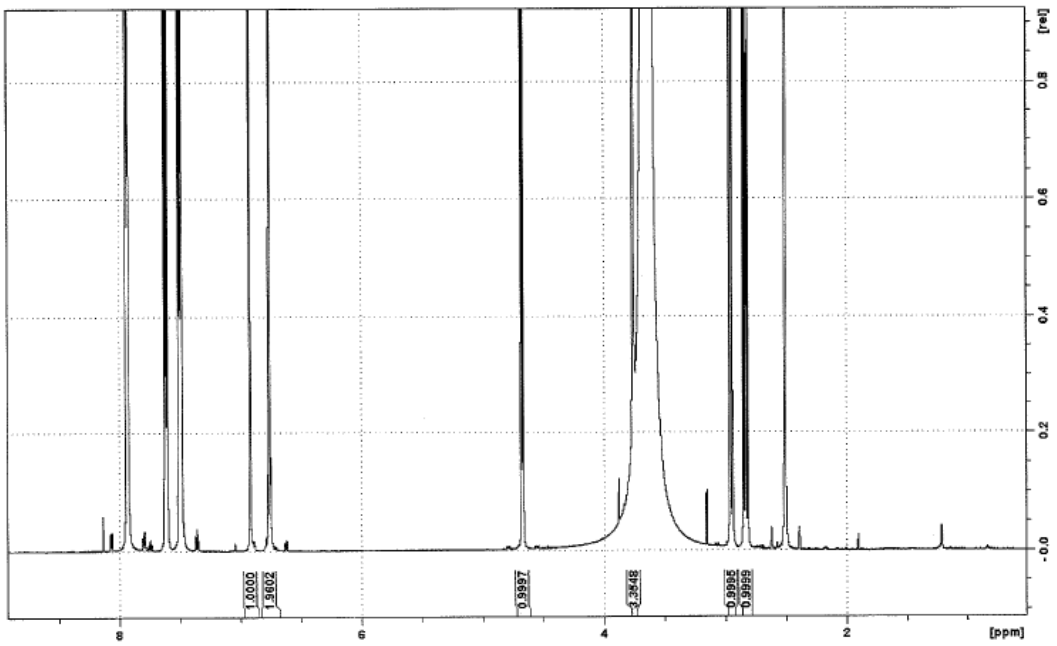
Uncertainty
The calculated uncertainty for the qNMR result was calculated as outlined in SOP INS/B1-0416 issue 5. The uncertainty calculations can be found in the below file:

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3114_qNMR Assay.xlsx

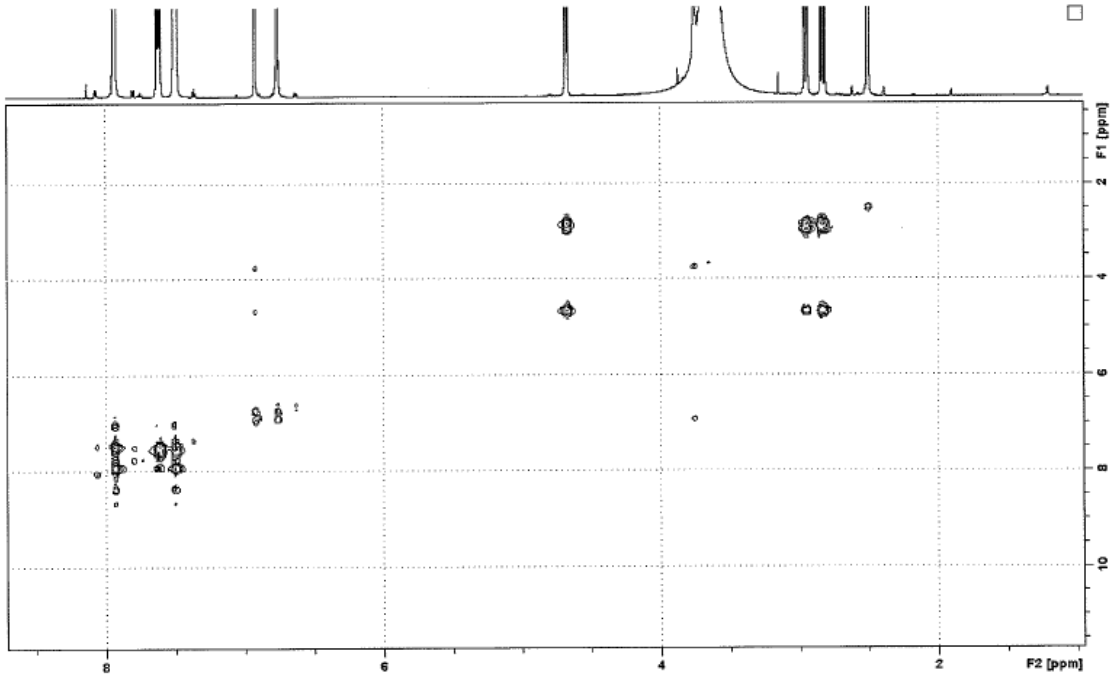
Traceability
Internal standard: Benzoic acid Standard Reference Material, NIST
Batch: PS1
Certified value: 99.992 ± 0.006% mass
Source of traceability: Traceability to the International System of Units (SI Units) via the kilogram is realised through the use of ISO 17025 calibrated balances and an internal standard certified for purity by % mass.

Conclusion
The NMR data is consistant with a salt of Normetanephrine. The purity of Normetanephrine Lot: SHBD2227V (OA/19/3114), as the free base, was determined to be 82.09% m/m ± 0.61% at the 95% confidence interval (k= 1.98) by qNMR assay using an internal standard.

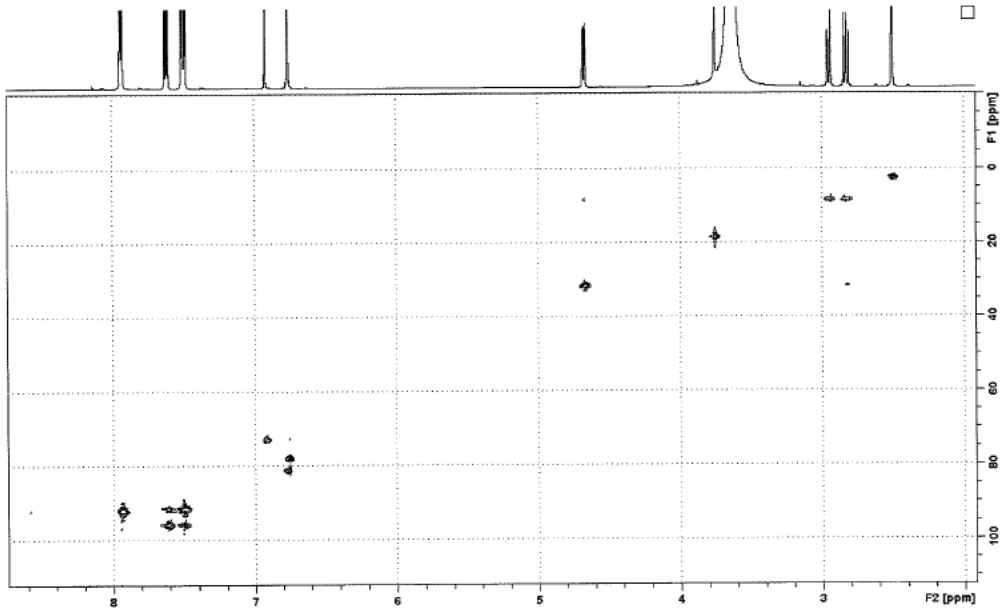
Appendix 1 ^{1}H NMR spectrum of OA/19/3114 with Benzoic acid in DMSO-d_6



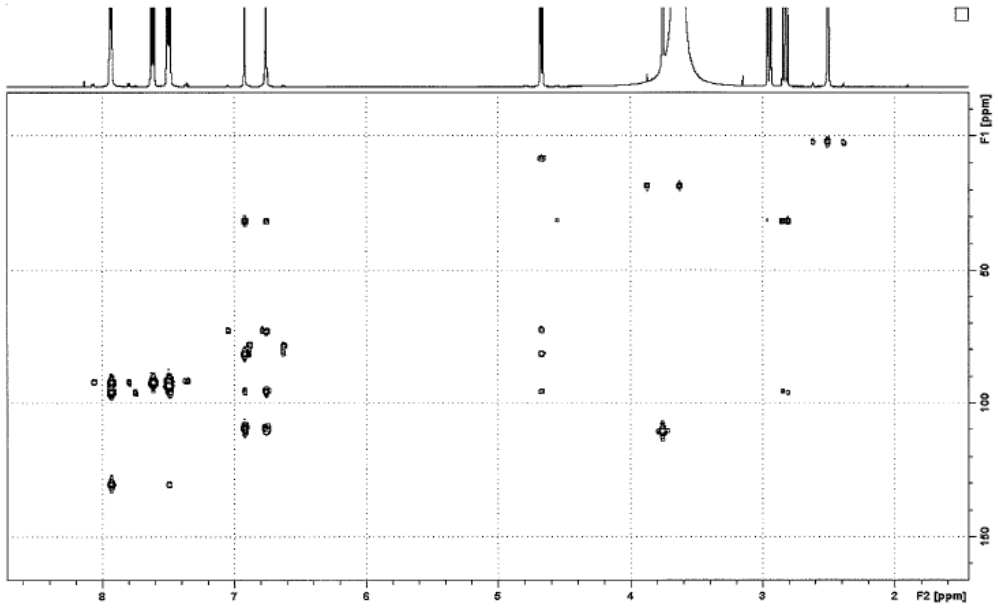
Appendix 2 COSY spectrum of OA/19/3114 with Benzoic acid in DMSO-d_6



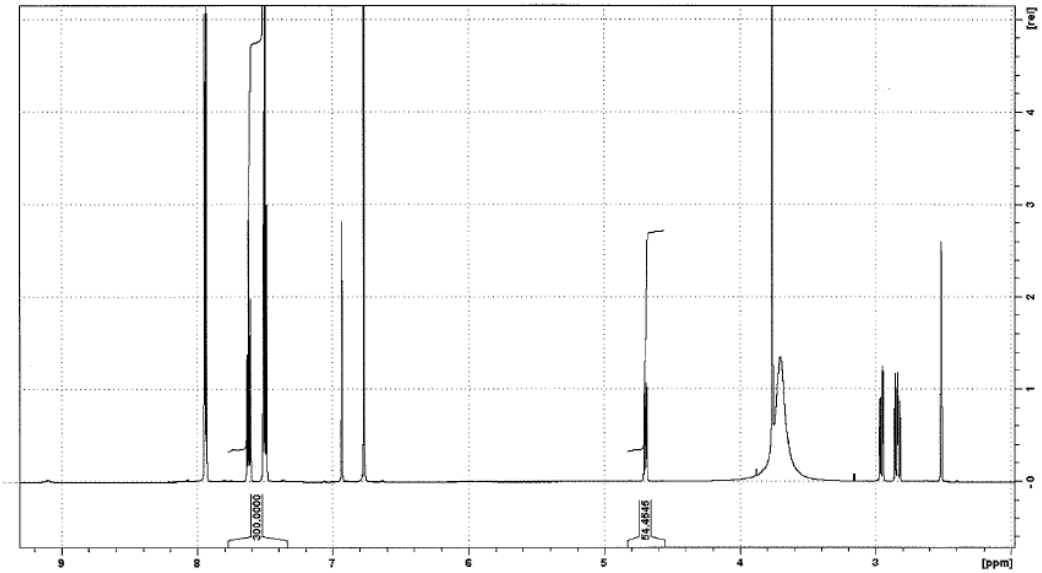
Appendix 3 HSQC spectrum of OA/19/3114 with Benzoic acid in DMSO-d_6



Appendix 4 HMBC spectrum of OA/19/3114 with Benzoic acid in DMSO-d_6



Appendix 5 ^{1}H NMR spectrum of OA/19/3114 with Benzoic acid in DMSO-d_6



7.1.4 Certificate of the traceable purity analysis of Metanephrine solid standard by the NML

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.docx

LGC

UKAS CALIBRATION 0423

^{1}H qNMR Assay of a Metanephrine salt Lot: CS-ME-AAA-0530-01 (OA/19/3115), using SOP INS/B1-0416 (version 5)

15th October 2019

LGC Report Number: OA_19_3115_qNMR assay

Prepared by:

Eli Achtar 15.10.2019

Eli Achtar

Checked by:

C. Clarkson 15 Oct 2019

Cailean Clarkson

Approved by:

C. Clarkson

Cailean Clarkson,
Science Leader, Purity and Calibration

Date: 15 Oct 2019.

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.docx

Contents

Revision History

Date	Version	Description	Author
15th October 2019	1	First version	Eli Achtar

Client
Dima Almekdad, Organic Analysis Team, LGC Teddington

Analyst
Eli Achtar, Purity and Calibration, Teddington, Lab 6.15

LGC data reference
Purity and Calibration, Lab 6.15, LNB Organic 0046, page 145
Date of sample receipt: 1st September 2019
Description of sample: White powder
Sample storage: Approximately 5°C
Date of analysis: 20th September and 4th October 2019

Sample information
A sample of Metanephrine hydrochloride salt (Lot: CS-ME-AAA-0530-01) was received for quantitative NMR (qNMR) analysis by SOP INS/B1-0416 issue 5 under UKAS calibration accreditation schedule 0423. Any opinion and interpretation within this report falls outside the scope of this accreditation.This sample was registered as OA/19/3115 and structure confirmation by NMR and assay by qNMR were performed on the sample. The reference numbers for the sample and corresponding NMR experiments are given in Table 1.

Table 1. Reference numbers for the sample and NMR experiments

Sample Reference	PCS Reference	NMR Experiment	NMR Dataset Reference
Metanephrine HCl	OA/19/3115	Identification	
Lot: CS-ME-AAA-0530-01		^{1}H	OA_19_3115_Sep20-2019 14
		COSY	OA_19_3115_Sep20-2019 15
		HSQC	OA_19_3115_Sep20-2019 16
Supplier: Clear Synth		HMBC	OA_19_3115_Sep20-2019 17
		qNMR Assay	
		Rep1	OA_19_3115_Rep 1_Oct04-2019 10
			OA_19_3115_Rep 1_Oct04-2019 11
			OA_19_3115_Rep 1_Oct04-2019 12
		Rep2	OA_19_3115_Rep 2_Oct04-2019 10
			OA_19_3115_Rep 2_Oct04-2019 11
			OA_19_3115_Rep 2_Oct04-2019 12
		Rep3	OA_19_3115_Rep 3_Oct04-2019 10
			OA_19_3115_Rep 3_Oct04-2019 11
			OA_19_3115_Rep 3_Oct04-2019 12
		Stability	
		^{1}H (T=0 hrs)	OA_19_3115_Rep 1_Oct04-2019 10
		^{1}H (T ~5 day)	OA_19_3115_Rep 1_Oct09-2019 10

Sample preparation – structure confirmation
Approximately 10 mg of sample and Benzoic acid were dissolved in 1 mL of dimethylsulfoxide-d_6 (DMSO-d_6, Sigma Aldrich) and 3 drops of D_2O were added. The solution was vortexed, sonicated and transferred to a 5 mm NMR tube for analysis.

NMR analysis – structure confirmation
To enable structural characterisation of the sample, ^{1}H NMR spectra and correlation spectroscopy (COSY), heteronuclear single quantum coherence (HSQC) and

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.docx

heteronuclear multi bond correlation (HMBC) spectra were obtained using a Bruker Avance 600 MHz NMR spectrometer with a 5 mm broadband inverse probe.

For the ^{1}H NMR analysis, the following experimental parameters were used:

Number of scans:	128
Receiver delay:	1 second
Temperature:	298 K
Spectral width:	20.0 ppm

Results – structure confirmation
The ^{1}H spectrum assignments of Metanephrine are given in Table 2, corresponding to the numbering system shown in Figure 1. The NMR spectra are shown in Appendix 1 to 4. The NMR data shows one main component in sample OA/19/3115 and is consistent with Metanephrine.

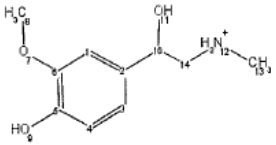


Figure 1. Structure of Metanephrine

Table 2. Assignment of ^{1}H NMR signals of the main component

Chemical Shift /ppm	No of Hydrogens	Multiplicity	Assignments
6.92	1	doublet	H1
6.76	2	multiplet	H3, H4
4.75	1	double doublet	H10
3.75	3	singlet	H8
3.76	-	-	Water
3.04	1	double doublet	H14a
2.98	1	double doublet	H14b
2.57	3	singlet	H13
2.50	-	quintet	DMSO-d_5

Method development – qNMR
The use of Benzoic acid (NIST, PS1) as an internal standard in DMSO-d_6 was found to be suitable with a few drops of D_2O. The longitudinal relaxation time (T1) of Metanephrine with Benzoic acid in DMSO-d_6 was investigated and the longest T1 value for a signal used for quantitation was 3.20 sec. Sufficiently long delay times (D1) were used in the standard qNMR experiments to cover this. The solution was shown to be stable over 4 days.

The analyte signal used for quantitation did not appear to overlap with any impurity signals and is representative of the Metanephrine molecule.

Sample preparation – qNMR
Approximately 10 mg of Metanephrine and 10 mg of the Benzoic acid (NIST, PS1) internal standard were accurately weighed and 1 mL of DMSO-d_6 with 3 drops of D_2O was added. The resultant solution was vortexed and sonicated for one minute and

transferred to a 5 mm NMR tube for analysis. Three independent solutions were prepared and qNMR experiments were acquired in triplicate for each solution.

NMR analysis – qNMR
The ^{1}H qNMR analyses were performed on a Bruker Avance 600 MHz NMR. The following experimental parameters were used:

Number of scans:	16
Relaxation delay:	60 seconds
Spectral width:	20.0 ppm
Temperature:	298.0 K
FID processing software:	Topspin 3.5 pl 2
Baseline correction:	Manual, polynomial baseline correction
Signal integration:	Manual, excluding ^{13}C satellites
Analyte signal:	4.75 ppm
Standard signal:	7.93 ppm
Sino value, Analyte signal:	2972
Sino value, Standard signal:	17438

The assay results are summarised in Table 3 and an example NMR spectrum is shown in Appendix 5. All calculations can be found in:

https://mylgc2010/scienceandtech/measres/mrteams/oateam/OrganicAnalysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.xlsx

Results – qNMR

Table 3. Results of ^{1}H NMR assay for OA/19/3115

qNMR Replicate	qNMR Experiment	Experiment % Purity (m/m)	Replicate Mean % Purity (m/m)
1	10	81.70	81.70
	11	81.70	
	12	81.70	
2	10	81.22	81.23
	11	81.21	
	12	81.25	
3	10	81.58	81.55
	11	81.55	
	12	81.53	
		Mean% Purity (m/m)	**81.49**

Uncertainty
The calculated uncertainty for the qNMR result was calculated as outlined in SOP INS/B1-0416 issue 5. The uncertainty calculations can be found in the below file:

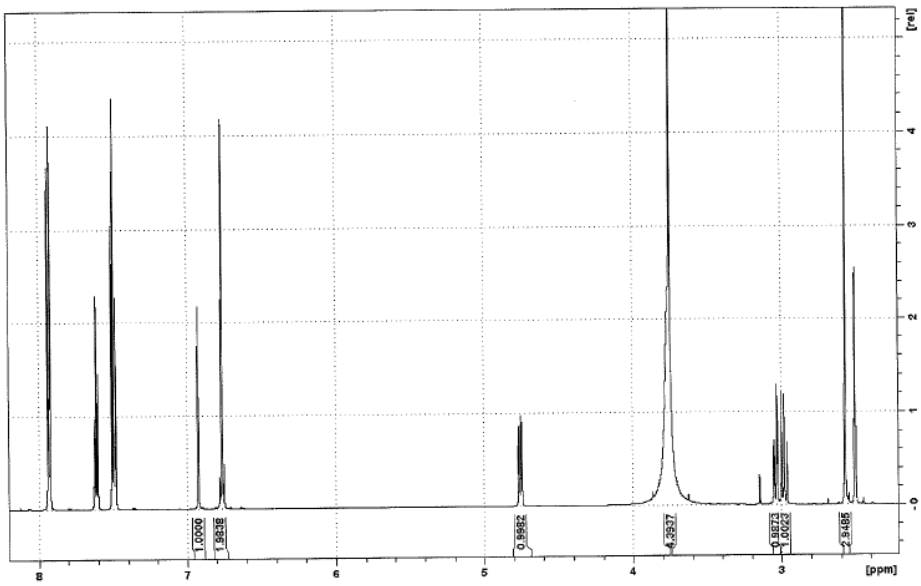
https://mylgc2010/scienceandtech/measres/mrteams/oateam/OrganicAnalysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.xlsx

https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.docx

Traceability
Internal standard: Benzoic acid Standard Reference Material, NIST
Batch: PS1
Certified value: 99.992 ± 0.006% mass
Source of traceability: Traceability to the International System of Units (SI Units) via the kilogram is realised through the use of ISO 17025 calibrated balances and an internal standard certified for purity by % mass.

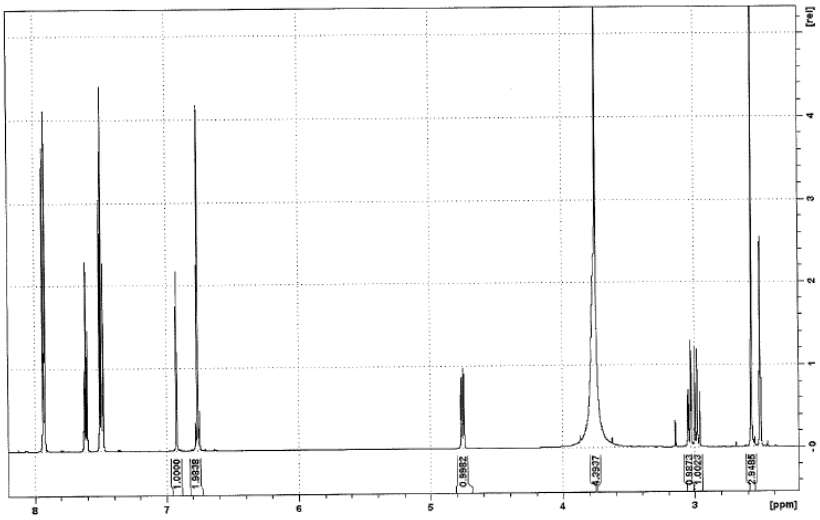
Conclusion
The NMR data is consistant with a salt of Metanephrine. The purity of Metanephrine Lot: CS-ME-AAA-0530-01 (OA/19/3115), as the free base, was determined to be 81.49% m/m ± 0.63% at the 95% confidence interval (k= 2.01) by qNMR assay using an internal standard.

Appendix 1 ^{1}H NMR spectrum of OA/19/3115 with Benzoic acid in DMSO-d_6



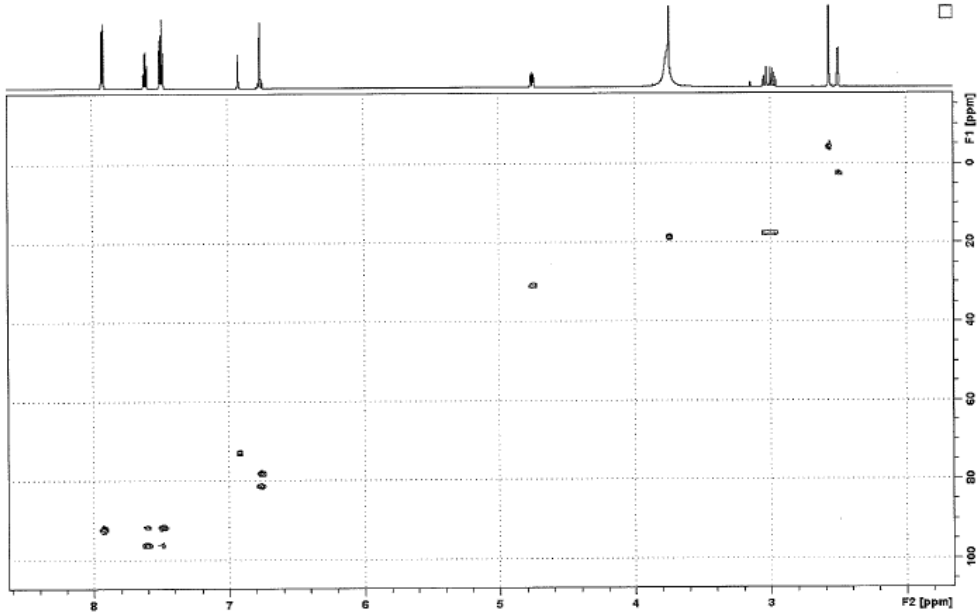
https://mylgc2010/scienceandtech/measres/mrteams/oateam/Organic Analysis Reports/Purity and Calibration/OA_19_3115_qNMR Assay.docx

Appendix 1 ^{1}H NMR spectrum of OA/19/3115 with Benzoic acid in DMSO-d_6

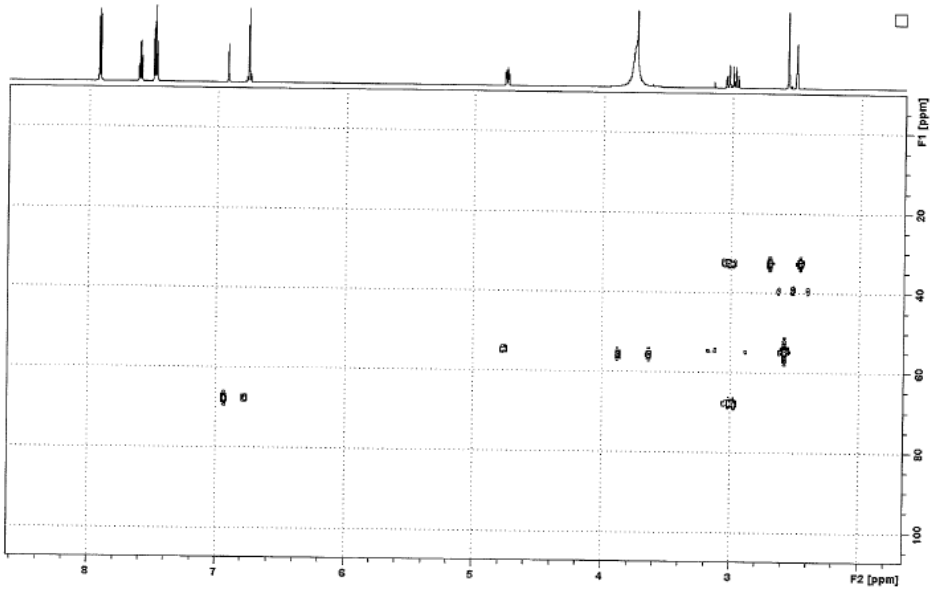


OA_19_3115_qNMR assay
Page 7 of 11
LGC Limited, Queens Road, Teddington, Middlesex TW11 0LY

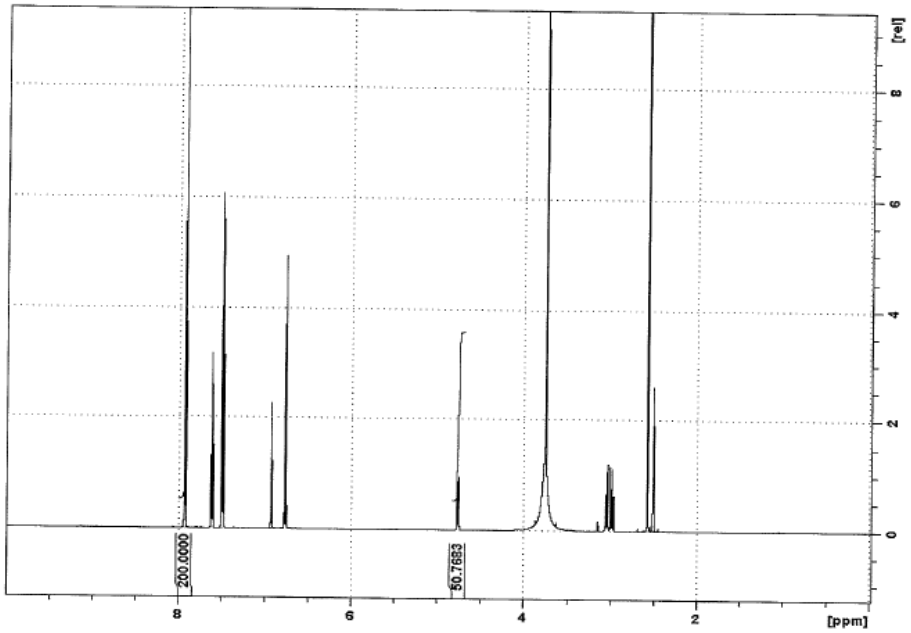
Appendix 3 HSQC spectrum of OA/19/3115 with Benzoic acid in DMSO-d_6



Appendix 4 HMBC spectrum of OA/19/3115 with Benzoic acid in DMSO-d_6



Appendix 5 ^{1}H NMR spectrum of OA/19/3115 with Benzoic acid in DMSO-d_6



7.1.5 CoA of the higher-order certified reference material of pure tacrolimus

For every ERM DA110a vial used a signed certificate of analysis was provided. Below is an example of the CoA.

LGC

ERM European Reference Materials

CERTIFICATE OF ANALYSIS

ERM®- AC022a

Tacrolimus		
	Certified Value [1, 2] (% mass)	Uncertainty [3] (% mass)
Purity	97.65	0.68

1) The measurand is defined as the mass fraction of anhydrous tacrolimus.

2) The certified value is traceable to the SI.

3) The quoted uncertainty is the half-width of the expanded uncertainty interval, calculated using a coverage factor (*k*) of 2, which gives a level of confidence of approximately 95 %.

This certificate is valid for 12 months from the date of shipment provided the sample is stored under the recommended conditions.

The minimum amount of sample to be used is 20 mg.

NOTE

European Reference Material ERM®-AC022a was produced and certified under the responsibility of LGC according to the principles laid down in the Technical Guidelines of the European Reference Materials® co-operation agreement between BAM-LGC-IRMM. Information on these guidelines is available on the Internet (http://www.erm-crm.org).

Accepted as an ERM®, Teddington, September 2016.
Certificate revised December 2017

Signed: _____

Dr Derek Craston, UK Government Chemist
LGC Limited
Queens Road
Teddington
Middlesex
TW11 0LY, UK

UKAS TESTING 4005

ERM® - AC022a
Page 1 of 5

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

SPECIMEN

ERM European Reference Materials

DESCRIPTION OF THE SAMPLE

This material was produced from a batch of tacrolimus monohydrate in powder form kindly donated by Sandoz International GmbH. Portions of at least 100 mg were dispensed into 1.25 mL amber glass vials with PTFE lined screw caps, and sealed in plastic bags containing desiccant.

Data obtained from the analysis of the material by NMR spectroscopy were consistent with the structure of tacrolimus and literature values, providing confirmation of its identity.

Tacrolimus IUPAC name:

[3S[3R*[E(1S*,3S*,4S*)],4S*,5R*,8S*,9E,12R*,14R*,15S*,16R*,18S*,19S*,26aR*]] 5,6,8,11,12,13,14,15,16,17,18,19,24,25,26,26a-hexadecahydro-5,19-dihydroxy-3-[2-(4-hydroxy-3-methoxycyclohexyl)-1-methylethenyl]-14,16-dimethoxy-4,10,12,18-tetramethyl-8-(2-propenyl)-15,19-epoxy-3H-pyrido[2,1-c][1,4]oxaazacyclotricosine-1,7,20,21(4H,23H)-tetrone

INTENDED USE

The intended use of this certified reference material is for the calibration of instruments, quality control and the validation of methods to determine the immunosuppressant drug tacrolimus. It can also be used in the training and evaluation of staff.

ANALYTICAL METHODS USED FOR CHARACTERISATION

The approach used to calculate the certified purity value entailed quantification of the impurities detected using a range of techniques. Water content was quantified by oven coulometric KF, HPLC-UV was used for organic impurities and TGA for inorganic residues and residual solvents. The purity value obtained was confirmed by qNMR. TLC was used to confirm the absence of large bio-organic molecules and polymers.

Organic impurities

HPLC

Organic impurities in the material were quantified using HPLC-UV with an Agilent Zorbax SB C18 column (4.6x150 mm, 3.5 µm) and the following gradient:

Time (min)	% A (90:10 H_2O:acetonitrile + acetic acid)	% B (100 acetonitrile + acetic acid)
0	60	40
30	55	45
50	30	70
50.1	60	40
60	60	40

The flow rate was 2.3 mL/min with a column temperature of 60 °C. The purity of the material was quantified by peak area normalisation using UV detection at λ = 220 nm.

In aqueous solution, tacrolimus exists as a mixture of three isomers that were resolved by the chromatographic method used.

The sum of the areas of the three peaks was used for quantification (see Figure 1).

ERM® - AC022a
Page 2 of 6

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

SPECIMEN

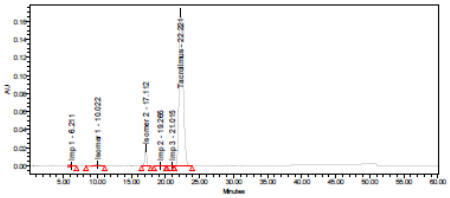


Figure 1: Typical HPLC chromatogram of ERM®-AC022a at 220 nm

The organic purity of ERM®-AC022a was found to be 99.81 % mass with a standard uncertainty of 0.33 % mass.

The method was tested for co-elution using HPLC-MS, and experimental information was obtained on the molecular weight and UV response factor of the organic impurities to confirm the integrity of the quantification approach taken. Based on this supplementary information, and also to allow for any unresolved peaks, contributions were added to the combined uncertainty associated with the certified value.

TLC

This technique was used to demonstrate the absence of significant bio-organic residues or polymers. This information was used for qualitative purposes only and the information was not incorporated in the calculation of the purity of the material.

Water content (% water)

The water content was determined using 10 units by coulometric KF titration. The average water content was found to be 2.16 % mass with a standard uncertainty of 0.09 % mass. This amount of water is consistent with the monohydrate state of the raw material.

Inorganic residues (% inorganic)

The inorganic residue content was determined using 3 units by TGA. The average inorganic residue content was found to be 0.00 % mass with a standard uncertainty of 0.04 % mass.

Residual solvents (% solvents)

The solvent residue content was determined from 3 units using TGA. The average solvent residue content was found to be 0.00 % mass with a standard uncertainty of 0.13 % mass.

Calculation of the Purity Value

The following formula was used to calculate the purity value:

$$\text{Purity} = \left[1 - \left[\frac{\%\ \text{water}}{100} + \frac{\%\ \text{inorganic}}{100} + \frac{\%\ \text{solvents}}{100}\right]\right] \times \text{organic purity}$$

Confirmation of Identity and Purity Value

The identity of the material was confirmed using ^{1}H NMR spectroscopy. A validated internally standardized ^{1}H NMR assay was used to confirm the purity value. The value obtained using qNMR was (97.67 ± 0.71) % mass (k = 2.26), which agrees with the certified value within the stated expanded uncertainties.

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

HOMOGENEITY

The material was considered to be homogeneous on the basis of HPLC-UV and KF measurements on ten randomly selected units analysed in triplicate.

STABILITY

A short-term stability study was carried out on the material where units in their final sale format were exposed to 4 °C, 18 °C, 37 °C and 60 °C temperatures for periods of up to four weeks. The results of the study showed no evidence of significant deterioration at elevated temperatures over a period of four weeks. Based on these results, the material is shipped at ambient temperature.

HPLC and KF measurements were carried out over a period of six months on samples stored at -20 °C, 4 °C, 18 °C and 37 °C. There was evidence of significant instability at 37 °C in terms of the water content of the material. However, no evidence was found of any significant change in the value of organic purity at any of the temperatures tested. Any potential instability has been accounted for by the addition of a stability component to the combined uncertainty. Post-certification stability monitoring tests will be performed on the material.

No data was available for the assessment of the stability of the material once a unit has been opened. Should users wish to use the material repeatedly over a period of time, they are advised to carry out their own stability assessment.

MEASUREMENT UNCERTAINTY

The magnitude of each of the components that contribute to the uncertainty associated with the certified value is given in the table below:

Standard uncertainty (u, % mass)	Characterisation	Homogeneity	Stability	Combined u
	0.33	0.027	0.11	0.34

TRACEABILITY

The certified value is traceable to the SI through appropriate calibration and control of the instrumentation used in accordance with the requirements of ISO/IEC 17025 for calibration. KF and TGA measurements are traceable to the SI through the use of NIST SRM 2890 and E2 calibrated weights traceable to the National Primary Standard of mass via the UK National Physical Laboratory (NPL).

LGC's capability to produce SI traceable measurements of purity is verified by participation in international key comparison studies organised by the Organic Analysis Working Group of the Consultative Committee for Amount of Substance (CCQM, http://www.bipm.org/metrology/chemistry-biology/).

SAFETY INFORMATION

According to Regulation (EC) No 1272/2008 this material is classified as toxic if swallowed. Refer to material safety data sheet for further safety information.

INSTRUCTIONS FOR USE

Before opening, the contents should be brought to ambient temperature (20 ± 5) °C and thoroughly mixed by inverting the vial several times.

Experience with the material at LGC has found it to be soluble in acetonitrile. Solutions of 3 mg/mL were used for the homogeneity study. When in solution, attention should be paid to isomers inter-conversion. Furthermore, the spiking of a blank blood matrix with a 50 µg/mL acetonitrile solution prepared with a previous batch of the material led to the successful production of a homogeneous matrix material.

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

STORAGE

The material should be stored at (-20 ± 10) °C in the original closed vial.

ABBREVIATIONS

HPLC-UV: High Performance Liquid Chromatography with UV detection

KF: Karl Fischer

TGA: Thermo Gravimetric Analysis

qNMR: quantitative Nuclear Magnetic Resonance

HPLC-MS: High Performance Liquid Chromatography with Mass Spectrometry detection

TLC: Thin Layer Chromatography

CERTIFICATE REVISION

In December 2017 revisions were made to the certificate as follows:

- The definition of the measurand was expanded in footnote 1 of the certified values box
- The ANALYTICAL METHODS USED FOR CHARACTERISATION was expanded
- The DESCRIPTION OF THE SAMPLE section was expanded
- The STABILITY section was expanded
- The SAFETY INFORMATION section was expanded
- The INSTRUCTIONS FOR USE section was expanded

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

Unit Number:	Shipment Date:

LEGAL NOTICE

The values quoted in this certificate are the best estimate of the true values within the stated uncertainties and based on the techniques described herein. No warranty or representation, express or implied, is made that the use of the product or any information, material, apparatus, method or process which is the subject of or referred to in this certificate does not infringe any third party rights. Further, save to the extent: (a) prohibited by law; or (b) caused by a party's negligence; no party shall be liable for the use made of the product, any information, material, apparatus, method or process which is the subject of or referred to in this certificate. In no event shall the liability of any party exceed whichever is the lower of: (i) the value of the product; or (ii) £500,000; and any liability for loss of profit, loss of business or revenue, loss of anticipated savings, depletion of goodwill, any third-party claims or any indirect or consequential loss or damage in connection herewith is expressly excluded.

Production of this reference material and certificate was funded by the UK government Department for Business, Energy & Industrial Strategy (BEIS).

This certificate may not be published except in full, unless permission for the publication of an approved extract has been obtained in writing from LGC Limited. It does not of itself impute to the subject of measurement any attributes beyond those shown by the data contained herein.

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

7.2 Appendix 3 – certificate of analysis of biological QC materials used

7.2.1 Plasma METs QCs certificates of analysis

Australian Scientific Enterprise
Ph (+61) 2 9477 3870 | Fax (+61) 2 9477 7646
83 Hunter St Hornsby NSW 2077

Plasma Metanephrines Control Level 1

STORE REFRIGERATED (2-8°C)

CAUTION: This material is prepared from human-based materials. Although source materials have been tested for certain diseases, no test can absolutely assure the absence of all infectious agents. Therefore, this product should be handled as a potential biohazard. For in vitro diagnostic use only.

Intended Use
The ASE Plasma Metanephrines controls are non-assay specific controls intended for professional use in monitoring the precision of the laboratory method and technique for Plasma Metanephrines.

Preparation
Allow material to come to room temperature before reconstituting. Reconstitute the material with exactly 1.5mL deionised water. Let stand for 15 minutes and swirl gently to ensure homogeneity.

Material Stability

The lyophilised product is stable until the expiry stated on the vial. Once reconstituted the material is stable for one day at 2-8°C.

Manufacturing Target Values

Batch 21331	Target
Metanephrine	0.05-0.25 nmol/L
Normetanephrine	0.5-0.7 nmol/L
3-Methoxytyramine (3-MT)	0.05-0.18 nmol/L

Contact & Ordering Information
Please contact us with any queries or orders via the following channels:

Product Orders: Sales@ASEonline.com.au
Technical Support: Support@ASEonline.com.au
General Queries: Info@ASEonline.com.au

Or visit our website for our full product range:

www.ASEonline.com.au

Cat No: 907115 Revision Jan-19 **12 x 1.5mL vials**

Australian Scientific Enterprise
Ph (+61) 2 9477 3870 | Fax (+61) 2 9477 7646
83 Hunter St Hornsby NSW 2077

Plasma Metanephrines Control Level 2

STORE REFRIGERATED (2-8°C)

CAUTION: This material is prepared from human-based materials. Although source materials have been tested for certain diseases, no test can absolutely assure the absence of all infectious agents. Therefore, this product should be handled as a potential biohazard. For in vitro diagnostic use only.

Intended Use
The ASE Plasma Metanephrines controls are non-assay specific controls intended for professional use in monitoring the precision of the laboratory method and technique for Plasma Metanephrines.

Preparation
Allow material to come to room temperature before reconstituting. Reconstitute the material with exactly 1.5mL deionised water. Let stand for 15 minutes and swirl gently to ensure homogeneity.

Material Stability

The lyophilised product is stable until the expiry stated on the vial. Once reconstituted the material is stable for one day at 2-8°C.

Manufacturing Target Values

Batch 21332	Target
Metanephrine	1.0-1.3 nmol/L
Normetanephrine	2.7-3.1 nmol/L
3-Methoxytyramine (3-MT)	2.5-3.5 nmol/L

Contact & Ordering Information
Please contact us with any queries or orders via the following channels:

Product Orders: Sales@ASEonline.com.au
Technical Support: Support@ASEonline.com.au
General Queries: Info@ASEonline.com.au

Or visit our website for our full product range:

www.ASEonline.com.au

Cat No: 907215 Revision Jan-19 **12 x 1.5mL vials**

7.2.2 Human Blood - Tacrolimus ERM DA11-a

For every ERM DA110a vial used a signed certificate of analysis was provided. Below is an example of the CoA.

CERTIFICATE OF ANALYSIS

ERM®- DA110a

Human Blood - Tacrolimus		
Constituent	Certified Value [1] (µg/kg)	Uncertainty [2] (µg/kg)
Tacrolimus	7.41	0.25

1) The certified value is the mass fraction of tacrolimus determined using exact matching isotope dilution mass spectrometry and is traceable to the SI through the use of a traceable, high accuracy tacrolimus standard for instrument calibration.

2) The quoted uncertainty is the half-width of the expanded uncertainty interval, calculated using a coverage factor (k) of 2.3, which gives a level of confidence of approximately 95 %.

This certificate is valid for 12 months from the date of shipment provided the sample is stored under the recommended conditions.

The minimum amount of sample to be used is 0.33 g.

NOTE

European Reference Material ERM®-DA110a was produced and certified under the responsibility of LGC according to the principles laid down in the Technical Guidelines of the European Reference Materials® co-operation agreement between BAM-LGC-IRMM. Information on these guidelines is available on the Internet (http://www.erm-crm.org).

Accepted as an ERM®, Teddington, September 2012.

Signed: _____

Dr Derek Craston, UK Government Chemist
LGC Limited
Queens Road
Teddington
Middlesex
TW11 0LY, UK

4005

ERM® - DA110a
Page 1 of 5

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

Additional Material Information		
Constituent	Concentration [1] (µg/L)	Uncertainty [2] (µg/L)
Tacrolimus	7.82	0.25

1) The concentration and uncertainty have been calculated using the density of whole blood (1.055 ± 0.002) kg/L, expanded uncertainty (k = 2) measured at LGC.

2) The quoted uncertainty is the half-width of the expanded uncertainty interval (k = 2.3), which gives a level of confidence of approximately 95 %.

DESCRIPTION OF THE SAMPLE

The material was prepared by Analytical Services International (London, UK) by spiking blank pooled human blood (provided by Biological Specialty Corporation, PA, USA), containing EDTA as a preservative, with 100 µL of a 50 µg/mL standard solution of tacrolimus in acetonitrile to give a nominal concentration of 8 µg/kg tacrolimus in blood. The blood had been pre-screened for HIV 1 & 2 antibodies, hepatitis B surface antigen and hepatitis C antibody, and found to be negative. The material was then mixed and aliquoted into 1 mL units in 2 mL cryovials and stored at (-70 ± 10) °C.

INTENDED USE

Subject to the commutability information given below, this reference material is intended for use in the calibration of instruments, the validation of new methods, and monitoring the performance of methods commonly used in clinical laboratories to determine the tacrolimus content of human blood samples. It can also be used in the training and evaluation of staff.

ANALYTICAL METHODS USED FOR CHARACTERISATION

Pure tacrolimus powder used as a calibrant was sourced from LC Laboratories (Woburn, USA). The purity was found to be (97.3 ± 1.3) % m/m using a combination of direct (qNMR) and indirect (100%-impurities) approaches. The organic purity was determined by reverse phase HPLC, the water content determined by Karl Fischer titration and solvents and inorganic residues determined by TGA (Thermo Gravimetric Analysis). Inorganic residues were confirmed by ICP-MS and ICP-OES. SEM/EDX analysis was carried out to determine elemental composition.

Isotopically labelled tacrolimus was obtained from ALSACHIM (Strasbourg, France).

Characterisation was carried out at LGC using exact matching isotope dilution mass spectrometry. All samples, calibrators and quality controls were prepared gravimetrically. To extract the tacrolimus from the blood, zinc sulfate in aqueous methanol was added and the mixture vortexed, which releases the tacrolimus into solution through effective lysing of the blood cells. The blood was centrifuged at 20 °C and 3500 g, the clear solution decanted and further diluted with deionised water to decrease the methanol content to less than 40 % allowing good retention on the C18 phase SPE cartridge. Sample clean up was carried out using a 3M Empore™ C18 SPE 96 well plate (Sigma Aldrich, Dorset, UK).

LC-MS/MS was carried out using a Thermo Scientific TSQ Vantage Tandem Mass Spectrometer with a TLX-1 liquid chromatography system (Thermo Scientific, Hypersil Gold C18, 50 x 3 mm, 5 µm particle size LC column) using a gradient methanol:water mobile phase. Two SRM transitions were acquired for both the natural and labelled tacrolimus.

ERM® - DA110a
Page 2 of 5

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

CONFIRMATORY DATA

Analytical Services International (St George's University Hospital, London, UK) provided confirmatory data for the material using routine LC-MS laboratory measurements.

Analytical Services International results

Constituent	Results µg/L	
	Concentration	*Standard error
Tacrolimus	7.80	0.14

*The standard error is based on observed variation in results and does not include uncertainties associated with, for example, calibration, stability or homogeneity.

Note: Using the value of 7.80 µg/L to convert to the mass fraction (µg/kg) and a density value of (1.055 ± 0.002) kg/L (expanded uncertainty (k = 2)) measured at LGC, a value of 7.39 µg/kg is obtained.

HOMOGENEITY

The material was tested for homogeneity by analysing randomly selected samples for tacrolimus using the characterisation method detailed above. The material was judged to be homogenous as the variation between samples was not significantly greater than the method variation.

STABILITY

Analyses have shown that deterioration is not anticipated over the lifetime of the material when stored under the recommended conditions; however ERM®-DA110a will be subjected to testing under the LGC stability monitoring programme. Purchasers will be informed of any changes affecting the certified value.

COMMUTABILITY

This material has been tested for commutability using the exact matching isotope dilution mass spectrometry reference method developed for its characterisation and routine laboratory LC-MS/MS applied to both the reference material and patient samples, and was found to be commutable. For use with other methods of analysis, the confirmation of commutability is the responsibility of the end user. In addition, attention is drawn to the following:

- The ERM was produced by spiking whole blood with tacrolimus. This material is not an 'incurred' pooled patients' blood material.
- In patient blood samples, low concentrations of tacrolimus metabolites will exist which may be detected and measured as part of the total 'tacrolimus' signal detected by some assays. These metabolites are not present in ERM®-DA110a.
- In patient whole blood samples, tacrolimus is found predominately within the erythrocytes, and a lysis stage is required to release tacrolimus for extraction and subsequent measurement; analysis of ERM®-DA110a will not assess the efficiency of the lysis process.
- The pure tacrolimus used to spike the whole blood was found to contain 0.4 g/100 g of ascomycin, giving a concentration of approximately 0.3 µg/L ascomycin in ERM®-DA110a. Any assay which uses ascomycin as an internal standard should take this into account.

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

ANALYTICAL METHOD USED FOR THE DETERMINATION OF DENSITY

The density of the whole blood was determined in order to calculate the concentration on a unit/volume basis. The weight of accurately pipetted aliquots of deionised water and whole blood were measured on an analytical balance. The volume of whole blood dispensed was calculated from the weight of deionised water, taking into account temperature and air pressure. The density of the whole blood was obtained by relating the weight observed with volume dispensed. The average ambient temperature of the biohazard cabinet during the analysis was 23.4 °C. The density value obtained for the whole blood was re-calculated using the thermal coefficient of blood to give a value at 20 °C. An oscillating-type density meter (KEM density DA-500) operated at 20 °C was used as a confirmatory technique.

SAFETY INFORMATION

The material is of human origin and should be handled with appropriate care. The material was tested for the presence of HIV 1 & 2 antibodies, hepatitis B surface antigen and hepatitis C antibody. All were found to be negative.

For further details refer to the material safety data sheet.

INSTRUCTIONS FOR USE

Prior to use, the material should be thawed by equilibrating at room temperature for at least 2 hours and mixed by inverting the vial several times. LGC has no data on the effect of further freezing and thawing of the material before use, and therefore it is recommended that, after thawing, the material is stored at refrigerator temperature (5 ± 4) °C, and used within one month.

STORAGE

The material should be stored at (-20 ± 5) °C in the original closed vial until it is first opened.

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com

Unit Number:	Shipment Date:

LEGAL NOTICE

The values quoted in this certificate are the best estimate of the true values within the stated uncertainties and based on the techniques described herein. No warranty or representation, express or implied, is made that the use of the product or any information, material, apparatus, method or process which is the subject of or referred to in this certificate does not infringe any third party rights. Further, save to the extent: (a) prohibited by law; or (b) caused by a party's negligence; no party shall be liable for the use made of the product, any information, material, apparatus, method or process which is the subject of or referred to in this certificate. In no event shall the liability of any party exceed whichever is the lower of: (i) the value of the product; or (ii) £500,000; and any liability for loss of profit, loss of business or revenue, loss of anticipated savings, depletion of goodwill, any third-party claims or any indirect or consequential loss or damage in connection herewith is expressly excluded.

Production of this reference material and certificate was funded by the UK National Measurement System.

This certificate may not be published except in full, unless permission for the publication of an approved extract has been obtained in writing from LGC Limited. It does not of itself impute to the subject of measurement any attributes beyond those shown by the data contained herein.

Queens Road, Teddington, Middlesex TW11 0LY Tel: +44 (0)20 8943 7000 Fax: +44 (0)20 8943 2767 Web: www.lgcgroup.com