

**STUDIES OF IMPRECISION, PERFORMANCE AND MEASUREMENT
UNCERTAINTY IN LABORATORY ASSESSMENT OF THROMBOTIC
THROMBOCYTOPENIC PURPURA**

by

Stephen MacDonald FRCPath MPhil

Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the
degree of Doctor of Philosophy.

School of Health Medicine and Life Sciences

University of Hertfordshire, Hatfield Hertfordshire, UK

AL10 9AB

Specialist Haemostasis Department, Cambridge University Hospitals NHS Foundations Trust,
Hills Road, Cambridge, CB2 0QQ

August 2025

Declaration

This thesis is the result of my own work and includes nothing that was the outcome of work done in collaboration other than that which is acknowledged below. Support was received from Werfen (Werfen UK, Warrington, UK) and Technoclone (Technoclone, Austria) for reagent provision for the initial verification studies of the ADAMTS13 activity rapid assays.

Some findings were reviewed in collaboration with colleagues within the CUH trust and colleagues nationally. All findings where discussions or data analysis was done in collaboration are presented in this thesis only as the work I personally conceived and performed. Additional work is restricted to that which has been published and author contributions acknowledged.

Collaborations

Throughout the course of this research, informal academic discussions were held with national and international experts in the field of ADAMTS13 testing. These discussions provided valuable context regarding current practices, emerging challenges, and developments within the field. However, no direct intellectual contribution, practical involvement, or collaborative input influenced the design, execution, or analysis of the work presented in this thesis.

Table 1: List of collaborative activities.

Chapters	Contributors	Extent of contribution	Thesis Chapter
Clinical and analytical comparison of ADAMTS13 activity platforms	Technoclone	Reagent provision for method comparison studies	5
Clinical and analytical comparison of ADAMTS13 activity platforms	Werfen	Reagent provision for method comparison studies	5
Clinical and analytical comparison of ADAMTS13 activity platforms	Dr Tina Dutt, Dr Jun Yong, Colin Downey (Royal Liverpool), Peter Baker (Oxford University Hospitals), Paul Murphy (Newcastle upon Tyne Hospitals), Rebekah Fretwell, Thomas Pitchford (Royal Hallamshire Hospitals, Sheffield)	Academic discussion	5

Acknowledgements

The work presented in this thesis would not have been possible without the guidance and support of my supervisory team. In particular Dr Enric Vilar has provided a huge amount of encouragement guidance and inspiration to enable this thesis to become complete. I am extremely grateful for all that he has contributed to this research and the knowledge he has provided during our weekly meetings has been of great benefit to not only for the completion of this thesis but also for my development in the research field. Prof. Ken Farrington has also helped me significantly with getting this over the line and his critical review of details within the study has benefitted me throughout.

My colleagues at CUH including clinical staff and laboratory staff have also been very supportive to getting the thesis completed. Dr Will Thomas is a constant source of valuable clinical information and sharing an office with him has benefitted by development. Drs Robinson, Symington and Besser have also been a constant source of valuable clinical support over the years. The laboratory staff have, of course, been a source of inspiration throughout, not just for listening to my constant “latest theories” but also in supporting the development of the service and progressing it to where it is now. Also, recognising the signs where I just needed some peace and quiet to get something done is a skill they have all developed over the years.

Finally, my family. My wife Kerry for constantly putting up with my need to get something done on my PhD, but also dragging me away from it when I needed it. My boy Alex, for his constant distraction about Minecraft, Roblox and Star Wars that provided a welcome break from ADAMTS13. And latterly in the thesis development Donnie, the crazy sproodle that while walking him in the fresh air added to some time to think clearly. Of course, my parents (Catherine and Evan) got me where I am today and without the determination and work ethic they instilled I would not have got this done.

Contents

Declaration	2
Collaborations	2
Acknowledgements.....	3
Contents	4
Abstract.....	10
List of Tables.....	11
List of Figures	13
List of Equations.....	16
Abbreviations	17
Glossary of Terms.....	20
Postgraduate certifications	22
1 Introduction.....	23
1.1 Clinical Vignette: Assessing ADAMTS13 Uncertainty.....	23
1.2 Thrombotic Microangiopathy (TMA)	24
1.3 Laboratory measurement of ADAMTS13.....	30
1.4 ADAMTS13 assay's role in clinical outcome performance specifications (Model 1 of the Milan criteria).....	46
1.5 Biological Variation of ADAMTS13 (Model 2)	50
1.6 State of the Art of ADAMTS13 assays (Milan Model 3)	57
1.7 Representing Clinical Performance in comparison studies.....	65

1.8	Measurement Uncertainty	72
2	Literature review	81
2.1	MU in the Haemostasis Laboratory	81
2.2	Abstract.....	81
2.3	Introduction	82
2.4	The Bottom-Up Approach to quantifying MU – the INR	83
2.5	Recent discussion of MU of INR	84
2.6	The Top-down method	88
2.7	The Top-down method in haemostasis assays	94
2.8	Using Monte Carlo Simulation for haemostasis MU.....	96
2.9	Discussion	97
2.10	Conclusions.....	99
	Rationale for the Thesis	101
	Limitations of Current Approaches to MU	101
	Gaps in APS in Haemostasis	101
	Why ADAMTS13 Activity is an Appropriate Exemplar	102
	Scope and Direction of the Thesis.....	102
	Statement of Originality and Contribution to Knowledge	102
	Thesis Structure	104
	Thesis Aims.....	106
3	Thesis Methods	108
3.1	Introduction to Methods	108

3.2	Laboratory Processes.....	109
3.3	Assay Platforms and Data Sources.....	111
3.4	Data Management and Governance.....	114
3.5	General Statistical Methods.....	116
3.6	Development of the ADAMTS13 Error Grid.....	124
3.7	APS Calculator for Simulating Clinical Risk and APS	120
3.8	Biological Variation Data Source	122
3.9	Biological Variation APS calculations	122
3.10	Reference Interval Determination Using the refineR Algorithm	125
3.11	Personalised Reference Intervals.....	127
3.12	Reference Change Value Calculations.....	130
3.13	Ethical and Regulatory Approvals	132
4	Chapter 4 - Clinical Thresholds and clinical Interpretation	134
4.1	Introduction.....	134
4.2	Aims	135
4.3	Defining Allowable Uncertainty at Clinical Decision Limits	136
4.4	Biological Variation and Individual Interpretation.....	147
4.5	APS for Treatment Response	168
4.6	Chapter 4 Discussion	171
5	Chapter 5 – Method Comparison of ADAMTS13 Assays: Linking Verification to Clinical Risk ...	175
5.1	Introduction.....	175
5.2	Aims	177

5.3	Limitations of Standard Verification Methods.....	178
5.4	Optimising clinical cut off from local verification data	188
5.5	ADAMTS13 Error Grid: Visualising Misclassification Risk	197
5.6	Chapter 5 Discussion	210
6	Chapter 6 – Measurement Uncertainty in ADAMTS13: A Practical Framework	214
6.1	Introduction.....	214
6.2	Aims	215
6.3	ANOVA-Based Measurement Uncertainty Estimation for ADAMTS13 Assays	217
6.4	Benchmarking Combined Uncertainty Using Standard Top-Down Method.....	222
6.5	Chapter 6 Discussion	227
7	Chapter 7 - Decomposing Internal Quality Control Data for MU	229
7.1	Introduction.....	229
7.2	Aims	231
7.3	Randomness of IQC data used for u_{RW} determination	232
7.4	Autocorrelation of IQC data used for u_{RW} determination.....	238
7.5	Decomposition of IQC data used for u_{RW} determination.....	248
7.6	Chapter 7 Discussion	260
8	Chapter 8 – Calibration as a Source of Measurement Uncertainty in ADAMTS13 Assays	263
8.1	Introduction.....	263
8.2	Aims	265
8.3	The Impact of Calibration Model Selection on Interpolation Uncertainty	265
8.4	Calibration Model Choice as a Structural Contributor to Run-to-Run Assay Variability....	278

8.5	Quantifying Calibration-Derived Measurement Uncertainty	284
8.6	Calibration Standardisation and Hidden Uncertainty in Automated Systems.....	304
8.7	Chapter 8 Discussion	315
9	Chapter 9 – Interpreting ADAMTS13 Results: A Clinically Risk-Based Framework.....	317
9.1	Introduction	317
9.2	Aims	318
9.3	Quantifying Diagnostic Evidence from Uncertainty at the Decision Threshold	319
9.4	Bayesian Updating of ADAMTS13 Results Using PLASMIC-Derived Risk	321
9.5	Interpretive Robustness: Quantifying the Clinical Consequences of MU	326
9.6	Chapter 9 Discussion	331
10	Thesis Discussion	334
10.1	Introduction	334
10.2	Summary of findings.....	336
10.3	Integration of Findings and Contribution to Knowledge	348
10.4	Recommendations for Future Research and Practice	351
11	References	356
12	Appendices	372
12.1	Appendix A - Comparison of measurand definitions of ADAMTS13 activity assays.....	372
12.2	Appendix B – Clinical outcome simulations for MAU	374
12.3	Appendix C – Reference Change Value proportions	376
12.4	Appendix D – ANOVA variance component analysis	377
12.5	Appendix E - ANOVA variance component analysis for Preliminary MU.....	380

12.6	Appendix F – Top-Down uncertainty of the Technozym [®] ELISA	381
12.7	Appendix G – Pooled SD Method for combined uncertainty of HemosIL AcuStar [®]	383
12.8	Appendix H - Top-Down uncertainty of the HemosIL AcuStar [®]	384
12.9	Appendix I – Published method comparison studies	386
12.10	Appendix J – Stability raw data	389

Abstract

ADAMTS13 activity testing is essential for diagnosing and managing thrombotic thrombocytopenic purpura (TTP). However, current methods for calibration, measurement uncertainty (MU), verification, and result interpretation are poorly suited to the complexity of ADAMTS13 assays. These tests rely on non-linear calibration models, show high variability between platforms, and lack well-defined analytical performance limits linked to clinical decision points.

This research challenges the adequacy of existing standards, including ISO 15189, ISO/TS 20914, and CLSI guidelines, when applied to ADAMTS13. It develops new statistical approaches that better describe assay behaviour and support clinical interpretation.

The project also establishes a framework for defining analytical performance specifications (APS) directly linked to clinical risk. This includes ADAMTS13-specific error grids that incorporate uncertainty through guard bands, improving the reliability of results near key clinical thresholds at 10 IU/dL, 20 IU/dL, and 50 IU/dL. Likelihood ratios are calculated across the reportable range, providing clinicians with tools to quantify diagnostic confidence and track disease progression.

Verification is restructured using risk-based methods that move beyond protocol-driven processes. This approach connects analytical performance directly to clinical outcomes. The framework developed delivers a more accurate understanding of the limitations and capabilities of ADAMTS13 testing. While focused on ADAMTS13, the methods are applicable to any complex diagnostic assay where current standards fail to link analytical quality with clinical decision-making.

List of Tables

Table	Page
Table 1: List of collaborative activities.....	2
Table 2: List of Abbreviations used in the thesis.....	17
Table 3: Definitions of terms used throughout the thesis	20
Table 4: Extra curricular post graduate certifications.	22
Table 5: PLASMIC Score Criteria and Risk Categories.	25
Table 6: Bentley Score Criteria and Risk Interpretation.....	26
Table 7: French Score Criteria and Performance	26
Table 8: Summary Comparison of Pre-Test Probability Scores for TTP	27
Table 9: ADAMTS13 activity methods.	33
Table 10: HemosIL AcuStar® comparison performance.	35
Table 11: ADAMTS13 activity assays.	39
Table 12: Model 1 and 2 of the Milan criteria – ADAMTS13 activity assays.....	50
Table 13: Published performance specifications for ADAMTS13 activity assays - Technozym ® ELISA.....	51
Table 14: Biological variation data from the EFLM database.	55
Table 15: Model 2 analytical performance specifications for ADAMTS13 activity.....	55
Table 16: Model 2 Biological Variation Models.	56
Table 17:ADAMTS13 activity assay performance characteristics.	60
Table 18: Comparison of Common Statistical Approaches Used in Method Comparison Studies	63
Table 19: Error Grid zones.	66
Table 20: Advantages and Disadvantages of MU Estimation Techniques.....	75
Table 21: MU considerations in published studies of MU in the INR.	87
Table 22: Comparative Overview of MU Approaches in Haemostasis Assays.	95
Table 23: Decision limits applied to the APS Calculator.	137
Table 24: Performance Specifications for Technozym® ELISA and HemosIL AcuStar®.....	138
Table 25: ADAMTS13 analytical performance specifications by Model 1b.	145
Table 26: ADAMTS13 reference intervals from population sample data.....	158
Table 27: RefineR-Derived Coefficients of Variation.....	159
Table 28: Longitudinal data summary for monitored patients.	161

Table 29: Personalised reference intervals (prRI) for ADAMTS13 activity.	163
Table 30: High probability relapse reference change values.	166
Table 31: The impact of MU on correlation statistics.	184
Table 32: Likelihood ratio of ADAMTS13 activity across varying decision limits.	193
Table 33: Definition of clinical zones for comparison when deriving the error grid.....	198
Table 34: Standard uncertainty for three lot numbers from variance components analysis.	221
Table 35: Assays and control material used as the focus of this study.	233
Table 36: IQC data characteristics used for the study of autocorrelation and partial correlation.....	238
Table 37: Random variation contribution for Technozym ® ELISA Quality Control (QC).....	256
Table 38: Random variability in internal quality control (IQC) measurements.	258
Table 39: True random variability in APTT and ADAMTS13 by two methods.....	259
Table 40: Raw data (absorbance at 450nm) for calibrator values in IU/dL.....	288
Table 41: Absorbance values across six calibrator levels over 12 independent ELISA runs.....	290
Table 42: ANOVA summary statistics for aggregated Technozym ® ELISA.	296
Table 43: Unknown samples results with measurement uncertainty.	302
Table 44: Calibration fit statistics across five independent calibrations.	306
Table 45: Calibration four-parameter logistic fitted parameters.	308
Table 46: Aggregated calibration fit parameter across five independent runs.	309
Table 47: Performance thresholds and associated limits used for interpretive robustness chart.....	328
Table 48: Measurand definition comparison for ADAMTS13 activity quantification.....	372
Table 49: RCV of ADAMTS13 activity based on four different models.	376
Table 50: Variance component analysis of Technozym ® internal quality control.....	377
Table 51: Variance components analysis ANOVA output for Technozym ® ELISA High QC.....	378
Table 52: Variance components analysis ANOVA output for Technozym ® ELISA Low QC	378
Table 53: Imprecision and ANOVA output data for high and low ADAMTS13 QC - AcuStar.	380
Table 54: Calculation of MU of the Technozym ELISA method.....	381
Table 55: The pooled standard deviation (SD_{pooled}) across ADAMTS13 quality control material.....	383
Table 56: Calculation of MU for the HemoSIL AcuStar ® method	384
Table 57: Summary of published method comparison studies for AcuStar.....	386
Table 58: Stability of ADAMTS13 activity.....	389

List of Figures

Figure	Page
Figure 1: Pathophysiology of TTP (2005).....	28
Figure 2: Domain organisation of ADAMTS13.....	30
Figure 3: Domain structure with Open (A) and Closed (B) conformation of ADAMTS13.	32
Figure 4: The chemiluminescent technology employed on the AcuStar.	34
Figure 5: ELISA methodology using the Technozym ® ADAMTS13 Activity (Technoclone, Austria).....	36
Figure 6: ADAMTS13 activity by Fluorescence Resonance Emission Transfer.	37
Figure 7: Monte Carlo simulation of increasing uncertainty to patient results.	49
Figure 8: The process of assigning analytical quality specifications.	58
Figure 9: The Clarke Error Grid.	67
Figure 10: The Ishikawa diagram.....	77
Figure 11: Apportioning of the MU budget.....	79
Figure 12: Linking performance specifications to measurement uncertainty.	89
Figure 13: Measurement uncertainty and the calibration hierarchy.....	91
Figure 14: Simulating MU in patient results.	121
Figure 15: Analytical Performance Specifications for overall agreement.....	140
Figure 16: Analytical Performance Specifications for agreement at 10 IU/dL.....	141
Figure 17: Analytical Performance Specifications for agreement at 10-20 IU/dL.....	142
Figure 18: Analytical Performance Specifications for agreement at 21-60 IU/dL.....	143
Figure 19: Analytical Performance Specifications for agreement at >60 IU/dL.	144
Figure 20: Population specific reference intervals from local data – Technozym ® ELISA	152
Figure 21: Population specific reference intervals from local data – AcuStar ®	153
Figure 22: Reference intervals from local data – Technozym ® ELISA, first presentation.	155
Figure 23: Population specific reference intervals from local data – AcuStar ®, first presentation	156
Figure 24: Proportion of samples giving positive RCV change.	165
Figure 25: Analytical Performance Specifications for ADAMTS13 replacement therapy.....	170
Figure 26: Stratified correlation of ADAMTS13 method comparison.	180
Figure 27: Pearson’s correlation in ADAMTS13 activity method comparisons.....	182
Figure 28: Correlation in ADAMTS13 method comparisons with measurement uncertainty.	183

Figure 29: Regression in ADAMTS13 method comparisons with measurement uncertainty.	186
Figure 30: Cut-off determined positive predictive value.	191
Figure 31: Cut-off determined specificity.	192
Figure 32: Cut-off determined false positive rate.	192
Figure 33: Cut-off determined positive likelihood ratio.	194
Figure 34: ADAMTS13 method comparison results.	200
Figure 35: ADAMTS13 activity Error Grid.	201
Figure 36: ADAMTS13 activity Error Grid with guard banding.	203
Figure 37: Local ADAMTS13 method comparison error grids.	206
Figure 38: ADAMTS13 method comparison error grids applied to published studies.	209
Figure 39: APTT IQC performance.	234
Figure 40: ADAMTS13 IQC run plots (Technozym ®).	235
Figure 41: ADAMTS13 IQC run plots (AcuStar ®).	236
Figure 42: Autocorrelation in APTT IQC.	241
Figure 43: Autocorrelation in ADAMTS13 Low IQC (Technozym ®).	242
Figure 44: Autocorrelation in ADAMTS13 Normal IQC (Technozym ®).	243
Figure 45: Autocorrelation in ADAMTS13 IQC (Technozym ®) single lot number.	244
Figure 46: Autocorrelation in ADAMTS13 Low IQC (AcuStar ®).	245
Figure 47: Autocorrelation in ADAMTS13 Normal IQC (AcuStar ®).	246
Figure 48: Autocorrelation in ADAMTS13 IQC (AcuStar ®) single lot number.	247
Figure 49: APTT High Abnormal Internal Quality Control data.	252
Figure 50: Decomposed APTT High Abnormal Internal Quality Control data.	253
Figure 51: Decomposed Technozym ®ELISA Low (Top) and Normal (Bottom) Quality control (QC).	255
Figure 52: Decomposed Internal Quality Control - HemoSIL AcuStar ADAMTS13 activity assay ®	257
Figure 53: Linear regression calibration fit - Technozym ® ADAMTS13	269
Figure 54: Linear calibration residual plot - Technozym ® ADAMTS13	270
Figure 55: Interpolation of linear calibration - Technozym ® ADAMTS13.	271
Figure 56: Aggregated absorbance responses - Technozym ® ADAMTS13.	273
Figure 57: Quadratic polynomial calibration - Technozym ® ADAMTS13	275
Figure 58: Cubic polynomial calibration - Technozym ® ADAMTS13	277
Figure 59: Relative Bias across 6 calibration modelling methods.	280
Figure 60: Variability of the absolute bias across the measurement range.	282

Figure 61: Aggregated Technozym [®] ELISA runs (n=12).....	292
Figure 62: Absorbances across aggregated Technozym [®] ELISA runs.....	293
Figure 63: Residual and %Bias from aggregated Technozym [®] runs.....	294
Figure 64: Aggregated reproducibility Technozym [®]	298
Figure 65: Technozym [®] aggregated unknown sample concentrations.....	299
Figure 66: Performance specifications for aggregated Technozym [®] ELISA runs.....	300
Figure 67: Bias correction using the HemosIL AcuStar [®] master curve	311
Figure 68: Comparison of Master and Calibrated RLU values at 10% ADAMTS13 activity.	313
Figure 69: Lot-to-lot variability in RLU adjustment at 10% ADAMTS13 activity.	314
Figure 70: Measurement Uncertainty impact on positive likelihood ratio.....	320
Figure 71: Post test probabilities using the PLASMIC score and simulated test performance.	324
Figure 72: Interpretation robustness of low, intermediate and high pre test probability as a function of analytical performance	330
Figure 73: Simulated clinical outcome MAU AcuStar [®]	374
Figure 74: Simulated clinical outcome MAU Technozym [®] ELISA.....	375

List of Equations

Equation	Page
Equation 1: Uncertainty of the International Normalised Ratio.	85
Equation 2: Calculation of MU according to guidance provided in ISO/TS 20914:2019.....	90
Equation 3: The Fraser model for analytical variation.	122
Equation 4: The Gowans model for analytical variation.....	122
Equation 5: The updated model of Fraser.	123
Equation 6: Maximum Allowable Measurement Uncertainty (MAU) from Biological Variation	123
Equation 7: Total Allowable Error (TA _E) calculated from Biological Variation	123
Equation 8: Derivation of the mean of the patient results.....	128
Equation 9: Total variation of the result.....	128
Equation 10: Variability of the homeostatic set point.....	129
Equation 11: Variability of the set point, including the set point.....	129
Equation 12: Personalised reference interval	129
Equation 13: reference Range Index (RRI).....	129
Equation 14: The common RCV method.....	130
Equation 15: RCV Increase according to the Fokkema model.....	131
Equation 16: RCV decrease according to the Fokkema model	131
Equation 17: Maximum allowable uncertainty for drug therapies.	169
Equation 18: Within run standard deviation.....	217
Equation 19: The between run standard deviation.....	217
Equation 20: The combined uncertainty.....	218
Equation 21: Interpolation Error in OLS when Estimating x from y.....	267
Equation 22: Determination of % concentration error.....	286
Equation 23: Calculating Pretest odds from Pre Test Probability (provided by the PLASMIC score)	321
Equation 24: Converting Pre Test odds to Post test Odds using the Likelihood ratio	321
Equation 25: Conversion of Post test odds back to Post test probability	322

Abbreviations

Table 2 lists the abbreviations used throughout this thesis. This is provided to support readability and ensure clarity, particularly for terms commonly used in the fields of haemostasis, ADAMTS13 testing, MU, and laboratory quality management.

Table 2: List of Abbreviations used in the thesis

Abbreviation	Definition	Abbreviation	Definition
ACD	Analytical Critical Difference	LOD	Limit of Detection
ACF	Autocorrelation Function	LPU	Law of Propagation of Uncertainty
ADAMTS13	a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13	LR ⁺	Positive Likelihood Ratio
ADF	Augmented Dickey-Fuller	LR-	Negative Likelihood Ratio
ANA	Antinuclear Antibody	MAHA	Microangiopathic Haemolytic Anaemia
ANOVA	Analysis of Variance	MAU	Maximum Allowable Uncertainty
APS	Analytical Performance Specification	MCS	Monte Carlo Simulation
APTT	Activated Partial Thromboplastin Time	MCV	Mean Corpuscular Volume
AUC	Area Under the Curve	MHRA	Medicines and Healthcare products Regulatory Agency
AUROC	Area Under the Receiver Operating Characteristic Curve	MNPT	Mean Normal Prothrombin Time
AVWS	Acquired von Willebrand Syndrome	MS	Mean Square
BA	Bland-Altman	MSbtw	Mean Square Between Groups
BCCL	Beijing Centre for Clinical Laboratories	MSwth	Mean Square Within Groups
BIPM	Bureau International des Poids et Mesures	MU	Measurement Uncertainty
BIVAC	The Biological Variation Critical Appraisal Checklist	NA	Not Applicable
BV	Biological Variation	NDC	Number of Distinct Categories
CAAPS	Clinically Acceptable APS	NEQAS	National External Quality Assessment Service
CCC	Concordance Correlation Coefficient	NHS	National Health Service
CDL	Clinical Decision Limit	NIBSC	National Institute for Biological Standards and Control
CEG	Clarke Error Grid	NPV	Negative Predictive Value
CELME	Cutting Edge of Laboratory Medicine in Europe	OIML	International Organization of Legal Metrology

Abbreviation	Definition	Abbreviation	Definition
CI	Confidence Interval	OLS	Ordinary Least Squares
CIA	Chemiluminescent Immunoassay	PACF	Partial Autocorrelation Function
CLSI	Clinical and Laboratory Standards Institute	PD	Percent Difference
CRM	Certified Reference Material	PEX	Plasma Exchange
CRP	C-Reactive Protein	PL	Phospholipid
CSA	Chromogenic Substrate Assay	PPP	Platelet Poor Plasma
cTP	Congenital Thrombotic Thrombocytopaenic Purpura	PPV	Positive Predictive Value
CUB	ADAMTS13 CUB domain	PT	Prothrombin Time
CUH	Cambridge University Hospitals	Pb	Passing Bablok
CV	Coefficient of Variation	PrRI	Personalised Reference Interval
CV _g	Within group (inter) biological variation	QFA	Quantitative Factor Assay
CV _i	Intraindividual Coefficient of Variation	RCV	Reference Change Value
CV _a	Analytical Coefficient of Variation	RLU	Relative Light Unit
DF	Degrees of Freedom	ROC	Receiver Operating Characteristic
DG-EIA	Enzyme Immunoassay	RRI	Reference Range Interval
DOR	Diagnostic Odds Ratio	RSSE	Residual Sum of Squares Error
ECAT	External quality Control of diagnostic Assays and Tests with a focus on Thrombosis and Haemostasis	RWD	Real World Data
EFLM	European Federation of Clinical Chemistry and Laboratory Medicine	SD	Standard Deviation
EGA	Error Grid Analysis	SDI	Standard Deviation Index
EHL	Extended Half-Life	SDP	Standard Deviation of the Population
ELISA	Enzyme-Linked Immunosorbent Assay	SD _a	Analytical Standard Deviation
EMS	Environmental Monitoring System	SEM	Standard Error of the Mean
FPR	False Positive Rate	SI	System of Internation Units
FRET	Fluorescence Resonance Energy Transfer	SPC	Summary of Product Characteristics
GCP	Good Clinical Practice	SPR	Surface Plasmon Resonance
GDPR	General Data Protection Regulation	SS	Sum of Squares
GST	Glutathione-S-Transferase	STL	Seasonal Trend and Decomposition using LOESS
GUM	Guide to the Expression of Uncertainty in Measurement	T-SE	Total Standard Error
HIL	Haemolysis, Icterus, Lipemia	TCV	Total Coefficient of Variation
HRP	Horseradish Peroxidase	TE	Total Error
HSP	Homeostatic Set Point	TEa	Total Allowable Error
ICSH	International Council for Standardization in Haematology	TMA	Thrombotic Microangiopathy
IEC	International Electrotechnical Commission	TPR	True Positive Rate

Abbreviation	Definition	Abbreviation	Definition
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine	TT	Thrombin Time
IFU	Instructions for Use	TTP	Thrombotic Thrombocytopenic Purpura
ILAC	International Laboratory Accreditation Cooperation	TVset	Total Variation of the Setpoint
INR	International Normalised Ratio	UK	United Kingdom
IQC	Internal Quality Control	UKAS	United Kingdom Accreditation Service
IQR	Interquartile Range	ULN	Upper Limit of Normal
IS	International Standard	ULQ	Upper Limit of Quantification
ISI	International Sensitivity Index	VHSP	Variability of Homeostatic Set Point
ISO	International Organization for Standardization	VIM	International Vocabulary of Metrology
ISTH	International Society on Thrombosis and Haemostasis	VTV	Variability of the Total Variation
iTTP	Immune Thrombotic Thrombocytopenic Purpura	VWF	von Willebrand Factor
IU	International Unit	WG-PRE	Working Group on Preanalytical Phase of the EFLM
JCGM	Joint Committee for Guides in Metrology	WHO	World Health Organization
JCTLM	Joint Committee for Traceability in Laboratory Medicine	WR	Within Run
LDH	Lactate Dehydrogenase	bA	Balanced Accuracy
LIMS	Laboratory Information Management System	i.i.d.	Independent and Identically Distributed
LLQ	Lower Limit of Quantification	pT	Predictive Threshold
LOA	Limits of Agreement	uRW	Uncertainty of the reproducibility within laboratory

Glossary of Terms

Table 3 provides definitions for key terms, concepts, and abbreviations used throughout this thesis.

These definitions are included to ensure clarity and consistency, particularly for terms that may have specific interpretations within the context of ADAMTS13 testing, MU, and laboratory method

evaluation. Where applicable, definitions are aligned with international standards, including ISO and

CLSI guidelines, to reflect current best practices in medical laboratory science.

Table 3: Definitions of terms used throughout the thesis

Term	Definition	Source / Note
ADAMTS13	A metalloprotease that cleaves von Willebrand factor; deficiency or inhibition is central to the diagnosis of TTP(TTP).	Haematology standard
Analytical Performance Specifications (APS)	A defined requirement for analytical quality, typically expressed as allowable bias, imprecision, or total error.	ISO 15198:2004
Balanced Accuracy (bA)	The average of sensitivity and specificity, used when classes are imbalanced.	Project-defined
Bias	The difference between the expectation of test results and a reference value.	ISO 5725-1:1994
Calibration	The process of establishing a relationship between the values indicated by an instrument and known reference values.	ISO/IEC 17025:2017
Clinical Decision Limit	A threshold used to distinguish between clinical categories (e.g., normal vs. abnormal).	IFCC-IUPAC 2015
Commutability	The property of a reference material to behave like a patient sample when measured by different methods.	ISO 17511:2020
Combined Standard Uncertainty (uc)	The standard uncertainty resulting from the combination of multiple sources of uncertainty (e.g., ucal, ubias, uRw), usually combined by the root-sum-of-squares method.	ISO/TS 20914:2019
CUB Domain	A structural domain (Complement C1r/C1s, Uegf, Bmp1) within ADAMTS13 important in protein interactions.	Structural biology
Decision Rule	A rule stating how measurement results are interpreted, taking uncertainty into account.	ISO/IEC 17025:2017
Error Grid	A graphical tool used in method comparison studies to assess clinical risk based on the magnitude and clinical relevance of measurement differences.	CLSI EP27-A
Expanded Uncertainty (U)	An interval around the result within which the true value is expected to lie with a stated probability (usually 95%), calculated as $U = k \times uc$.	ISO/TS 20914:2019
Guard Band	A safety margin around a clinical or specification limit, adjusted for uncertainty, to reduce false accept/reject outcomes.	ISO/TS 20914:2019
Measurement Uncertainty (MU)	A non-negative parameter characterising the dispersion of values that could reasonably be attributed to the measurand.	ISO/IEC Guide 99:2007 (VIM)
Measurand	The quantity intended to be measured, including its specification (analyte, matrix, quantity type).	ISO/IEC Guide 99:2007

Term	Definition	Source / Note
Metrological Traceability	The ability to relate a measurement result to a reference through an unbroken chain of calibrations, each contributing to the MU.	ISO/IEC Guide 99:2007
Passing-Bablok Regression (Pb)	A non-parametric statistical method for method comparison, robust to outliers and assumptions about distribution.	CLSI EP09c
Precision	The closeness of agreement between independent results obtained under specified conditions.	ISO 5725-1:1994
preliminary Reference Interval (prRI)	A statistically derived interval for interpreting results using small or local datasets.	prRI framework
Reference Change Value (RCV)	The minimum difference between two consecutive results in the same individual that can be considered clinically significant, accounting for both biological and analytical variability.	Fraser, IFCC
Reference Interval (RI)	The range of values expected for a defined healthy population, typically encompassing 95% of that group.	CLSI C28-A3
Repeatability	Precision under identical operating conditions (e.g. same day, operator, reagent lot).	ISO 5725-1:1994
Reproducibility	Precision across varied conditions (different days, operators, instruments, labs).	ISO 5725-1:1994
Sigma Metric (σ)	A measure of analytical performance defined as $(TEa - bias) \div CV$; higher values indicate better quality.	Westgard
Standard Uncertainty	The uncertainty of a result expressed as a standard deviation.	ISO GUM (Guide to the Expression of Uncertainty in Measurement)
Trueness	The closeness of the mean value obtained from a large series of test results to a reference value.	ISO 5725-1:1994
TVset (Total Variation of Setpoint)	The variation of a test's target value over time due to both systematic and random effects.	Project-defined
ubias (Uncertainty of Bias Correction)	The uncertainty from correcting for systematic bias, incorporating the uncertainty of the reference value and the laboratory result.	ISO/TS 20914:2019
uRw (Within-Laboratory Reproducibility)	An uncertainty component derived from long-term precision (intermediate precision) data, incorporating variation due to different runs, days, or operators.	ISO/TS 20914:2019
ucal (Uncertainty of Calibrator Assigned Value)	The uncertainty in the value assigned to a calibrator, including contributions from the calibration hierarchy.	ISO/TS 20914:2019
Validation	The process of confirming that a method meets performance requirements for its intended clinical use.	ISO 15189:2022
Verification	The confirmation that a specific method performs as expected in a given laboratory, often involving local data collection.	ISO 15189:2022
VHSP (Variability of Homeostatic Set Point)	The biological variability of a homeostatic marker across individuals; relevant in determining allowable bias.	Project-defined
VTV (Variability of the Total Variation)	The total expected variability in results from both analytical and biological sources.	Project-defined
Within-Run Variation (WR)	The variability observed within a single run, used to estimate repeatability.	CLSI EP05-A3
Type A Evaluation of Uncertainty	Based on statistical analysis of repeated measurements (e.g., SD of IQC data).	ISO/IEC Guide 98-3
Type B Evaluation of Uncertainty	Based on other information, such as manufacturer specifications, literature, or expert judgment.	ISO/IEC Guide 98-3

Postgraduate certifications

Table 4 presents the postgraduate certifications and qualifications completed during the course of this thesis. These qualifications reflect formal training undertaken to support the development of advanced expertise in the relevant methodological, clinical, and analytical aspects of the research.

Table 4: Extra curricular post graduate certifications.

These certifications were undertaken during the timeframe of the PhD registration. PGCert – post graduate certification

Qualification	Awarding Institution	Delivery	Award	Date
PGCert (Research Methods for Health)	University of Aberdeen	Distance learning	Distinction	September 2019
Advanced Clinical Assessment Skills (Level 7)	Anglia Ruskin University	Distance Learning, practical assessment	Distinction	December 2020

1 Introduction

1.1 Clinical Vignette: Assessing ADAMTS13 Uncertainty

Accurate laboratory measurements are crucial for diagnosing and managing thrombotic microangiopathies (TMAs). However, uncertainty in these measurements can present significant clinical challenges. Such uncertainty arose in a 38-year-old female who presented with thrombocytopenia, red cell fragmentation, and neurological symptoms suggestive of TMA.

Her other laboratory results included an ADAMTS13 (*a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13*) activity of 12 IU/dL. Published guidelines define levels <10 IU/dL as highly indicative of TTP (1), (2, 3). The clinician asked, 'What is the likelihood that the ADAMTS13 activity is actually below 10 IU/dL?'. This question highlights a fundamental challenge in laboratory diagnostics: communicating the impact of MU on clinical decision-making. As simple, and very common occurrence this case underscores not only the critical role of ADAMTS13 assays in TTP diagnosis but also the broader importance of addressing MU in laboratory medicine to ensure accurate, reliable, and clinically actionable results. This discussion formed the foundation for the research questions explored in this thesis including managing MU in ADAMTS13 activity assays for diagnosing and treating TMAs. Specifically, the research seeks to answer the following questions:

How does MU impact clinical decision-making in TTP diagnosis?

What frameworks can optimise the calibration and standardisation of ADAMTS13 assays?

How can harmonisation of laboratory practices reduce diagnostic variability and improve patient outcomes?

By bridging clinical, analytical, and methodological perspectives, this thesis contributes to the growing body of work emphasising the importance of MU in laboratory medicine.

1.2 Thrombotic Microangiopathy (TMA)

Thrombotic Microangiopathy (TMA) is a clinicopathological diagnosis requiring integration of laboratory findings and clinical examination to determine its precise nature (4). Organ dysfunction or damage is indicated by elevated creatinine and troponin (2). Common haematological manifestations include thrombocytopenia, haemolysis—defined using various laboratory techniques—and absence of coagulopathy assessed using widely available tests such as prothrombin time (PT) or international normalised ratio (INR) (2); (3). Autoimmune dysfunction may be identified through tests like anti-nuclear antibody assays (5). The presence of ongoing haemolysis is determined by elevated reticulocyte count, reduced or absent haptoglobin and variably lactate dehydrogenase (LDH) elevation above the upper limit of normal (ULN) (6).

Combining these laboratory markers with clinical history, including factors such as the presence of cancer or prior solid organ or stem cell transplant, assists clinicians in determining whether microangiopathic haemolytic anaemia (MAHA) or a specific diagnosis, such as TTP is likely.

1.2.1 Pre-Test Probability Scoring Systems for TTP

Pre-test probability scores are vital in diagnosing TTP. These tools combine clinical and lab data to assess a patient's likelihood of severe ADAMTS13 deficiency, helping guide early treatment choices.

1.2.1.1 *Scoring Systems in Thrombotic Microangiopathy*

In haemostasis other clinical diagnostic scoring systems, such as the Wells score (7) and 4T score (8), offer structured frameworks for estimating disease probability. The PLASMIC, Bentley, and French scores follow similar principles, assigning weighted points to defined criteria and categorising patients by risk.

1.2.1.2 The PLASMIC Score

The PLASMIC score consists of seven readily available variables (Table 5) and stratifies patients into low, intermediate, or high risk for severe ADAMTS13 deficiency (defined in this case as <15 IU/dL). Its strengths lie in ease of use, applicability across clinical settings, and strong diagnostic performance.

Table 5: PLASMIC Score Criteria and Risk Categories.

MCV – Mean Corpuscular Volume, INR – International Normalised Ratio

Parameter	Criterion	Score
Platelet count	<30 × 10 ⁹ /L	1
Haemolysis (reticulocytes)	>2.5%	1
Haptoglobin	Undetectable	1
Indirect bilirubin	>2.0 mg/dL	1
MCV	<90 fL	1
INR	<1.5	1
Creatinine	<2.0 mg/dL	1
Active cancer or transplant	Present	0

1.2.1.2.1 PLASMIC Score Interpretation

Cumulative scores are classified as 0–4 (Low risk - 0%), 5–6 (Intermediate -~6%) and 7 (High risk - 72%) (9)

Validated in different settings, the PLASMIC score shows high sensitivity (92%) and specificity (79%) (10) and has also proven useful in paediatric and atypical cases (11), (12). Limitations include reduced specificity in elderly patients and those with comorbidities (3). Adjustments to criteria, such as MCV and INR thresholds, have further improved predictive value (13, 14).

1.2.1.3 The Bentley Score

The Bentley score, the first developed pre-test algorithm for TTP, uses a weighted scoring system with five variables (Table 6). It demonstrated excellent predictive performance (AUC = 0.98) for ADAMTS13 deficiency <10% in its derivation study (15).

Table 6: Bentley Score Criteria and Risk Interpretation

Parameter	Threshold	Score
Platelet count	$>35 \times 10^9/\text{L}$	-30
Creatinine	$>2.0 \text{ mg/dL}$	-11.5
D-Dimer	$>4.0 \mu\text{g/mL}$	-10
Reticulocytes	$>3\%$	+21
Indirect bilirubin	$>1.5 \text{ mg/dL}$	+20.5

1.2.1.3.1 Bentley Score Interpretation

Cumulative scores are classified as <10 - 0% risk, 20–30 - ~40% risk and >30 - 100% risk.

Despite its high diagnostic accuracy, the Bentley score's complexity and reliance on assays not validated for this clinical setting, like D-dimer, limit its widespread use.

1.2.1.4 The French Score

Developed from registry data in settings lacking ADAMTS13 testing, the French score uses only three parameters (Table 7) and requires just one to meet criteria to suggest high risk (5).

Table 7: French Score Criteria and Performance

Parameter	Criterion	Score
Platelet count	$<30 \times 10^9/\text{L}$	1
Creatinine	$\leq 200 \mu\text{mol/L}$	1
ANA	Detected	1

A score of ≥ 1 yields a sensitivity of 98.8% and a negative predictive value of 93.3%, though specificity is limited (48.1%) (5). Its utility lies in rapid exclusion of TTP, particularly where laboratory support is limited.

1.2.1.5 Comparative Summary

Each scoring system offers distinct advantages (Table 8). PLASMIC stands out for its balance of accuracy and simplicity, Bentley for predictive strength, and French for accessibility. No score is universally best; local resources often guide choice. Future improvements could refine thresholds

and incorporate markers like LDH/ULN ratio to enhance discrimination among TMA subtypes (16), (6).

Table 8: Summary Comparison of Pre-Test Probability Scores for TTP

Feature	PLASMIC	Bentley	French
Number of variables	7	5	3
Validation status	Multicentre, external	Derivation only	Registry cohort
Sensitivity	92%	Not reported	98.8%
Specificity	79%	AUC = 0.98 (high)	48.1%
Practicality	High (routine tests used)	Moderate (needs D-dimer)	High (minimal inputs)
Limiting factors	Reduced specificity in older adults	Complexity, assay standardisation (D-Dimer)	Low specificity

1.2.2 Thrombotic Thrombocytopaenic Purpura (TTP)

1.2.2.1 Evolution of Understanding TTP

The understanding of TTP has significantly advanced over the past century (a diagram from 2005 is shown in Figure 1). First described by Dr Eli Moschowitz in 1925 in a 16-year-old female presenting with petechiae, pallor, and coma, histological examination revealed thrombi composed primarily of platelets (17). At that time, TTP was fatal in most cases. Improved understanding of TTP pathophysiology and the introduction of targeted therapies have transformed TTP from a fatal disease into a treatable condition. A comprehensive review by Lämmle (2024) commemorates the 100th anniversary of TTP's discovery, highlighting these milestones (18).

1.2.2.2 Pathophysiology of TTP

TTP is characterised by a severe deficiency (typically < 10 IU/dL) of ADAMTS13, a metalloprotease responsible for cleaving von Willebrand factor (VWF) multimers. Both acquired (immune mediated)

and congenital forms are present with 95% of cases being acquired (3). Deficiency leads to the accumulation of ultra-large VWF multimers, which promote excessive platelet aggregation (Figure 1) and systemic microvascular thrombus formation (19). The ensuing MAHA and thrombocytopenia result in widespread endothelial injury and impaired organ perfusion. The clinical manifestations of TTP are driven by ischaemic injury to vital organs such as the brain, kidneys, and heart (20); (21). Common symptoms include fever, neurological abnormalities (e.g., confusion, seizures), and acute renal impairment (22).

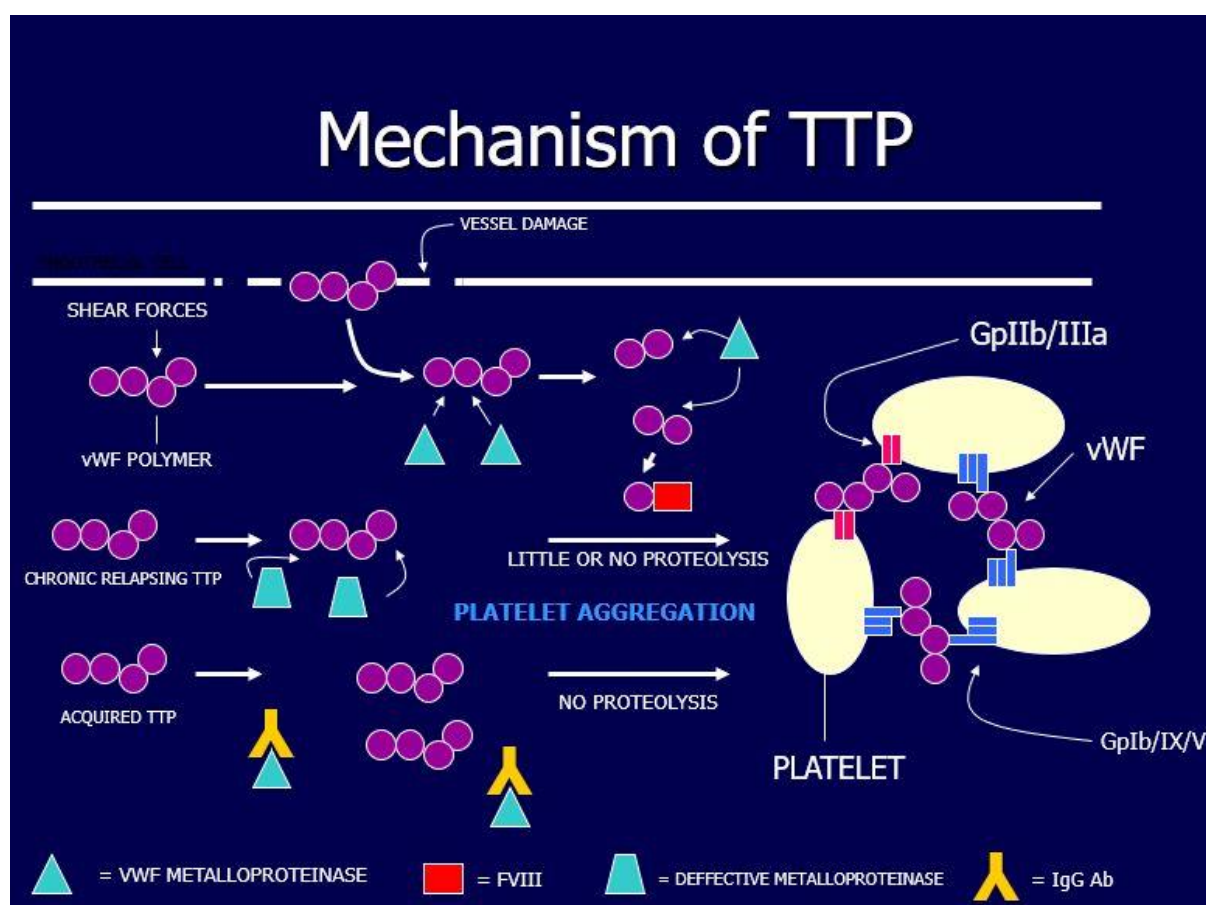


Figure 1: Pathophysiology of TTP (2005).

This diagram reflects the understanding of TTP in 2005, highlighting congenital ADAMTS13 deficiency (chronic relapsing TTP) and immune-mediated inhibition of ADAMTS13 (acquired TTP). Both mechanisms lead to the accumulation of ultra-large VWF multimers, platelet microthrombi, thrombocytopenia, and subsequent organ dysfunction. Note at this time ADAMTS13 wasn't formally recognised as the metalloproteinase responsible. Image created by S. MacDonald (2005).

Contributing mechanisms include direct platelet-VWF thrombi damage, oxidative stress, and activation of the Fas pathway, with emerging evidence suggesting a role for complement activation (21).

1.2.2.3 Autoantibodies in immune TTP

Autoantibodies in immune TTP (iTTP) neutralise ADAMTS13 activity or promote its clearance (23).

Most target the N-terminal domains, particularly the spacer domain (24). These antibodies are associated with enhanced disease severity, reduced ADAMTS13 levels, and poorer prognosis (25).

Understanding autoantibody profiles has informed development of therapies such as recombinant ADAMTS13, which shows promise in saturating autoantibodies and improving clearance (26).

1.2.2.4 Treatment of TTP

The treatment of TTP has undergone significant evolution, beginning with the introduction of plasma therapy. Early plasma transfusions substantially improved survival rates, marking the first major breakthrough in TTP management (18, 27); (28). Plasma exchange (PEX) later became the cornerstone of therapy, offering dual benefits of replenishing ADAMTS13 levels and removing inhibitory autoantibodies (29).

The discovery of ADAMTS13 deficiency as the underlying mechanism of TTP has refined diagnostic and treatment strategies, allowing differentiation between cTTP and iTTP forms (30); (31, 32). This distinction has enabled tailored approaches, including the use of immunosuppressants such as corticosteroids and rituximab for iTTP, alongside PEX. Due to the high mortality risk of untreated TTP, diagnosis is time critical. Current guidelines advocate initiating PEX and corticosteroid therapy without awaiting ADAMTS13 activity results, ensuring early intervention (33); (3).

The introduction of caplacizumab (Sobi (Swedish Orphan Biovitrum AB)), an anti-VWF nanobody raised in llamas, has further improved outcomes by accelerating platelet count normalisation and achieving faster clinical remission (34); (35). Guidelines from the International Society on Thrombosis

and Haemostasis (ISTH) recommend considering caplacizumab prior to the availability of ADAMTS13 results, supported by evidence from the Titan (36) and Hercules (37) studies.

For cTTP, advances such as recombinant ADAMTS13 and gene therapy are becoming viable treatment options, offering long-term solutions for previously unmet needs (38).

1.3 Laboratory measurement of ADAMTS13

1.3.1 Properties of ADAMTS13

ADAMTS13 plays a critical role in regulating VWF activity, particularly its platelet-activating functionality. ADAMTS13 is predominantly synthesised in hepatic stellate cells (39), with additional expression in tissues such as the placenta and skeletal muscle (40). mRNA for ADAMTS13 has also been identified in the kidneys, pancreas, spleen, and other organs, suggesting a broad physiological role (41).

1.3.2 Structure

The ADAMTS13 protein consists of 14 domains: a metalloprotease (M), disintegrin-like (D), eight thrombospondin type 1 repeats (T1–T8), cysteine-rich (C), spacer (S), and two CUB domains (CUB1, CUB2) (40). These domains are essential for its enzymatic activity and substrate specificity (Figure 2).

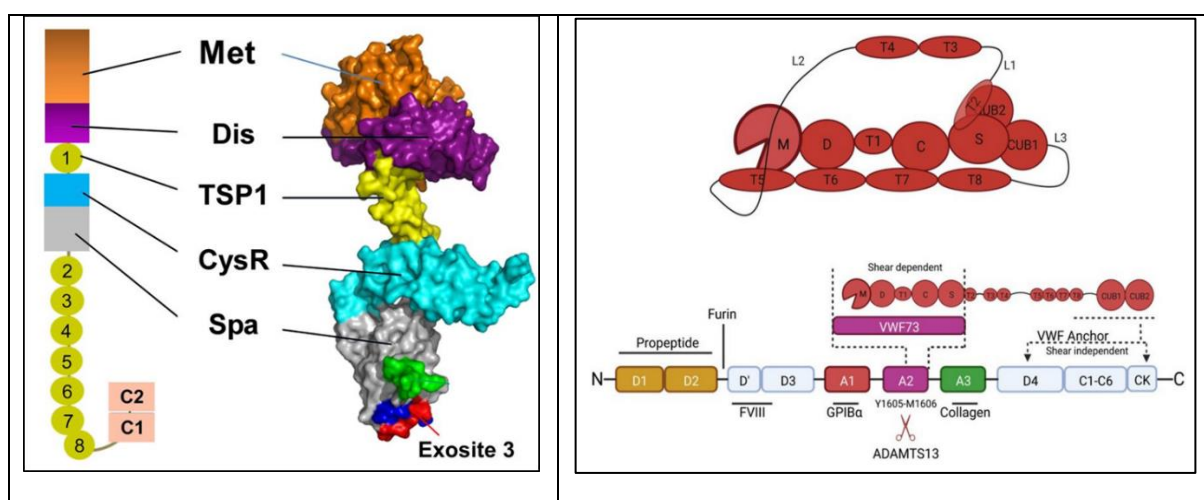


Figure 2: Domain organisation of ADAMTS13.

Domain organisation of the structure of ADAMTS13 (left) with partial crystal structure of ADAMTS13 as reported by (42).

Right – Top - the folded structure of ADAMTS13 showing the interactions of the spacer domain with CUB1 and CUB2 and the interaction (right -bottom) with vWF A2 domain as shown by (43).

The metalloprotease domain contains binding sites for calcium and zinc, with zinc coordinating the active site. Calcium is critical for catalytic activity, as demonstrated by reduced activity in the presence of the chelating agent EDTA (44); (45). Specific residues, such as Arg349 in the disintegrin-like domain, optimise ADAMTS13's cleavage of VWF by stabilising its scissile bond (46). The cysteine-rich domain and spacer domain contribute to substrate recognition, with the spacer domain primarily mediating binding affinity for VWF (Figure 2, right) (46). The CUB domains are unique to ADAMTS13 and are implicated in interactions with autoantibodies in iTTP (47); (48).

1.3.3 Interaction with von Willebrand Factor

ADAMTS13 cleaves VWF without requiring activation, relying instead on exposure of the VWF A2 domain cleavage site, which is induced by shear stress (43). Two shear-dependent mechanisms facilitate this. In circulation, shear forces reduce inter-monomer interactions within VWF, enabling ADAMTS13 access to the A2 domain (49). At the endothelial surface, VWF bound to collagen experiences high shear forces, resulting in elongation and exposure of the cleavage site (50). The A2 domain, stabilised by a disulphide bond (C1669/C1670), unfolds under shear stresses between 35–70 dyn/cm² (51), transiently revealing the Tyr1605-Met1606 bond for cleavage (49). This process predominantly occurs at vessel walls where shear is maximal, localising ADAMTS13 activity to endothelial surfaces. Once exposed, ADAMTS13 binds through its spacer, cysteine-rich, and disintegrin-like domains, stabilising the unfolded conformation and aligning the cleavage site for its metalloprotease domain (52).

1.3.4 Open and Closed Conformations of ADAMTS13

ADAMTS13 exists in open or closed conformations, determined by interactions between its spacer and CUB domains (Figure 3, below, (53); (54). The open conformation is associated with iTTP episodes, while the closed conformation predominates in non-immune TTP (55), (54, 56).

Studies of congenital TTP have shown that structural mutations, such as p.Arg1060Trp, affect ADAMTS13 conformation and activity. In a French cohort, 82% of adult-onset cTTP cases carried this mutation (55).

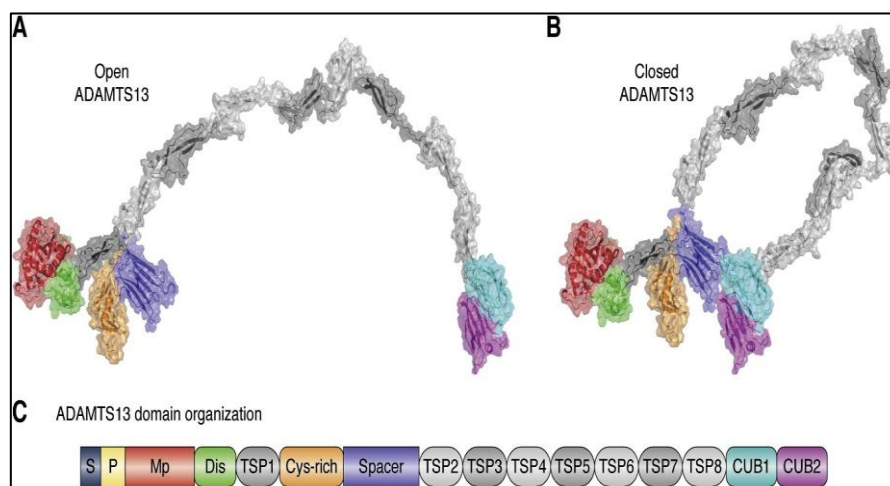


Figure 3: Domain structure with Open (A) and Closed (B) conformation of ADAMTS13.

Image from (55)

1.3.5 Regulation of ADAMTS13

Unlike many proteases, ADAMTS13 is secreted as an active enzyme with a half-life of 2–3 days. Regulation primarily occurs at the substrate level (42). Shear forces influence ADAMTS13 activity through biphasic kinetics, reflecting its adaptability to varying hydrodynamic conditions (57). Allosteric interactions between distal CUB domains and proximal domains further modulate its activity (58), with conformational changes triggered by pH and ligand binding (59). Inhibitors regulating ADAMTS13 activity remain a subject of active research, with none yet identified.

1.3.6 ADAMTS13 testing

Routine testing for ADAMTS13 activity remains limited to specialised centres due to the complexity and technical demands of available assays. These limitations often place ADAMTS13 testing beyond the capabilities of most routine hospital laboratories, despite its critical importance in diagnosing TTP.

ADAMTS13 activity is assessed by measuring its ability to cleave either the full-length von VWF or truncated peptide fragments of the VWF A2 domain (Table 9). First-generation assays utilised full-length VWF, while second-generation assays predominantly employ truncated peptides, such as vWF-73 (60). Other peptide fragments, including vWF-71 and vWF-86, are also used (61). These fragments mimic the natural cleavage site of ADAMTS13 without the complications of using full length VWF.

Table 9: ADAMTS13 activity methods.

Overview of various ADAMTS13 activity assay methodologies, including characteristics that have evolved over time improving performance and accessibility.

Assay types	Details
First-generation Assays	Utilised full-length VWF
Second-generation Assays	Predominantly employ truncated peptides, such as vWF-73
Other Peptides	vWF-71 and vWF-86
FRET Assays	Employ quencher and fluorophore flanking VWF A2 cleavage site, generating fluorescent signal proportional to ADAMTS13 activity
Chromogenic Endpoint Assays	Technozym® ELISA detects horseradish peroxidase (HRP)-linked chromogens bound to cleaved VWF-73

Assay types	Details
Chemiluminescent Assays	AcuStar® uses isoluminol-labelled monoclonal antibodies and magnetic particles to generate chemiluminescent signal within 33 minutes
Alternative Rapid Assays	Technofluor® (using the FRET principle) and Technoscreen® (Lateral flow immunoassay) provide rapid or semi-quantitative results

1.3.7 ADAMTS13 Activity by Chemiluminescence Immunoassay (CIA)

The HemosIL AcuStar® Chemiluminescent Immunoassay (CIA), developed by Werfen (Bedford, USA), is a fully automated, two-step immunoassay that advances traditional ELISA-based methods. This assay uses magnetic particles as the solid phase to isolate ADAMTS13 activity (Figure 4), providing results within ~33 minutes.

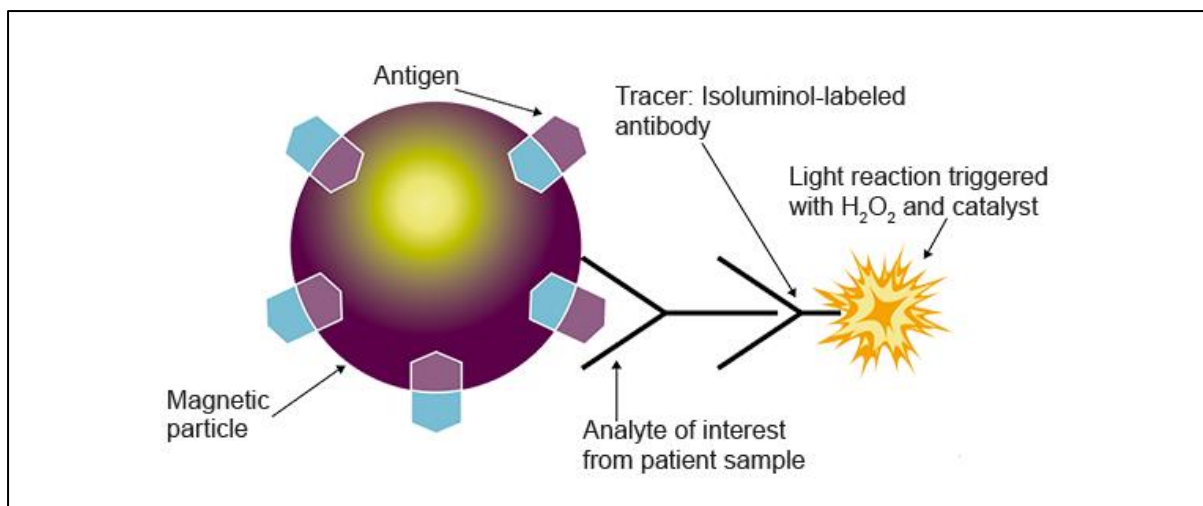


Figure 4: The chemiluminescent technology employed on the AcuStar.

The target antigen is bound specifically by the tracer. Upon addition of an oxidiser (H₂O₂) and a catalyst, light is emitted. The antigen coated magnetic particle serves as the dock for the measurand of interest and isolates the measurand during the washing steps. (Image taken from <https://www.werfen.com/uk/en/acl-acustar>)

The magnetic particles are coated with Glutathione-S-Transferase (GST)-vWF73, which mimics the ADAMTS13 cleavage site (Figure 4). In the first step, patient ADAMTS13 cleaves the substrate, enabling detection. The second step employs isoluminol and an anti-GST antibody to generate a chemiluminescent signal proportional to the cleavage activity.

Table 10: HemosIL AcuStar® comparison performance.

Method comparison results for HemosIL AcuStar® with ELISA and FRET methods, taken from the Werfen instructions for use (Werfen, Bedford, USA)

N	Slope	Intercept	Pearson's Correlation (r)	Comparative Method
105	0.975	2.617	0.943	ELISA
100	0.970	7.617	0.970	FRET

The HemosIL AcuStar® shows strong correlation with ELISA and FRET methods (

Table 10). Performance characteristics shown below show the reliance on correlation and regression for determining comparability of methods, an approach that is to be studied in depth in chapter 5.

1.3.8 ADAMTS13 by Technozym® ELISA

The Technozym® ELISA (Technoclone, Austria) quantifies ADAMTS13 activity by measuring the extent of vWF-73 substrate cleavage. The ELISA kit is often provided by local distributors such as DiaPharma (Surrey, UK) and Eurobioscientific (Dorking, UK), until recently. A colour change proportional to ADAMTS13 activity is generated through a horseradish peroxidase (HRP)-linked reaction (Figure 5).

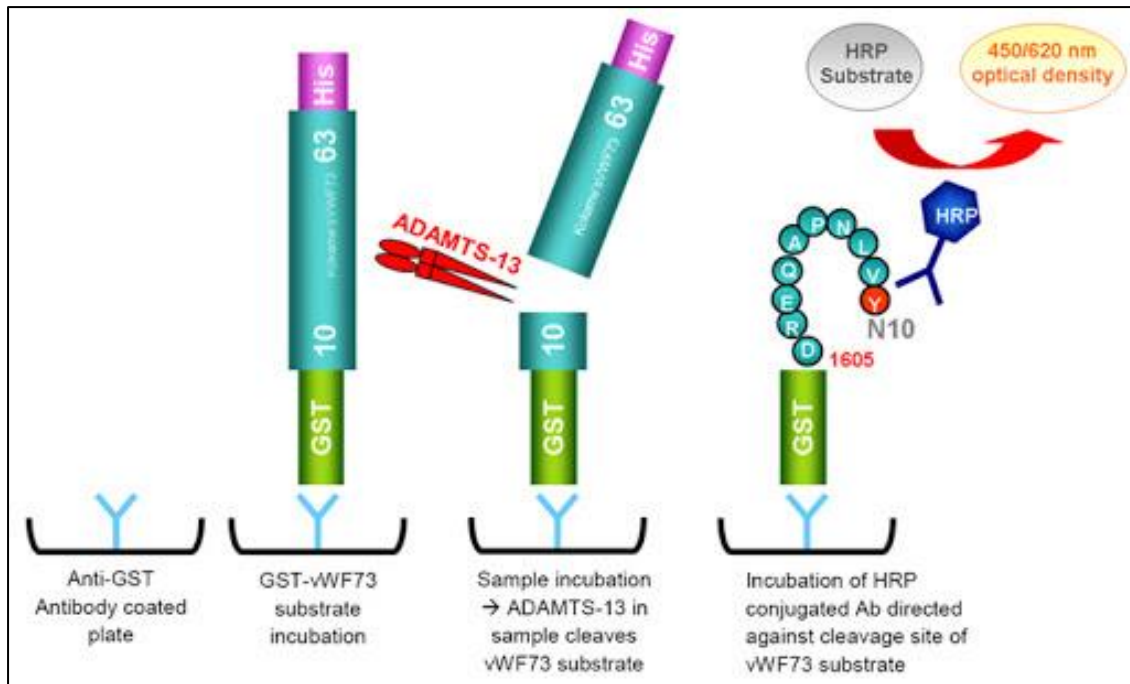


Figure 5: ELISA methodology using the Technozym[®] ADAMTS13 Activity (Technoclone, Austria).

GST-vWF73 is isolated on a microtitre plate, cleavage of the vWF73 substrate liberates a binding site (N10) for a HRP conjugate to bind and initiate a colour change proportional to ADAMTS13 activity level. Image from <https://diapharma.com/news-and-events/adamts13-activity/>

Assays are performed in duplicate, with calibration curves constructed using six diluted calibrators traceable to the first ADAMTS13 WHO international standard (12/252, MHRA (formerly NIBSC), Potters Bar, UK).

1.3.9 ADAMTS13 Activity by Fluorescence Resonance Energy Transfer (FRET)

Fluorescence Resonance Energy Transfer (FRET) is a powerful spectroscopic technique widely applied in ADAMTS13 testing to assess enzymatic activity and conformational states. In activity assays, synthetic peptide substrates such as FRETs-VWF73 are utilised. These peptides are labelled with a fluorophore and a quencher at opposite ends (Figure 6). In the intact substrate, the quencher suppresses the fluorescent signal. When cleaved by active ADAMTS13, the fluorophore is released from the proximity of the quencher, allowing fluorescence to be emitted. This enables real-time, quantitative monitoring of ADAMTS13 activity (60).

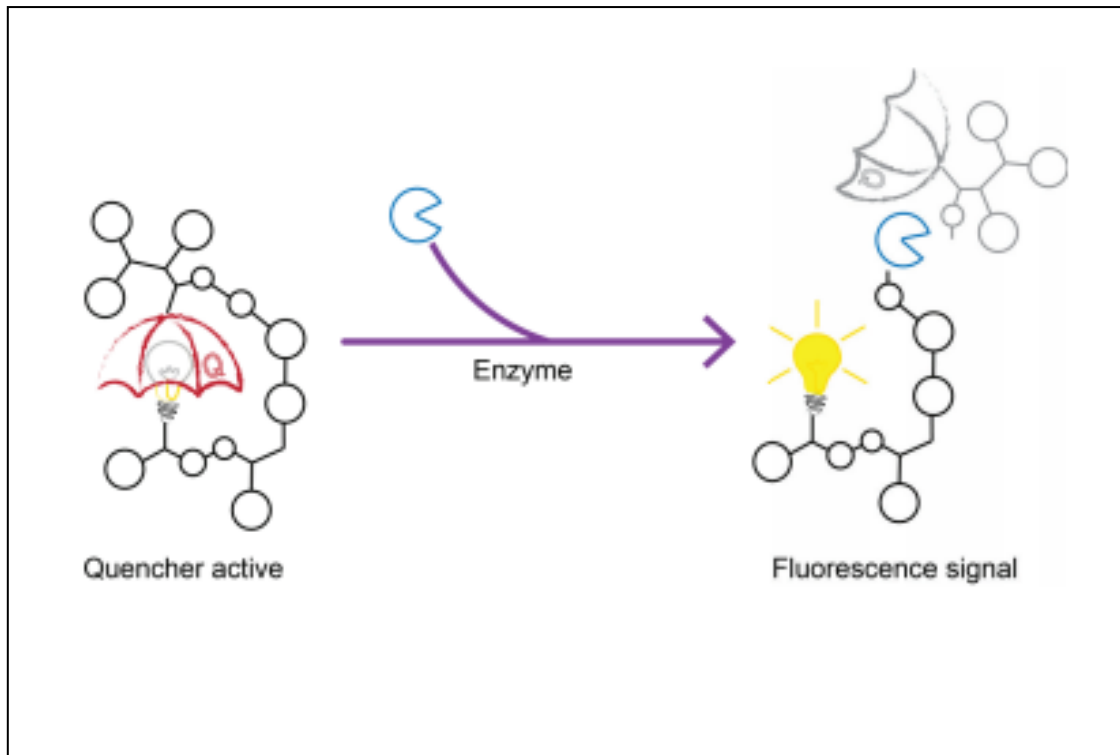


Figure 6: ADAMTS13 activity by Fluorescence Resonance Emission Transfer.

Presence of the quencher inhibits fluorescence signal from being generated (left side). Once the quencher is removed by enzymatic cleavage of the VWF A2 domain, the signal is able to be captured, proportional to the extent of cleavage, and therefore concentration of enzymatic activity. Image source: <https://bicconference.org/2019/posters/2019/07/31/p96-feasibility-of-a-new-fully-automated-adamts-13-activity-assay/>

FRET is also used in conformation-sensitive assays to distinguish between the open and closed forms of ADAMTS13. FRET-labelled monoclonal antibodies targeting distant sites on ADAMTS13 generate a signal in the closed conformation, which is lost when the protein opens. This assay helps differentiate between immune-mediated and cTTP by assessing conformational changes (62). A FRET-based method (Technofluor[®]), available on platforms such as the Ceveron Analyser[®], advances manual FRET methods and measures ADAMTS13 activity using fluorophore and quencher pairs flanking the vWF73 cleavage site.

1.3.10 Flow through single use assays

Lateral flow-based assays have been developed to provide rapid and accessible diagnostic tools, moving the ELISA based immunoassay principle into a faster time to result method. One such method, the Technoscreen[®] rapid assay, enables semi-quantitative detection of ADAMTS13 activity in a simplified format. Its key advantages are speed and ease of use, making it particularly well-suited to point-of-care settings and emergency departments where a quick presumptive diagnosis of TTP can guide urgent treatment decisions (63); (64).

1.3.11 Comparison of ADAMTS13 Activity Assays

All assays are traceable to the WHO 1st ADAMTS13 international standard (65), ensuring consistency in calibration and measurement (Table 11). While Technozym[®] ELISA, AcuStar[®] and Technofluor[®] offer high precision, rapid assays like Technoscreen[®] provide faster results, beneficial in time-critical settings with the trade off, of a semi-quantitative result. The choice of assay depends on the clinical context, resource availability, and turnaround time requirements.

Table 11: ADAMTS13 activity assays.

Comparison of commercially available ADAMTS13 activity assays. WHO – World Health Organisation, VWF – von Willebrand Factor, FRET – Fluorescent Resonance Energy transfer, ELISA – Enzyme Linked Immunosorbent Assay

Method	Manufacturer	Platform	Calibration	Substrate	Detection Method
Technozym® ELISA	Technoclone	Manual ELISA	6 levels traceable to WHO standard	vWF73	Chromogenic
HemosIL AcuStar	Werfen	AcuStar®	3-point calibration adjustment from master calibration, traceable to WHO standard	vWF73	Chemiluminescence
Technofluor®	Technoclone	Ceveron s100®	Serial dilution of reference plasma, traceable to WHO standard	vWF73	FRET
Technoscreen®	Technoclone	Manual (flow- based)	None, albeit with values traceable to WHO standard	vWF73	Visual observation

The biological complexity of ADAMTS13, including variable molecular forms and protein interactions, makes full SI traceability difficult (66); (67). In such cases, harmonisation of the measurand definition, even without reference methods, is the most effective route toward meaningful comparability (68). A detailed comparison of the principal differences between commercially available ADAMTS13 assays, as derived from CLSI EP29-A principles, is provided in Appendix A.

ELISA-based assays, such as Technozym®, rely on immobilised VWF73 and enzyme-linked colourimetric detection. These are robust but moderately sensitive and susceptible to matrix interference such as sample turbidity. Chemiluminescent assays, such as the AcuStar®, offer

enhanced sensitivity and speed, but depend on calibration systems that may not align with WHO standards. FRET-based assays, including FRETs-VWF73 and Technofluor®, use synthetic substrates in solution-phase formats, allowing high sensitivity and real-time detection (60); (61), though they remain vulnerable to interference from bilirubin and haemoglobin (69); (70).

Autoantibody interference and the presence of truncated or modified ADAMTS13 can also affect results differently depending on substrate structure (Mancini (71); (69). Additionally, differences in calibrant source, diluent composition, and dilution steps may lead to dissociation of immune complexes and falsely elevated activity (69).

Calibration traceability and reporting units further contribute to discordant results. While Technozym® and Ceveron® assays report in IU/dL traceable to WHO 12/252, the AcuStar® reports % activity, and the Technoscreen® provides only semi-quantitative results (72). These differences complicate direct comparison and may affect clinical decisions, especially when fixed cut-offs (e.g. 10 IU/dL) are used in TTP diagnosis or monitoring. Lateral flow-based assays like Technoscreen® offer benefits in accessibility and speed but sacrifice precision and traceability.

Clinical chemistry has made significant progress in harmonisation, (73). In haemostasis, such practices are less widely adopted. The comparison of assay performance across platforms—framed within harmonised definitions and supported by robust uncertainty estimation—remains a necessary and scientifically justified approach.

1.3.12 Calibration of ADAMTS13 Activity Assays to demonstrate Metrological traceability

In ADAMTS13 activity assays, calibration curves are used to interpolate patient results based on known calibrator concentrations. However, the selection of calibration models and parameters plays a critical role in determining both MU and bias. The first WHO standard for ADAMTS13 activity

(12/252) was created to promote harmonisation (65). Assigned a consensus value of 0.91 IU/vial via a 38-lab study, it aids in aligning assays across platforms. However, its lack of an associated uncertainty value restricts its utility for full metrological traceability (authors' opinion). With a geometric CV of 12.4% in the assignment process, standardisation remains limited by calibration variability and method differences. Although many of these assays are traceable to the WHO international standard 12/252, there remains considerable variation in calibration practices, which are often left to the discretion of the end user. This inconsistency can significantly affect the reliability of results, particularly near clinically relevant thresholds.

A wide range of calibration models are used across ADAMTS13 platforms. Polynomial regression, for example, is a common approach in which a single model fits the entire calibration curve. While this can be effective, the model's global nature means that any modification affects the whole curve, limiting its flexibility—particularly in regions where clinical accuracy is critical. In contrast, cubic spline functions provide greater adaptability (74). These models use localised basis functions, allowing specific regions of the curve to be adjusted independently. This is especially beneficial when targeting calibration around clinical decision points, as it enables reductions in uncertainty in those critical zones without compromising the entire curve's integrity.

Logistic curve fitting, including 4- and 5-parameter logistic (4PL and 5PL) models, is widely used in immunoassays due to the sigmoidal nature of the signal response. These models allow for accurate interpolation across a broad range of concentrations and offer improved fit for assays like ADAMTS13, where signals plateau at both high and low extremes (74). Nevertheless, even well-fitted curves can introduce bias if residual errors remain high. For instance, challenges have been noted with residual errors in the Technozym® ELISA and with the master curve fitting approach employed in the AcuStar® system, both of which may contribute to systematic error and variability across assay batches.

To optimise accuracy, calibration efforts should concentrate on minimising error around clinical decision points—particularly at the 10 IU/dL threshold that defines severe ADAMTS13 deficiency in TTP (1); (3). This involves ensuring a dense distribution of standards in this critical range, thereby reducing uncertainty in parameter estimates (75). Broader strategies include maximising the overall concentration range of standards, dynamically selecting calibration models based on observed signal behaviour, and validating calibration performance near decision limits (74).

Weighted uncertainty estimation in proximity to clinical cut-offs enhances diagnostic precision. By assigning greater importance to calibration accuracy near the 10 IU/dL threshold, laboratories can better manage the clinical risk associated with misclassification in acute TTP diagnosis or remission monitoring.

1.3.13 Uncertainty of End-User Calibrator assigned values

Although ADAMTS13 activity calibrators are provided with assigned values and, in some cases, associated uncertainties, their use varies depending on the calibration strategy—whether for direct calibration or the construction of serial calibration curves. However, one of the key challenges lies in the lack of transparency and consistency in how information on prior calibration steps is communicated. In practice, in vitro diagnostic (IVD) manufacturers rarely provide comprehensive documentation detailing all preceding stages in the calibration hierarchy. As a result, estimates of calibration uncertainty (u_{cal}) are often limited to the imprecision of the methods used by each manufacturer to assign the calibrator value, without accounting for accumulated uncertainty throughout the traceability chain.

1.3.14 Calibration Hierarchies for ADAMTS13 Activity Assays

1.3.14.1 *Technozym® ADAMTS13 Activity ELISA*

The Technozym® ADAMTS13 activity ELISA assay uses direct calibration. These six calibrators are traceable to the 1st WHO International Standard (12/252) via an internal master lot maintained by Technoclone®, Austria.

Of note in this case is that the uncertainties for assigned calibrator values are reported as remarkably low, in the absence of detailed information of how they were derived.

0.78 IU/mL: 1.2%

0.42 IU/mL: 1.9%

0.20 IU/mL: 3%

0.10 IU/mL: 3%

The uncertainty of the blank calibrator (0.00 IU/mL) is not provided. For clinical decision-making, particularly near critical thresholds (e.g., 10%/10 IU/dL), the uncertainty of the 0.10 calibrator (3%) is used as a reference.

1.3.14.2 *HemosIL® ADAMTS13 Activity Assay (AcuStar®)*

The calibration hierarchy for the HemosIL® assay is similar to the Technozym® method. It employs a local measurement procedure (by the manufacturer) and standard master lot for calibrator value assignment. According to the manufacturer the uncertainty of the calibrator is 10%, with no other context available, despite repeated conversations between the author and the company (personal communication).

Information is limited to the fact that the uncertainty for calibrators is calculated using pooled standard deviations from replicate measurements. A coverage factor ($k=2$) is applied, providing 95%

confidence intervals. However, these values are influenced by reagent and calibrator lot variability, as acknowledged by the manufacturer (Werfen, personal communication).

The manufacturer's statement that uncertainty values are reported in the same units (i.e. %) as patient results raises conceptual issues. While results are reported in percentage activity, it is unclear whether the uncertainty is expressed as an absolute or relative value, creating potential ambiguities in interpretation. The question of why % is used when the manufacturer calibrator is traceable (apparently) to the 12/252 WHO international standard has yet not been satisfactorily answered.

1.3.15 ADAMTS13 Antigen Assays

ADAMTS13 antigen assays are designed to quantify the concentration of the ADAMTS13 protein in plasma, regardless of its enzymatic activity. Measuring antigen levels is not often employed in the differential diagnosis of TTP. However, it may distinguish between iTTP, where ADAMTS13 protein is typically present but inactivated by circulating autoantibodies, and cTTP, where the enzyme is either absent or markedly reduced due to genetic mutations affecting protein synthesis or secretion (76); (77).

Several commercial and in-house ELISA-based antigen assays are available, with varying degrees of analytical sensitivity and specificity. The performance of some of these assays has been shown to be affected by the presence of ADAMTS13–autoantibody complexes, which may mask epitopes and interfere with antibody binding in the assay (78).

1.3.16 Antibody Assays and Inhibitor Detection

Autoantibodies, primarily of the IgG class, inhibit ADAMTS13 activity by either steric hindrance at functional sites or by promoting accelerated clearance of the enzyme from the circulation (79). Two main approaches are used to assess the presence of these antibodies: inhibitor (functional) assays and immunoassays (binding assays).

Functional inhibitor assays, typically based on the Bethesda-like mixing study format, evaluate whether patient plasma can inhibit ADAMTS13 activity in normal pooled plasma. A significant reduction in activity following incubation indicates the presence of neutralising antibodies. While useful, this method is limited by its low sensitivity and can yield false negatives in patients with non-inhibitory (clearing) antibodies or when samples are collected after therapeutic plasma exchange (80).

Binding assays, on the other hand, directly detect ADAMTS13-specific antibodies using enzyme-linked immunosorbent assays (ELISA) or similar immunoassay formats. These tests do not require functional inhibition to be present, thereby detecting both inhibitory and non-inhibitory antibodies. Furthermore, recent developments in epitope mapping have identified specific ADAMTS13 domains targeted by pathogenic autoantibodies, such as the spacer, cysteine-rich, and TSP2-8 domains, allowing for refined diagnostic and prognostic insights (81); (28).

In clinical practice, combining activity testing, antigen quantification, and antibody assays provides a comprehensive diagnostic framework for TTP. This integrated approach not only facilitates accurate classification into iTTP or cTTP but also supports risk stratification for relapse and monitoring response to therapies such as rituximab or caplacizumab.

1.3.17 Emerging Technologies for ADAMTS13 activity measurement

A novel multimer ratio based on vWF structure has recently been proposed as a marker of disease activity in TTP (82). This approach involves quantifying changes in vWF multimer distribution, offering a potential improvement in disease monitoring. When used alongside conventional ADAMTS13 activity assays, this multimer ratio could enhance the assessment of TTP progression or remission, particularly in immune-mediated forms of the disease.

Surface Plasmon Resonance (SPR) is another emerging technology applied in the evaluation of ADAMTS13. SPR allows real-time, label-free monitoring of molecular interactions and is being

investigated for its utility in assessing both ADAMTS13 enzymatic activity and conformational state. Recent studies have demonstrated that SPR can distinguish between open and closed conformations of ADAMTS13, a feature that may help predict relapse in patients with iTTP (83, 84); (85). This conformation-sensitive analysis represents a promising advancement in both diagnostics and longitudinal disease monitoring.

1.4 ADAMTS13 assay's role in clinical outcome performance specifications (Model 1 of the Milan criteria)

While laboratory results play a pivotal role in guiding clinical decisions, patient outcomes depend on several intermediate steps, including treatment decisions and the timeliness of interventions (86). Thus, the role of analytical performance specifications (APS) must be evaluated within this broader clinical context.

A critical theme from the Cutting Edge of Laboratory Medicine in Europe (CELME) Prague meeting in 2023, an important international gathering of experts, was the alignment of APS with clinical outcome in mind. Different clinical applications for the same measurand may necessitate distinct APS, challenging the “one-size-fits-all” approach. This principle aligns with ISO 15189:2022, which mandates that quality processes mitigate patient risk by linking analytical performance to clinical utility.

However, disparities exist between disciplines. Clinical biochemistry has led the adoption of APS, while fields like haemostasis face significant challenges due to the lack of reference methods and materials.

1.4.1 Clinical Outcomes of TTP

Measuring clinical outcomes in TTP remains a clinical challenge, with outcomes varying widely depending on the subtype and timeliness of treatment. The impact of ADAMTS13 measurement on those outcomes is therefore, additionally complex. iTTP patients treated in intensive care units experience a mortality rate of approximately 42%, even with modern therapeutic approaches (87). In contrast, cTTP exhibits lower mortality rates, ranging from 0% to 13.4% across studies (88).

Cardiovascular complications, including stroke and myocardial infarction, are a leading cause of premature death in TTP patients, particularly those with iTTP (88, 89). Delays in initiating PEX, the cornerstone of TTP treatment, are associated with increased mortality and worse clinical outcomes (90).

Beyond its clinical impact, TTP places a considerable burden on healthcare systems. Patients often require prolonged hospitalisation, with associated higher healthcare costs and resource utilisation (91). Long-term management is critical, as patients in remission may experience neurocognitive deficits and cardiovascular complications, which can significantly affect quality of life (30);(20). Regular ADAMTS13 measurements play a pivotal role in monitoring disease progression and guiding follow-up care.

1.4.2 Clinical Outcome Studies for ADAMTS13 Activity

Despite the clinical significance of ADAMTS13 activity assays, robust data linking assay performance to patient outcomes are limited and definition of outcomes remains somewhat unstandardised (92). Cuker *et al.* noted that improvements in analytical performance have enabled better detection of relapse in immune TTP (93). However, the study does not detail the specific performance metrics underlying these claimed improvements. Simulations could address this gap, allowing laboratories to model the effects of MU on decision limits for initiating or discontinuing therapies like plasma exchange or caplacizumab.

For cTTP, long-term complications such as chronic kidney disease and neurological issues highlight the importance of optimised assay performance (94, 95). The introduction of recombinant ADAMTS13 (rADAMTS13) for cTTP treatment further underscores the need for APS tailored to long-term outcomes, a topic that remains an area for future research.

1.4.3 Applying Clinical Outcome Models (Milan Model 1)

The determination of performance requirements for analytical methods based on clinical outcomes is the cornerstone of Model 1 of the Milan criteria (96). Despite its conceptual simplicity, the implementation of clinical outcome models is limited by the scarcity of robust published data (86).

For haemostasis assays, the link between assay performance and clinical outcomes remains underexplored. Model 1a is suitable for measurands directly associated with disease diagnosis (97) and treatment, such as ADAMTS13 activity in TTP.

Interestingly, studies have shown that reduced analytical performance does not always worsen patient outcomes. For example, point-of-care INR testing, while exhibiting decreased analytical precision compared to central laboratory testing, has improved patient outcomes due to the increased frequency of testing and more effective anticoagulation management (98); (99). These findings suggest that APS must account for both the technical performance of a test and its clinical utility in practice.

To address the challenges of minimal data being available, Model 1 is divided into two subcategories. Model 1a is direct clinical outcome studies, which assess the performance of an analytical method through direct correlation with patient outcomes, such as that above with point of care INR. Model 1b is focused on indirect clinical outcome studies, using simulation-based approaches such as Monte Carlo methods or decision theory.

1.4.4 Simulation in Clinical Outcome Modelling

In cases where direct outcome data are unavailable, simulation-based methods offer a valuable alternative (Figure 7, (100)). Monte Carlo simulation enables laboratories to model the effects of varying degrees of MU on patient classification against clinical decision limits or reference intervals. This method evaluates the critical question: how much uncertainty is tolerable before an unacceptable proportion of patients are misclassified?

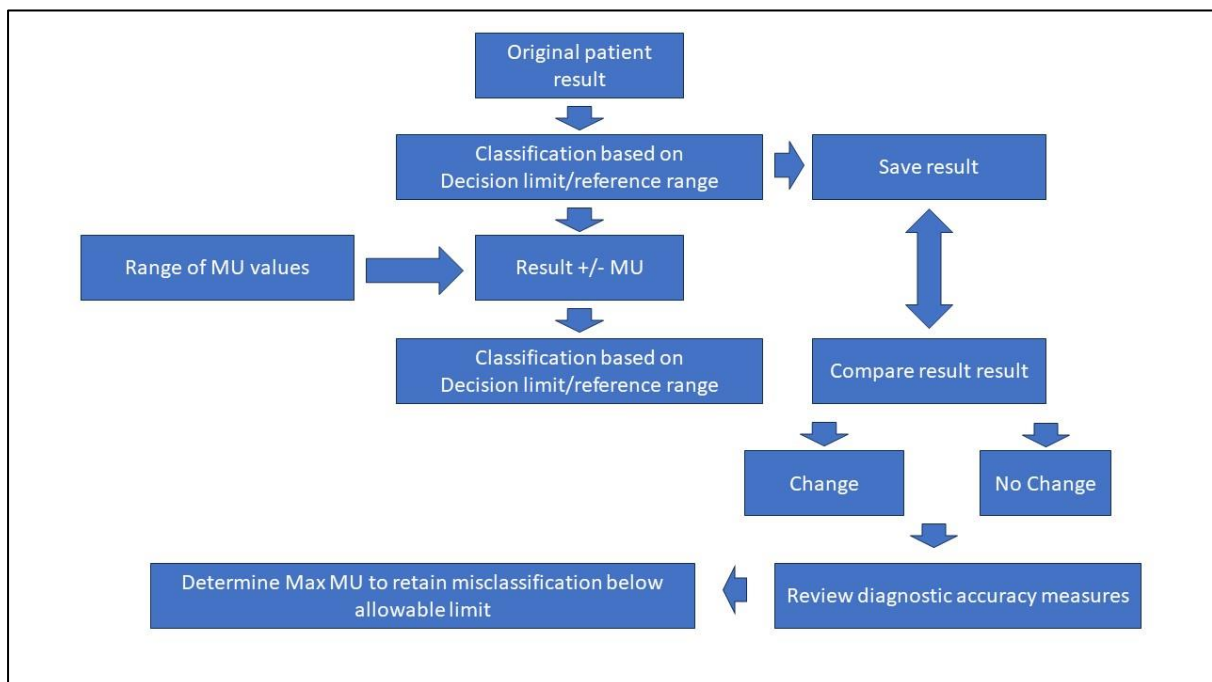


Figure 7: Monte Carlo simulation of increasing uncertainty to patient results.

Figure adapted from (100).

Recent advancements have provided laboratories with tools to apply Monte Carlo simulations to local patient populations (100). By selecting clinically relevant decision limits, laboratories can derive performance specifications tailored to their needs. This method was applied in subsequent sections of this thesis to demonstrate its utility in defining APS for ADAMTS13 activity.

1.5 Biological Variation of ADAMTS13 (Milan Model 2)

Model 2, which derives APS from biological variation (BV), has been a cornerstone of laboratory quality assurance since its early use in external quality assurance (EQA) schemes (101). BV-based APS rely on two key components: within-individual variability and between-group variability (Table 12). From these, specifications for allowable analytical imprecision and bias are calculated. The popularity of BV in APS is largely due to the Total Error (TE) approach, which incorporates BV as its primary basis. However, this approach is not without critics, particularly concerning its universal applicability (102, 103).

Table 12: Model 1 and 2 of the Milan criteria – ADAMTS13 activity assays.

As the two primary APS modelling methods limited data exists for application of both model 1 and model 2 in ADAMTS13 testing

Milan Model	Key Points	References
Clinical Outcome Studies for ADAMTS13 Activity	Limited robust data linking assay performance to patient outcomes. Improvements in analytical performance for immune TTP detection. Need for simulations to model MU effects on decision limits.	(93), (94)
Biological Variation	Model 2 derives APS from BV, used in EQA schemes. Relies on within-individual and between-group variability. Popularity due to TE approach. Challenges in applying BV models to ADAMTS13 due to dynamic regulation.	(104), (102, 105),(106), (107)

The correct application of Model 2 depends on identifying measurands with tight homeostatic control and stable intra-individual set points. ADAMTS13 activity is dynamically regulated by its interaction with VWF, transitioning between latent and active forms (105). This dynamic regulation

presents challenges when applying BV models to ADAMTS13, as its activity is substrate-dependent and not at a steady state. Moreover, while physiological inhibitors of ADAMTS13 remain unidentified, emerging research suggests potential regulatory mechanisms involving thrombin and plasmin (106); (107).

Table 13 highlights published performance specifications for ADAMTS13 activity assays using the Technozym[®] ELISA platform. These include within-run CV (12.7%) and total allowable error (TEa, 14.5%). Of note is that this data is based on a single BV study (108).

Table 13: Published performance specifications for ADAMTS13 activity assays - Technozym[®] ELISA.

CV = Coefficient of variation. TEa = Total Allowable Error, LOD – Limit of Detection, ULQ – Upper Limit of Quantification, LLQ – Lower Limit of Quantification, EDTA = Ethylene Diamine Tetra-acetic Acid, PPV – Positive Predictive Value, NPV – Negative Predictive value published by ECAT

Specification	Expected Performance	Source
Repeatability/precision		
Within run CV	12.7%	ECAT (Voorschoten, The Netherlands)
Between run CV	9.6%	ECAT (Voorschoten, The Netherlands)
Monitoring analytical CV	6.4%	ECAT (Voorschoten, The Netherlands)
Diagnostic analytical CV	9.3%	ECAT (Voorschoten, The Netherlands)
Bias	4.0%	ECAT (Voorschoten, The Netherlands)
TEa	14.5%	ECAT (Voorschoten, The Netherlands)
LOD	0.0071 U/mL (0.7U/dL)	Kit insert (Technozym ELISA)
ULQ	Highest calibrator value	Lot by lot basis
LLQ	Lowest Calibrator or LOD	Lot by lot basis
Linearity	Highest calibrator value	
Interfering substances	EDTA, HIL - as per kit insert	Kit insert (Technozym ELISA)
Sensitivity	98.73 (93.15-99.97)	Kit insert (Technozym ELISA)
Specificity	92.86 (80.52-98.5)	Kit insert (Technozym ELISA)
PPV	96.30 (89.73-98.72)	Kit inserts (Technozym ELISA)
NPV	97.50 (84.74-99.64)	Kit inserts (Technozym ELISA)

Diagnostic and monitoring imprecision targets are reported separately because they reflect different clinical needs. Diagnostic use relies on a single test result to make immediate clinical decisions, often near critical thresholds such as the 10 IU/dL cut-off for ADAMTS13 activity in TTP. As such, bias impacts misclassification significantly. In monitoring, where trends over time are evaluated, imprecision is the major consideration when trying to identify clinically meaningful changes in patient serial results.

The Biological Variation Critical Appraisal Checklist (BiVAC; (109)) ensures the standardisation of BV studies. This checklist provides a structured framework for appraising BV studies by scoring fifteen key methodological components from A (highest quality) to D (lowest quality). The Kilercik study (108) received an overall BIVAC grade of C, with several elements achieving the highest quality rating (A), but others constrained by methodological limitations that reduced the overall strength of evidence.

The Kilercik study involved a cohort of 15 healthy adult volunteers, comprising eight males and seven females, with an age range of 21 to 40 years and a mean age of 30 years. All participants were in good health and free from potentially confounding influences such as pregnancy, medications, special diets, or athletic conditioning. This clearly defined, healthy population resulted in high BIVAC scores for study population (A), measurand (A), and pre-analytical processes (A). The study duration was five weeks, with one blood sample collected per subject per week, resulting in a total of five samples per individual. The gradings do not directly correlate clinical utility of these assays for study evaluation. Weekly testing may be appropriate for frequently tested individuals, for example those undergoing acute treatment post diagnosis however, monitoring of patients is significantly less frequent, up to months between visits.

1.5.1 Sampling Strategy and Analytical Methodology

The sample collection strategy was consistent and appropriate for a BV study, earning an A score for sample handling. However, replicate analyses were not performed; only a single measurement per

sample was reported. This limited the ability to assess analytical imprecision (CV_a) and introduced greater uncertainty into the estimates of within- and between-subject variability. As such, the replicate analysis component was graded as B.

The analytical method used for determining ADAMTS13 activity was not explicitly specified, although the reported analytical coefficient of variation (CV_a) was 4.21%, with confidence intervals ranging from 2.78% to 8.57%. This level of analytical precision is broadly acceptable for biological variation studies, although the absence of analytical replicates remains a notable limitation.

1.5.2 Estimates of Biological Variation and Statistical Evaluation

The calculated within-subject biological variation (CV_i) was 12.73%, with a 95% confidence interval (CI) of 9.72% to 15.77%. Between-subject variation (CV_g) was estimated at 9.63% (CI: 5.62% to 16.48%). These estimates were derived from data collected consistently across all five sampling points, but the confidence intervals for CV_i and CV_g were only partially reported. While calculated CIs were provided, they were not consistently stated in the dataset fields designated for reporting. Nonetheless, the inclusion of CIs contributed to an A score for this component.

Some key details were either insufficiently detailed or omitted entirely. The method for outlier detection and treatment was not adequately described, resulting in a C score for outlier analysis. Assessments of normality and the use of ANOVA for variance component analysis were scored as B, reflecting reasonable but suboptimal statistical treatment. However, variance homogeneity was addressed appropriately and scored A.

The number of subjects and samples was adequate for preliminary estimation but fell below the optimal thresholds for high-quality BV data. The limited sample size (15 subjects with five samples each) led to a C score for the number of results. Increasing the number of subjects and, again, incorporating replicates would have strengthened the reliability of the variance estimates.

1.5.3 Measurand Concentrations and Data Quality

Rather confusingly, the mean ADAMTS13 activity across the dataset was reported as 120%, with a range of 50% to 103%, itself raising questions about the analytical assessment performed in the EFLM database. While the range is clinically relevant, the maximum value appears lower than the mean, suggesting a possible transcription or reporting error.

1.5.4 Implications

According to the BiVAC criteria (109) the Kilercik study (108) represents a methodologically sound dataset on the biological variation of ADAMTS13 activity, with strong performance in areas related to participant selection, pre-analytical control, and sampling standardisation. However, the absence of replicate measurements, incomplete statistical detail regarding outlier analysis and normality assessment, and the relatively small sample size contributed to an overall BIVAC grade of C. It is also worthy of note that the study was also designed to determine the BV of VWF antigen and activities so was not designed with ADAMTS13 specifically (or at least uniquely) in mind.

Despite these limitations, the study contributes valuable preliminary data to the currently limited evidence base for ADAMTS13 biological variation. These estimates may be useful for informing APS or evaluating significant change in serial measurements, particularly in research or early-stage clinical applications.

From the data produced within-subject variability ($CV_i=12.7\%$) and between-subject variability ($CV_g=9.6\%$) are provided (Table 14). Using these metrics, desirable, minimal, and optimal MAU levels can be calculated (Table 15) from a single BV based model (a summary of additional available models is in the following section).

Table 14: Biological variation data from the EFLM database.

Data extracted from <https://biologicalvariation.eu/>

Metric	Biological Variation
% Within-subject (CV_i) estimate	12.7
% Between-subject (CV_g) estimate	9.6

Table 15: Model 2 analytical performance specifications for ADAMTS13 activity.

Desirable, minimal and optimal MU acceptability for ADAMTS13 activity assay according to the EFLM biological variation database (https://biologicalvariation.eu/meta_calculations/.)

Specification	Imprecision (CV_a , %)	Bias (%)	Maximum Allowable Uncertainty (MAU, %)	Total Error (%)
Minimum	9.5	6.0	19.0	21.7
Desirable	6.3	4.0	12.7	14.5
Optimal	3.2	2.0	6.3	7.2

The operational feasibility as well as the conceptual justification of biological variation data requires further study and is addressed in this thesis. It is clear, with different goals set by BV, adherence to these specifications may be variably achievable using modern methodologies.

1.5.5 APS from Biological Variation derived models

APS for analytical variation (CV_a) for ADAMTS13 activity using different BV models are summarised in Table 16, below. From this, bias is included, or not, to derive total allowable error or maximum allowable MU respectively according to Equation 6 and Equation 7.

Table 16 presents a comparative overview of different models used to define allowable analytical imprecision (CV_a) and their corresponding impact on maximum allowable uncertainty (MAU). The

linear model by Fraser (110) incorporates both bias and imprecision components, offering the most lenient specification with a CV_a of less than 14.5%. It reflects an upper threshold for total allowable error and is useful where both random and systematic errors are considered. The Gowans diagnosis model (111), focusing specifically on the role of imprecision in diagnostic misclassification, proposes a tighter limit of <9.6%. For diagnosis, it aligns closely with data from de Maat et al (9.3%) (112). Fraser's minimum model (113) builds on this with a threshold of <9.5%, based on CV_a being no greater than 0.75 times the within-subject biological variation (CV_i). His desirable and optimum models tighten this further to <6.4% and <3.2%, respectively, corresponding to $CV_a \leq 0.5 \times CV_i$ and $CV_a \leq 0.25 \times CV_i$. These models represent increasingly aspirational performance goals.

Table 16: Model 2 Biological Variation Models.

Three published models for derivation of APS based on biological variation data. CV_a – acceptable analytical coefficient of variation, CV_i – intra individual coefficient of variation (- Data verified by the ELFM database).*

Model	Basis of Calculation	Calculation	Specification
Linear model (110)	Includes both bias (top line, right) and imprecision (bottom line, right) terms; represents an upper limit for desirable total error.	$0.25 * (12.7^2 + 9.6^2)^{1/2} + 1.65 * (0.5 * 12.7)$	< 14.5%
Gowans' diagnosis model (111)	Focuses on analytical variation relative to diagnostic variability.	$0.6 * (12.7^2 + 9.6^2)^{1/2}$	< 9.6%
Fraser's minimum model (113)	Uses $0.75 \times CV_i$ as an acceptable upper bound for CV_a .	$CV_a \leq 0.75 * 12.7$	< 9.5%
Fraser's desirable model (113)	Based on $CV_a \leq 0.5 \times CV_i$.	$CV_a \leq 0.5 * 12.7$	< 6.4%
Fraser's optimum model (113)	Most stringent model with $CV_a \leq 0.25 \times CV_i$.	$CV_a \leq 0.25 * 12.7$	< 3.2%
Minimum MAU	Appropriations 50% of MAU to CV_a Frasers minimum model	$MAU < 2 * (0.75 * 12.7)$	< 19
Desirable MAU	Appropriations 50% of MAU to CV_a Frasers desirable model	$MAU < 2 * (0.5 * 12.7)$	< 12.7
Optimum MAU	Appropriations 50% of MAU to CV_a Frasers desirable model	$MAU < 2 * (0.25 * 12.7)$	< 6.4

Table 16 also links these CV_a specifications to maximum allowable uncertainty (MAU) by assuming that analytical imprecision accounts for 50% of the total MAU. Under this assumption, the minimum MAU (based on Fraser's minimum model) is <19%, the desirable MAU (from Fraser's desirable model) is <12.7%, and the optimum MAU (from Fraser's optimum model) is <6.4%. This structured approach allows laboratories to align analytical performance expectations with acceptable levels of clinical risk, depending on whether a minimum, desirable, or optimum level of performance is targeted. The choice among these models should reflect both clinical utility and the feasibility of achieving such performance within routine laboratory practice.

When considering the APS derived from these calculations it is challenging for any method, current or past, to achieve the specifications, common to many haemostasis measurands and resulting in poor adoption of this method for APS derivation.

1.6 State of the Art of ADAMTS13 assays (Milan Model 3)

The assignment of APS for measurands lacking strong biological variation data or clear links to clinical outcomes remains a challenge. In such cases, Model 3—state-of-the-art performance—provides a pragmatic framework for establishing interim specifications. This model is particularly useful for measurands where neither outcome-based targets (Model 1) nor biological variation-derived goals (Model 2) are feasible (97, 114).

Panteghini and colleagues have played a pivotal role in shaping the application of Model 3 (115-117).

Recognising that certain measurands, such as C-reactive protein (CRP), do not align neatly with Models 1 or 2 due to their involvement in a wide range of clinical conditions, they proposed using the best achievable analytical performance as a benchmark. Two papers from the same group evaluated common CRP methods and selected the assay with the lowest MU as the optimal standard. Desirable performance for other methods was then defined as 1.5 times this baseline MU(114, 118). This structured approach not only facilitates realistic target setting but also allows

laboratories and manufacturers to identify and address specific sources of performance variation, such as calibration quality (u_{cal}) and within-laboratory imprecision (u_{RW}).

Beyond CRP, this methodology has been extended to other measurands with similar limitations. For instance, (119) applied a state-of-the-art approach to serum ferritin, using achievable analytical performance to define APS in the absence of sufficient biological variation or outcome data. These applications demonstrate that Model 3 is not simply a fallback but a systematic, evidence-based tool for managing uncertainty and guiding quality improvements until stronger data emerges.

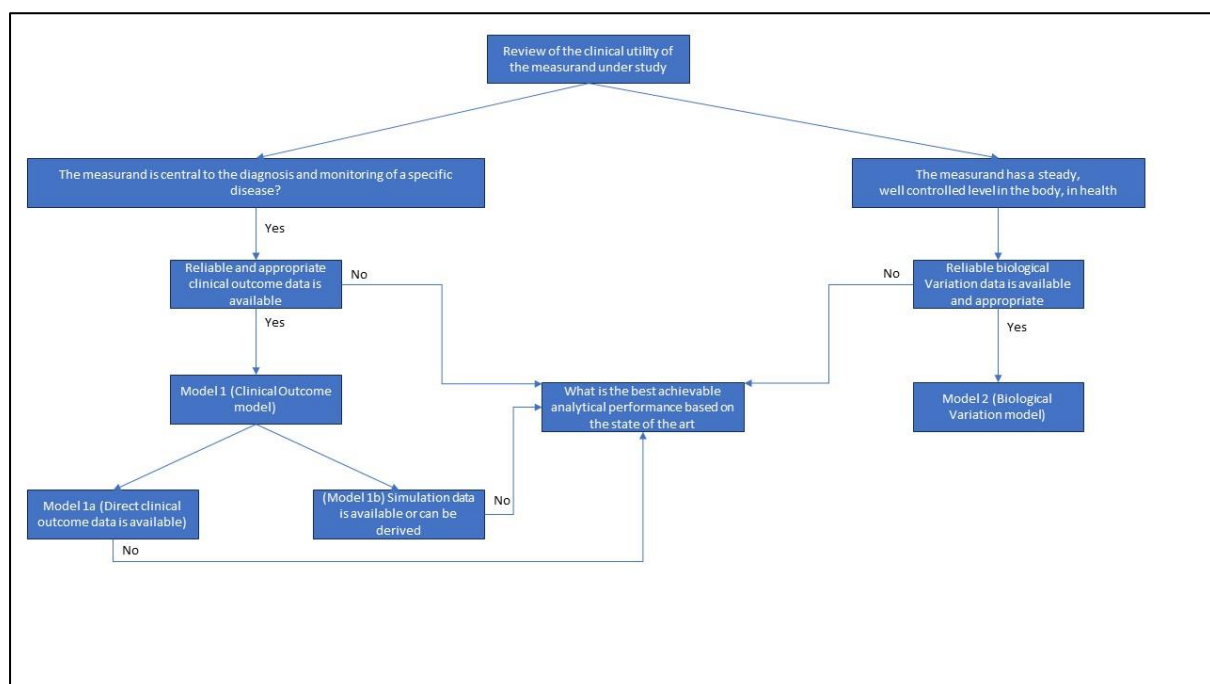


Figure 8: The process of assigning analytical quality specifications.

Figure modified from (120)

Figure 8 outlines the decision-making process for assigning APS to a measurand (modified from (120)). The initial step is to evaluate its clinical role: Is it central to diagnosing or managing a specific condition? Does it exhibit a stable intra-individual baseline in health? If either criterion is met,

Models 1 or 2 may be appropriate, depending on the availability of outcome or biological variation data. Otherwise, Model 3 offers a structured, temporary solution. Crucially, Model 3 is not intended for measurands where existing specifications are simply difficult to meet; it is a methodologically sound framework used while more definitive data are developed (121, 122).

1.6.1 Manufacturer Performance Characteristics of Commonly Available

ADAMTS13 Activity Assays in the UK

In the absence of guideline-driven APS, initial verification of ADAMTS13 activity assays relies on manufacturer-provided validation data (Table 17). These characteristics, validated under controlled conditions, offer insights into assay state of the art precision, sensitivity, specificity, and traceability.

Table 17: ADAMTS13 activity assay performance characteristics.

Summary validation performance characteristics for established and rapid assays of ADAMTS13 activity quantification.

Characteristics shown are selected as those relevant to further studies into clinical and analytical performance. MU –

Measurement Uncertainty, ELISA – Enzyme Linked Immunosorbent Assay, vWF – von Willebrand Factor, QC – Quality

Control, WR CV – within run coefficient of variation, TCV – Total coefficient of variation, IFU – instructions for use. *- as

stated in IFU. **- at time of writing (05/2023) from IFU available from Technoclone website.

Specification	Technozym® ELISA	HemosIL AcuStar®	Ceveron Technofluor®	Technoscreen®
Intended Use	Diagnosis and Monitoring	Diagnosis and Monitoring	Diagnosis and Monitoring	Diagnosis and Monitoring
Principle of Method	Two-step immunoassay	Two-step immunoassay - chemiluminescence	Fluorescence Resonance Energy Transfer (FRET)	Lateral flow based on the Immunoassay principle
Reference Range (from manufacturer)	0.40-1.30 IU/mL	60.6-130.6 %	0.6-1.21 IU/mL (equivalent to 60-121% according to manufacturer)	Semi quantitative results produced < 0.40 IU/mL used to indicate abnormality
Quality Control (QC)	ADAMTS13 High, ADAMTS13 Low	Low/Normal vWF/ADAMTS13 controls	Coagulation controls N and A	Control well on lateral flow through device to detect reagent acceptability
Traceability of Calibrators	WHO standard (12/252) via in house standard	WHO standard (12/252) via in house standard	WHO standard (12/252) via in house standard	None – reference chart calibrated against 12/252 for reference and interpretation
QC Low Performance	CV - 2.9-6.8%	WR CV: 3.2%, TCV: 6.6%	WR CV: 3.3% >1.2 IU/mL, 10.6% at 0.075 IU/mL	NA

Specification	Technozym® ELISA	HemosIL AcuStar®	Ceveron Technofluor®	Technoscreen®
Patient Sample Performance	CV: 5.2–7.8%	Total reproducibility 9.4-13.3%	CV 5.3% - 11.60% (1.2-0.075 % respectively)	Not quoted – comparison study performed (see below)
Clinical Performance	Sensitivity: 84.8% (28/33) 95% CI: 69.1% to 93.3% Specificity: 97.1% (101/104) 95% CI: 91.9% to 99.0% PPV: 90.2% (28/31) 95% CI: 75.2% to 96.6% NPV: 95.3 (101/106) 95% CI: 90.0% to 97.8%	Sensitivity: 88.7%, Specificity: 90.4%	Not quoted – comparison data – slope 1.059, Intercept - -0.005, Correlation – 0.972	Based on a cut off of 0.1IU/ml: positive predictive value 99%, negative predictive value 80%, sensitivity 94% and specificity 97%.

The Technozym® ELISA demonstrates excellent precision, with within-run coefficients of variation (WR CV) consistently below 10%, although its manual format restricts its use in non-specialist laboratories. The fully automated HemosIL AcuStar® offers faster processing but exhibits greater variability, particularly at lower activity levels as does the Ceveron Technofluor®, employing a FRET-based approach. Meanwhile, the Technoscreen® assay is advantageous for rapid, point-of-care testing, though its absence of quantitative calibration diminishes its utility for detailed disease monitoring. The data provided in the Instructions for Use (IFU) of methods is usually the first point of contact with expected performance of a specific assay. It is this data that is often used as a benchmark for new assay implementation. From there comparison and verification studies are required to confirm analytical and clinical acceptability of performance. However, as will be shown throughout this thesis, study design as well as choice and interpretation of statistical methods actually has a more significant impact on interpretation.

1.6.2 Interpreting verification of ADAMTS13 assays

Statistical methods play a crucial role in method comparison studies, helping laboratories evaluate new measurement procedures and validate manufacturer performance claims. However, one major limitation in current practice is the absence of universal guidance on how to incorporate MU into these verifications. This lack of standardisation has resulted in varied approaches across laboratories and platforms, highlighting a need for further methodological development and harmonisation. Understanding the assumptions, strengths, and limitations of available statistical tools is essential for selecting appropriate techniques that balance analytical rigour with clinical applicability.

Correlation, typically measured by Pearson's correlation coefficient (r), is one of the most widely used statistics in method comparison but is frequently misapplied. It quantifies the strength of association between two variables but does not provide any information about agreement (Table 18). High correlation can exist even when the methods produce systematically different results. This issue was demonstrated by Valsecchi *et al.*, who observed an r value of 0.96 in two method comparisons, despite one relationship being non-linear (123). To overcome this limitation, Lin's Concordance Correlation Coefficient (CCC) extends Pearson's r by evaluating both precision and accuracy. CCC compares data against the 45° line of perfect agreement, offering a more comprehensive assessment (124), however, it is not widely implemented in standard software packages, and its calculations are often considered complex for routine use.

Table 18: Comparison of Common Statistical Approaches Used in Method Comparison Studies

Method	Purpose / Description	Assumptions	Advantages	Challenges / Disadvantages
Correlation (Pearson's r)	Strength of association between two variables.	Interval or ratio data; normal distribution.	Commonly used and simple to compute.	Misleading in method comparison; does not assess agreement.
Lin's CCC	Extends Pearson's r to assess both precision and accuracy.	Same as correlation, plus concordance assumptions.	More robust than correlation alone.	Complex to calculate; not widely available in commercial software.
OLS Regression	Minimises vertical residuals between measured points.	Linearity; homoscedasticity; independent errors.	Frequently used and familiar to most practitioners.	Ignores MU in comparator; can lead to biased slope and unrealistic extrapolations.
Deming Regression	Accounts for MU in both comparator and candidate methods.	Constant variance ratio (λ); homoscedasticity.	More appropriate for method comparison; considers error in both axes.	Requires estimation of λ ; assumes constant variance ratio across the range.
Passing-Bablok Regression	Non-parametric regression robust to outliers; calculates slope and intercept.	Linearity; continuously distributed data.	Robust to outliers; slope and intercept fall within confidence intervals.	May require bias correction for non-independent observations.

Method	Purpose / Description	Assumptions	Advantages	Challenges / Disadvantages
Bland-Altman Analysis	Assesses agreement by plotting differences vs means; detects systematic and proportional bias.	Normally distributed differences.	Offers intuitive visualisation of agreement and bias.	Wide limits of agreement; confidence intervals must be interpreted cautiously.

Ordinary Least Squares (OLS) regression is another commonly applied method, yet it is poorly suited to method comparison due to its focus on minimising vertical residuals. This focus on one-axis error introduces distortion when both the comparator and candidate methods carry MU, often leading to unrealistic extrapolations such as negative values when slope coefficients are below 1. Deming regression addresses this issue by accounting for uncertainty in both axes, assuming a constant variance ratio (λ) between the methods (125, 126). For data with proportional variances, weighted Deming regression provides a more appropriate solution. Passing-Bablok regression is a non-parametric alternative that is robust to outliers and well-suited to clinical method comparisons (126). It provides estimates of slope and intercept, with ideal values of 1 and 0, respectively, being interpreted within confidence intervals for clinical acceptability.

Bland-Altman analysis complements these methods by offering a graphical approach to assess agreement (127, 128). It plots the differences between two methods against their mean, allowing for the identification of systematic and proportional bias across the measurement range. The visual nature of Bland-Altman plots makes them particularly effective for detecting trends or shifts not easily captured through numerical statistics alone. However, confidence intervals for the limits of agreement must be interpreted with caution, particularly in the presence of high MU or near critical

decision thresholds. Ultimately, combining statistical analysis with visual tools and clinical insight ensures that method verification is both analytically robust and relevant to patient risk.

1.7 Representing Clinical Performance in comparison studies

1.7.1 Error Grid Analysis

Traditional statistical methods for method comparisons, such as those discussed above, fail to account for the clinical impact of errors, treating all deviations equally regardless of their significance. Error grid analysis (EGA) addresses this limitation by incorporating clinical significance into the evaluation of method performance, offering a risk-based approach that aligns with ISO 15189:2022 principles.

EGA provides a two-dimensional plot where measured values are compared against reference values, dividing the plot into zones representing varying levels of clinical significance (129). Unlike traditional methods that focus on statistical and analytical variation, EGA prioritises clinical utility, assessing how differences in measurement results affect clinical decisions and patient outcomes.

1.7.1.1 *Clinical Zones of EGA*

EGA typically divides measurement ranges into zones, each reflecting a specific level of clinical significance (Table 19). The zones categorise errors based on their potential to cause patient harm.

For ADAMTS13 activity assays, deviations in Zones D or E could lead to inappropriate clinical decisions, such as unnecessary PEX or withholding life-saving interventions like caplacizumab.

Table 19: Error Grid zones.

Classification of Error Grid Zones by Magnitude of Result Deviation and Associated Clinical Impact

Zone	Classification	Deviation	Clinical Impact
A	Clinically acceptable errors	Small, direction unimportant	No significant impact on patient outcomes
B	No significant impact on clinical outcomes	Larger than Zone A	Minimal or no clinical significance
C	Potential for significant clinical impact	Moderate to large deviations	May require clinical intervention
D	Significant impact	Substantial deviations	Likely to affect patient care decisions
E	Severe errors requiring immediate action	Extreme deviations	Potentially harmful, requiring urgent correction

1.7.1.2 Clarke and Parkes Error Grids

The Clarke Error Grid (CEG) was originally developed for evaluating blood glucose meters and is designed to assess the clinical significance of discrepancies between a test method and a reference (129-132). It divides the plot of reference versus test results into five zones (A–E) (Figure 9), each representing different levels of clinical acceptability. Zone A includes values that are clinically accurate or differ slightly but would not lead to inappropriate treatment. Zone B reflects benign errors with little or no clinical consequence. Zones C, D, and E represent increasingly serious errors, from unnecessary treatment (Zone C), failure to detect hypo- or hyperglycaemia (Zone D), to errors

that could result in dangerous treatment decisions (Zone E). The CEG provides a visual tool to understand how measurement inaccuracies translate into potential clinical risks.

The Parkes (Consensus) Error Grid was introduced as an enhancement of the Clarke model, incorporating expert consensus to refine the interpretation of glucose meter accuracy (133). It is designed for both type 1 and type 2 diabetes and divides the error grid into 15 zones grouped into five broader risk categories, focusing on the severity of clinical consequences resulting from incorrect treatment decisions. Unlike the Clarke grid, the Parkes model accommodates a broader range of clinical perspectives and treatment strategies, providing a more nuanced assessment of patient risk. The Parkes grid is especially valuable for regulatory and clinical decision-making, offering improved sensitivity to clinical practice variability while still visually displaying the agreement between methods.

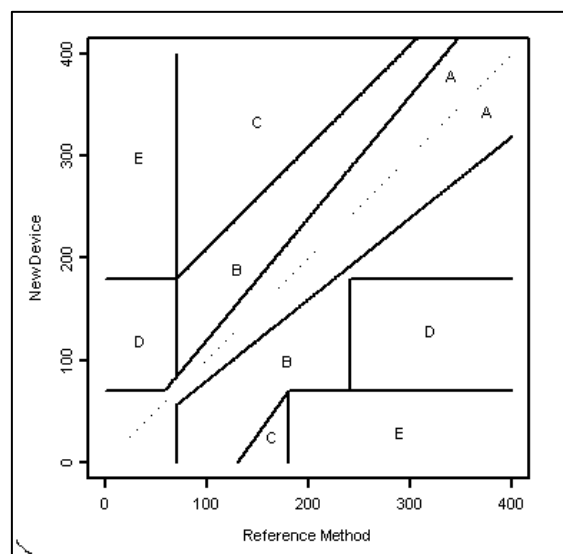


Figure 9: The Clarke Error Grid.

The clinical significance of differences between test and reference measurements are represented by Zones A–E indicating increasing levels of risk associated with inaccurate results. Image sourced from:

https://en.wikipedia.org/wiki/Clarke_Error_Grid

1.7.1.3 Direction of Errors

In most traditional Error Grid Analyses, the direction of measurement errors—whether a result is falsely elevated or depressed—is not explicitly considered. However, in the context of assays like ADAMTS13, the direction of error is critically important due to the significant clinical consequences associated with both underestimation and overestimation of enzyme activity. A falsely low ADAMTS13 result, for instance, may lead clinicians to initiate or unnecessarily prolong PEX therapy and administer caplacizumab inappropriately. These interventions, while potentially life-saving, carry risks of complications such as bleeding, allergic reactions, and increased healthcare burden. On the other hand, a falsely high ADAMTS13 result may delay or prevent essential treatment, thereby exposing the patient to the risk of ongoing or recurrent TTP potentially resulting in avoidable morbidity or mortality. These examples illustrate how the directionality of error can profoundly affect clinical decision-making, underscoring the importance of the design of EGAs for ADAMTS13 assays later in the thesis.

1.7.1.4 Interpretation of Error Grids

Error grids are most useful when they provide a clear visual summary of how well a test method aligns with a reference across clinically significant thresholds (133). When interpreting an EGA, several key principles apply. First, a well-performing method should display the majority of its data points within Zone A, which indicates clinically acceptable agreement with the reference method. Zones B and C suggest areas where the method may deviate slightly from the reference, but such deviations are usually still acceptable for clinical use. In contrast, Zones D and E represent serious discrepancies that may lead to harmful clinical outcomes (129, 132, 133). These zones highlight where the method fails to provide safe or reliable information and should trigger immediate investigation and corrective action.

The structured layout of error grids aids clinicians in visually assessing diagnostic performance, linking analytical error directly to potential clinical consequences. This mitigates overconfidence in

numerical accuracy by contextualising the error in terms of patient safety. EGAs also allow laboratories to identify specific thresholds or conditions under which a test may underperform, thereby informing quality improvement efforts. However, the utility of EGAs depends heavily on the availability of robust reference methods and sufficient data to populate all relevant zones accurately. This can be a limiting factor in fields like haemostasis, including ADAMTS13 testing, where validated reference methods are limited or non-existent. Consequently, while EGAs are a powerful tool for method evaluation, their application in specialised areas requires careful consideration of these limitations and, where necessary, the adaptation of grid design and zone definitions to fit the clinical context.

1.7.2 Likelihood Ratios for clinical interpretation of Laboratory results

The interpretation of laboratory measurements requires assessment of both analytical and clinical equivalence, which are often influenced by the methodologies applied during test verification and validation. Analytical equivalence is typically assessed using statistical tools such as correlation coefficients, regression analyses (e.g., Deming or Passing-Bablok), and Bland-Altman plots. These approaches quantify agreement and bias between methods, offering insight into systematic differences. Clinical equivalence, by contrast, frequently relies on measures of diagnostic accuracy—such as EGA but most often sensitivity, specificity, and classification metrics relative to a predefined clinical decision limits.

These clinical decision limits are themselves based on reference studies using standardised measurement procedures. For these limits to be clinically useful, they must be commutable across different analytical platforms. A significant limitation arises when continuous quantitative results are dichotomised using such thresholds. While dichotomisation offers pragmatic value in clinical workflows, it often results in a loss of diagnostic information. This is especially problematic near the cut-off, where test performance is least discriminatory and the influence of MU is greatest. As a

result, interpretation becomes increasingly ambiguous in precisely the region where clarity is most needed.

1.7.3 Likelihood Ratios: A Bridge Between Test Performance and Clinical Utility

To overcome the limitations inherent in binary classification, likelihood ratios (LRs) provide a robust alternative. LRs quantify how a test result modifies the probability of disease and are derived directly from diagnostic sensitivity and specificity.

Positive Likelihood Ratio (LR^+) indicates how much more likely the disease is to be present following a positive test result, while negative Likelihood Ratio (LR^-) indicates how much less likely the disease is to be present following a negative test result (134-136).

Crucially, these ratios can also be extended beyond binary interpretation. Instead of relying solely on a fixed cut-off, LRs can be calculated across a continuum of results, allowing more granular assessment of diagnostic power (137). This is particularly advantageous for quantitative haemostasis tests such as ADAMTS13 activity, where interpretation is not always clear-cut, and clinical outcomes may be heavily influenced by results near the threshold.

1.7.4 Bayesian Reasoning in Diagnostic Interpretation

The clinical value of likelihood ratios lies in their integration within a Bayesian framework. Bayes' Theorem allows clinicians to update a pre-test probability—often based on clinical assessment tools such as the PLASMIC, Bentley or French scores—into a post-test probability that reflects the new diagnostic information provided by the laboratory result. This approach aligns well with intuitive clinical reasoning and supports evidence-based decision-making.

For instance, a test with an LR^+ of 10 and a pre-test probability of 30% yields a post-test probability of approximately 77%, offering substantially greater diagnostic certainty. Conversely, a LR^- of 0.1 would decrease the probability to around 4%, suggesting strong evidence against the condition. These examples illustrate how LRs bridge the gap between statistical analysis and clinical utility.

1.7.5 Application to ADAMTS13 and Haemostasis Testing

The interpretation of ADAMTS13 activity is a clinically critical use case for this framework. Severe deficiency (typically <10 IU/dL) is a key diagnostic feature of TTP (1-3). Yet the discriminatory power of results in the 5–20 IU/dL range is often unclear (138). Using likelihood ratios, these results can be expressed in terms of their ability to shift the probability of TTP diagnosis, particularly when incorporated alongside PLASMIC scoring. This provides a more continuous, evidence-based interpretation model.

Moreover, recent work describes a distribution-independent approach to calculating likelihood ratios from quantitative results using Bézier curves fitted to ROC data (139). This method allows for test-specific likelihood ratios to be estimated even in the absence of clearly defined result distributions. In parallel, Habibzadeh and Habibzadeh demonstrated the graphical equivalence of LRs and the slope of the tangent to the ROC curve at a given result point, reinforcing the mathematical and interpretive foundations of this approach (135).

1.7.6 MU and Likelihood Ratios

Explored for the first time in this thesis, MU may play a role in refining the interpretation of likelihood ratios. First, it affects the reliability of the test value used to derive an LR. Second, it can be incorporated directly into the LR framework by calculating interval LRs or applying guard bands, which define a diagnostic confidence range for a given result. ISO 20914:2019 emphasises that laboratories must consider uncertainty when interpreting results near clinical decision limits as does CLSI EP29-A-2012.

Likelihood ratios offer a statistically rigorous yet clinically intuitive method for interpreting quantitative test results, particularly when binary classification may be insufficient. In haemostasis testing—especially for assays such as ADAMTS13 activity—this approach allows for more precise, probability-based decision-making that incorporates both clinical context and analytical uncertainty.

An additional application of combining MU with LR is that in order to maintain an expected clinical performance of a test, in the LR context, adjustment of the clinical decision limit to ensure diagnostic information is not lost by local laboratory practices is explored in this thesis. This means that, by changing the cut off, a result around that cutoff can be reported in LR and in doing so the posterior likelihood is retained despite different performance when an assay is used in any given laboratory. Doing so can harmonise clinical interpretation and even reporting units if successful.

1.8 Measurement Uncertainty

MU is a foundational concept in laboratory medicine and is formally defined by the International Vocabulary of Metrology (140), 2.26) as “a non-negative parameter characterising the dispersion of quantity values attributed to a measurand.” In essence, it reflects the doubt associated with the result of a measurement and quantifies the range within which the true value is believed to lie, with a given level of confidence. This concept is particularly critical in the medical laboratory setting, where small numerical differences can have significant clinical implications.

MU is deeply embedded within the quality management framework of ISO 15189:2022, the international standard for quality and competence in medical laboratories. Several clauses within the standard directly reference the need to understand, estimate, and manage MU. Notably, Clause 7.3.2.1 requires laboratories to determine MU for each quantitative measurement procedure used in patient testing, and to make this information available to users of laboratory services. MU also plays a role in Clause 7.7.1, which links MU to the validity of examination results, and in Clause 7.10, which addresses the interpretation and communication of results—emphasising the importance of understanding uncertainty when results approach clinical decision thresholds.

The importance of MU goes beyond regulatory compliance. Clinically, it ensures that patient results are not only accurate but also reliable when used in diagnosis, monitoring, or therapeutic decision-making. For example, in cases such as therapeutic drug monitoring (141, 142) or the evaluation of

coagulation factors (143, 144), the precision and uncertainty of reported results can directly affect result interpretation. Understanding MU supports metrological traceability and linking laboratory results through a chain of calibrations to a recognised reference or international standard (145-147).

In practice, the estimation of MU in medical laboratories typically involves combining information on within-laboratory imprecision—often derived from IQC data—with information about calibration and bias, where available. Guidelines such as ISO/TS 20914:2019 and CLSI EP29-A provide practical frameworks for this process. These documents recommend using a top-down approach, leveraging routine laboratory data to estimate long-term performance characteristics, and incorporating uncertainty components from calibrators (u_{cal}), reference materials or procedures (u_{ref}), and internal imprecision (u_{RW}). This allows laboratories to ensure that their measurement systems are both analytically sound and clinically meaningful.

1.8.1 The Standard Uncertainty

The standard uncertainty is the building block of any MU estimate. It is derived from one of two approaches: Type A or Type B, as outlined in the Guide to the Expression of Uncertainty in Measurement (JCGM:100 (GUM), Section 2.3.2). Type A evaluation, widely applied in medical laboratories, uses statistical methods based on repeated measurements. It assesses imprecision without needing to identify the source of variability and typically draws data from IQC, EQA and method comparison studies.

Key statistical tools in Type A evaluation include the arithmetic mean, standard deviation (SD), variance, relative standard deviation (RSD or CV), and the standard deviation of the mean. The arithmetic mean helps estimate systematic error (bias) in top-down MU approaches and serves as the central tendency for normally distributed data (JCGM 100:2008). SD quantifies random variation, while variance (SD^2) enables combination across datasets. CV expresses imprecision relative to the mean and is often used at the final stage of MU estimation. The standard error of the mean, calculated by dividing SD by the square root of the sample size, reflects uncertainty in the mean.

Type B evaluation, in contrast, uses non-statistical sources such as historical IQC data, manufacturer specifications, calibration certificates, and expert judgement. It addresses both random and systematic errors but often lacks raw data. The data may not follow a normal distribution and typically require transformation into standard uncertainties before combining with Type A estimates.

Different probability distributions are applied in Type B evaluations. The normal distribution, defined by a mean (μ) and standard deviation (σ), includes 68.2% of values within $\pm 1\sigma$ and 95.5% within $\pm 2\sigma$.

The rectangular distribution assumes uniform likelihood across a range and calculates standard uncertainty by dividing the range by $\sqrt{3}$. The triangular distribution, favouring central values, uses $\sqrt{6}$ in the denominator. While the log-normal distribution is more common in patient data than in analytical MU, it may apply where skewed data exist. When distribution type is unclear, GUM advises choosing a conservative option, often the rectangular distribution, to avoid underestimating uncertainty.

1.8.2 Frameworks for Uncertainty Estimation

Multiple methods exist to estimate MU, each grounded in different assumptions and computational approaches. These methods are vital for ensuring diagnostic reliability, particularly near clinical decision thresholds. Despite their relevance, comparative analysis of MU approaches in medical testing remains limited.

The table overleaf summarises the advantages and disadvantages of three widely used methods for MU estimation in medical laboratories: the Top-Down Method, the Law of Propagation of Uncertainty (LPU, GUM Method), and the Law of Propagation of Distributions (Monte Carlo Method) (Table 20).

Table 20: Advantages and Disadvantages of MU Estimation Techniques.

MU Estimation Method	Advantages	Disadvantages
Top-Down	- Simple and easy to implement. - Practical: relies on readily available laboratory data. - Transparent: traceable to reference methods and standards.	- May not accommodate complex measurement procedures. - Estimates can be overly conservative.
Law of Propagation of Uncertainty (GUM Method)	- Comprehensive: incorporates all known uncertainty sources. - Metrologically valid according to measurement standards. - Based on rigorous mathematical principles.	- Complex: requires advanced mathematical knowledge. - Assumes Gaussian distributions, which may not always apply. - Additional data often needed for comprehensive analysis.
Law of Propagation of Distributions (Monte Carlo Method)	- Highly flexible: accommodates non-linear relationships. - Simulates multiple scenarios for realistic uncertainty representation. - Explores the entire variable space.	- Computationally intensive. - Requires expertise in statistical modelling. - Input data must be carefully validated for distributions and correlations.

1.8.3 Challenges in MU Estimation

Uncertainty analysis in medical laboratories, regardless of the chosen methodology, is fundamentally constrained by the quality and scope of available data at the time of evaluation. Routine sources such as IQC results and calibrator performance data provide a practical basis for estimating MU, but they often fall short in representing values at or near critical clinical decision thresholds (148, 149). This disconnect can lead to challenges when attempting to interpolate uncertainty estimates for

results that fall outside the directly assessed calibration or control points, potentially undermining the clinical relevance of the uncertainty assigned to patient results.

To overcome these limitations, more robust and adaptable approaches are needed—particularly those capable of addressing data gaps and accounting for variability across the full measurement range. This study highlights several key considerations to enhance MU estimation frameworks in medical laboratories. First, uncertainty estimates should be explicitly aligned with clinical decision thresholds to ensure they are meaningful in real-world diagnostic and monitoring contexts, even when IQC and calibrator concentrations do not match these critical values (148). Second, methods must be capable of handling data that deviate from a Normal distribution, as is often the case with novel or specialised assays; incorporating non-Gaussian models can significantly improve the robustness of uncertainty estimates. Finally, computational flexibility is essential, approaches such as Monte Carlo simulations offer a powerful means to model complex sources of variability and provide more representative uncertainty estimates that represent modern laboratory measurement systems better (143, 150).

Estimating MU accurately depends on identifying key sources of variability affecting the final patient result—especially near clinical decision thresholds. ISO/TS 20914:2019 supports a top-down approach that simplifies estimation by grouping uncertainty components, though this can obscure their interactions. For example, imprecision may also affect calibration uncertainty (u_{cal}), bias (u_{bias}), or reference uncertainty (u_{ref}), depending on their derivation.

Typically, 3–4 main contributors are included in the MU model, if this satisfies propagation rules (JCGM 100:2008). Lesser contributors can be excluded if they represent under 30% of the largest uncorrelated or 10% of a correlated source (JCGM 100:2008). Overlooking relevant factors risks underestimating MU, while double counting can inflate it. A balanced approach focuses on contributors that significantly affect clinical outcomes, excluding well-controlled elements unlikely to impact interpretation.

Long-term IQC data often serve as a proxy for within-laboratory imprecision in top-down models. However, some methods or clinical contexts may require detailed analysis along the full traceability chain to capture all relevant uncertainties (103, 151, 152).

The Ishikawa diagram (153), is a visual tool used to categorise causes of measurement variation. Its use in MU modelling is growing, offering a structured way to display how components contribute to uncertainty.

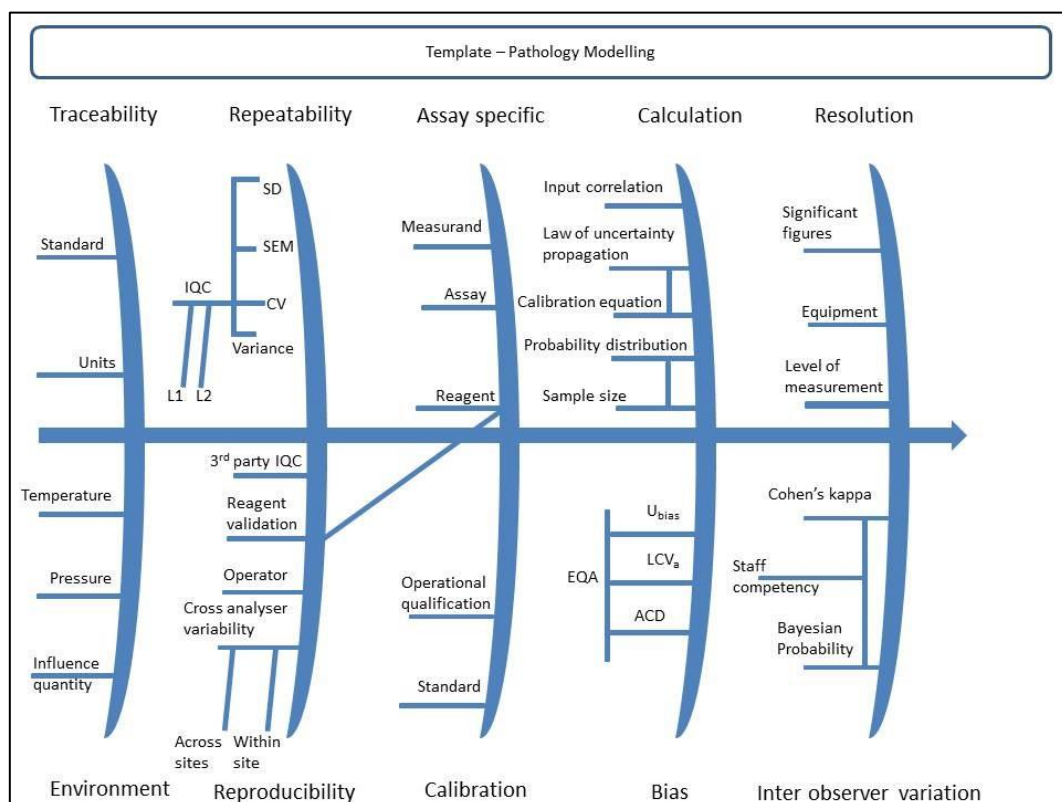


Figure 10: The Ishikawa diagram.

Contributors to uncertainty in medical laboratory tests. Diagram conceived and developed by the author, available at: <https://www.pathologyinpractice.com/story/20476/assay-modelling-the-next-step-in-measurement-uncertainty-evaluation>

Figure 10 shows an Ishikawa diagram developed by the author to represent uncertainty sources in a laboratory assay. While not routine in clinical labs, such diagrams are established in analytical sciences (Ellison, 1998) and can help systematically assess and reconcile MU components before model finalisation.

1.8.4 Application in Medical Laboratories

Ishikawa diagrams offer a useful starting point for identifying and categorising sources of MU in medical laboratories. However, their traditional design assumes linear, mutually exclusive contributors—an oversimplification that rarely reflects the complexity of clinical assays. In practice, contributors such as imprecision can influence other components like bias (u_{bias}), making it difficult to represent interactions clearly. Despite these limitations, Ishikawa diagrams can be valuable in the early stages of MU analysis. For example, they can help clarify how long-term IQC data may already account for short-term systematic shifts, helping laboratories distinguish between absorbed and unaccounted sources of variability (154).

A more nuanced challenge in MU estimation is the correlation between contributors. Whether contributors are fully correlated, partially correlated, or uncorrelated significantly affects how their uncertainties combine (155). Fully correlated sources amplify overall uncertainty and require a “penalty” in the combined estimate. For partially correlated contributors, conservative assumptions are used—treating partial positive correlations as fully positive and partial negative correlations as uncorrelated—to avoid miscalculating the impact (155). When contributors are uncorrelated, their effects are considered independent and combined using standard statistical methods. In practice, if the difference between assuming full correlation and no correlation is small (typically within 20–30%), the correlation’s influence may be considered negligible and excluded from the final calculation. This practical boundary helps laboratories balance complexity and accuracy in uncertainty modelling.

1.8.5 Using APS to Budget MU

The Milan models initially did not specify how total MU should be distributed across the traceability chain. To address this, Braga and colleagues proposed a structured approach that allocates the MU budget among three key contributors, helping laboratories and manufacturers manage and monitor sources of uncertainty more effectively (156).

The first contributor is the reference measurement procedure (u_{ref}), which is limited to no more than 33% of the total maximum allowable uncertainty (MAU). This represents the uncertainty associated with the highest-level, most accurate reference system. The second component, calibration uncertainty (u_{cal}), is capped at 50% of the total MAU (Figure 11). This includes the contribution from u_{ref} and reflects the uncertainty introduced through the calibration chain, encompassing both manufacturer and reference calibration processes. The third and final component is within-laboratory variation (u_{RW}), which accounts for the remaining portion of the MU budget. This component captures the combined effects of baseline system performance and local laboratory variation, such as reagent lot changes, equipment drift, and environmental influences.

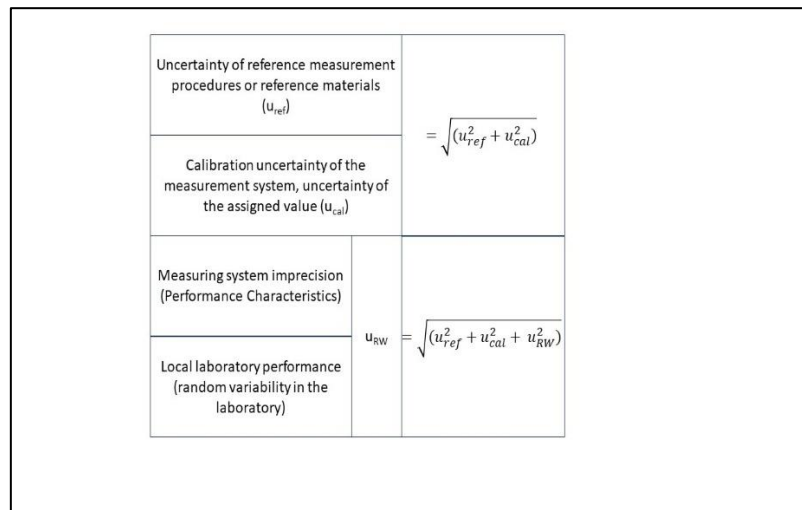


Figure 11: Apportioning of the MU budget.

Budget requirements are determined according to the traceability chain of the measurand. Defined limits are applied for each of u_{ref} , u_{cal} and u_{RW} . In doing so areas for improvement can be identified

Figure 11 illustrates the distribution of the MU budget, identifying areas where improvement efforts can be targeted. In particular, u_{RW} includes not only variation inherent to the system as delivered by the manufacturer but also variation introduced by the local laboratory environment. As such, this component requires continual oversight and control through well-designed IQC to minimise long-term imprecision and ensure the stability of performance over time.

In practical terms, conformance with APS is primarily monitored and maintained through IQC processes. While u_{cal} may vary across different calibrator lots—especially if traceability chains are updated or revised— u_{ref} is generally stable unless new international standards are adopted. This stability underlines the importance of ensuring that APS are not only analytically achievable within the current state of the art but also clinically meaningful, supporting decisions that directly affect patient outcomes.

2 Literature review

A discussion of the application of MU in Haemostasis assays has been submitted as an invited review to Seminars in Thrombosis and Haemostasis in 2025, and is presented below (157).

2.1 MU in the Haemostasis Laboratory

2.2 Abstract

Clinical laboratory results are inherently uncertain. Variability arises when repeating measurements of the same sample under different conditions. MU is a metrological framework that quantifies this variability, enabling clinical interpretation and comparison of results within and between patients. In haemostasis assays, MU describes uncertainty propagation across the entire traceability chain—from international standards to final patient results.

The GUM offered a statistically rigorous but complex method for estimating MU, limiting its widespread adoption. ISO/TS 20914:2019 introduced a Top-down approach that balances rigour with operational feasibility, supporting ISO 15189 compliance. However, its applicability to specialist areas such as haemostasis has been questioned. This review explores the implementation of different approaches for MU in haemostasis laboratories.

2.3 Introduction

MU refers to the quantified doubt about the result of a measurement and is defined as “a non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used” (158). As every measurement is susceptible to error, a measurement result is only complete if accompanied by a statement of its uncertainty (159). This uncertainty informs the user of its appropriateness for its intended clinical use. MU helps to understand that every reported result lies within a range of possible true values, shaped by both random variation (imprecision) and systematic influences (bias). MU enhances interpretability when comparing results to clinical decision limits or monitoring serial results in a patient over time. In addition, calibration and metrological traceability as part of MU also contribute significantly to setting performance specification for laboratory assays (122).

Since 2012, medical laboratories have had access to CLSI EP29A for guidance on MU (160). However, it was hoped that a more concise, universally applicable standard would emerge. ISO/TS 20914:2019 partially fulfils this need (161), translating complex metrological principles into a more accessible framework. Yet, as with many broad guidance documents, discipline-specific nuances limit its universal applicability (162).

In haemostasis, this challenge is particularly significant to demonstrate compliance to ISO15189 (163). The body of literature applying MU principles to haemostasis assays is limited, with a variety of different methods employed, particularly for the INR even if this work was not historically labelled as MU (164-166).

The measurand, as defined by the International Vocabulary of Metrology (158), is the quantity intended to be measured. Incomplete definitions can compromise result interpretation. This is evident in measuring extended half-life (EHL) factor replacement therapies and non-replacement therapies, where discrepancies occur between one-stage assays (OSA) and chromogenic substrate

assays (CSA) in both diagnosis (167) and treatment monitoring. Factor VII activity assays are influenced by thromboplastin sources due to differences in rabbit versus human tissue factors. The FVII Padua mutation shows sensitivity to reagent selection (168). Similar considerations affect thrombin time, antithrombin activity, and chromogenic Factor VIII assays. Reagent origins and assay conditions significantly impact results, emphasizing the need for clear measurand definitions and standardisation in reporting. While extensive descriptions may add complexity, key methodological distinctions—such as one-stage clotting assays versus chromogenic assays—should be clearly defined at both the requesting and reporting stages. ISO/TS 20914:2019 provides an example with the INR defining the measurand as: “Tissue Factor-induced relative time to form a clot in citrated plasma” (169).

Regardless of the measurand definition, there are many ways to calculate MU. Historically, different methods have been applied even for the same measurand. The simplest approach is using internal quality control, calibration, and bias data through the top-down method. The bottom-up method, although more complex and metrologically valid, is not widely used in clinical laboratories. There is limited published data on these methods and alternative Monte Carlo-based approaches.

2.4 The Bottom-Up Approach to quantifying MU – the INR

The bottom-up method constructs uncertainty using fundamental principles by modelling each input that affects the final result. It describes a formal metrological process that starts by defining the mathematical model of a measurement system and then estimating the contribution of each input to the overall uncertainty of the result (159). The bottom-up method provides a complete characterisation of MU through the propagation of uncertainty from its sources (159, 170, 171).

According to the Guide to the Expression of Uncertainty in Measurement (GUM) (172), partial derivatives (sensitivity coefficients) model how variability in each input affects the final measurement result. When calculating a measurand's output value using other measured inputs like ratios or

logarithmic transformations, the uncertainty of each input must be quantified and propagated mathematically.

This approach focuses on derived quantities. Like in other laboratory disciplines, many haemostasis parameters are calculated using mathematical relationships that include constants, which have inherent uncertainties (159). These uncertainties should be included in the MU budget, but are often excluded. For assays based on computed ratios or correction formulas, such as lupus anticoagulant tests or inhibitor screens, the propagation of individual uncertainties must be accounted for. According to the GUM excluding the standard uncertainties of these constants prevents accurate assessment of their full impact on the final result (172).

The bottom-up approach is rarely used in clinical labs due to its mathematical complexity, involving differential calculus, sensitivity analysis, and error propagation. This is further complicated by manufacturers' lack of transparent data on uncertainties. Despite these challenges, it provides the most accurate uncertainty estimation and is crucial for reference methods, calibration of certified materials, and inter-lab comparisons.

The INR is one of the few assays where LPU has been studied in depth, often used as an example. Ironically, INR is among the most complex assays for MU characterisation, with frequent methodological shortcomings reported (161, 173, 174).

2.5 Recent discussion of MU of INR

ISO/TS 20914:2019 includes the calculation of INR uncertainty in appendix A9, using the WHO 4th IS 2009 thromboplastin for traceability (169). The data is derived from internal quality control of PT, with PT's standard uncertainty factored into the INR equation. The MNPT uncertainty was 4.0% and 3.92% for two thromboplastin lots. The ISI is assumed to be constant with no uncertainty and is combined with the MNPT and PT (IQC) uncertainty as shown on page 85:

$$u_{rel}^2 INR = ISI^2 \times [u_{rel}^2 MNPT + u_{rel}^2 IQC]$$

Equation 1: Uncertainty of the International Normalised Ratio.

Taken from ISO/TS 20914:2019, where u_{rel} is the relative uncertainty (of either the INR or the IQC data, $u_{rel} MNPT$ is the standard uncertainty of the Mean Normal Prothrombin Time, ISI = International Sensitivity Index (taken as a constant).

Using Equation 1, provided estimates for the INR of 12.0% and 11.6% with 95% confidence at PT QC levels of 13.2 seconds and 25.3 seconds respectively. This is within the tolerance of 20% for therapeutic INRs using model 1 of the Milan criteria (175). Similar results of 11.6% and 11.1% were found with a different lot of thromboplastin with different ISI (169).

Despite being a useful guide for approaches to MU estimation, this methodology has faced criticism (161). The discussion highlighted the inappropriate application of LPU rules when an exponential function is involved, as is the case with the INR equation. A detailed example is provided in the commentary, using GUM principles to the INR equation and offering sensitivity coefficients for each input variable.

Two alternative methods were identified to simplify MU quantification in this case. The first method assumes that the uncertainty of the ISI and MNPT is small and contributes negligible amounts to overall uncertainty, as described by Taberner and colleagues (176). Due to the complexity of the geometric mean calculation, verifying the uncertainty of the MNPT may be challenging and requires further investigation (165, 166, 177). The second approach is simpler: it involves deriving MU for the INR from PT IQC data using baseline PT results to calculate corresponding INR values and determining the MU from the standard uncertainty of the INR values at the specific QC value.

The assumption that the ISI could be considered as a type B uncertainty based on manufacturers' data, and that this might be insignificant, requires clarification. This was addressed by Lim *et al.* in

2023 (178). The study focused on estimating the MU of the International Sensitivity Index (ISI) for twelve commercially available thromboplastins using a Monte Carlo Simulation (MCS) approach.

The ISI for each thromboplastin was determined via local calibration using orthogonal regression, where PT values from the reference thromboplastin (rTF/16, ISI = 1.0) and the test thromboplastins were logarithmically transformed to derive regression slopes. The ISI was then calculated using the slope and the assigned ISI of the reference standard. Precision profiling was performed to evaluate the relationship between PT values and the coefficient of variation (CV) for each thromboplastin, based on five replicate measurements over five days. These precision models were then used to simulate 100,000 pseudo-PT datasets per thromboplastin in the MCS, yielding robust estimates of ISI values, standard uncertainties, and their relative uncertainty. The study found relative uncertainties in the ISI estimates ranging from 9.7% to 12.1% when using plasma, and 11.6% to 12.0% when using ISI Calibrate (178). Notably, the study found discrepancies between manufacturer-claimed ISI values and those derived from MCS, particularly for reagents such as PT Rec and PT-Fibrinogen. The relative expanded uncertainty in the INR increases with INR value, but remains below 20% within the typical clinical range of 1.0–4.5. These findings support the use of MCS as a powerful tool to improve the accuracy and comparability of INR results by explicitly incorporating ISI-related uncertainties.

The variability in MU across the measurement range is determined by the method of INR calculation. Using a local ISI calibration method, INR MU ranged from 15.8% to 26.4% (for INR values between 1.03 and 5.40). In contrast, direct INR determination reduced MU to 10.9–12.0% across the same range, suggesting that missing uncertainty sources, rather than inflated local ISI calibration, contributed to the discrepancy. Across analytical platforms, ISI assignment accounted for ~50% of total INR MU, while local PT variability contributed more significantly to direct INR determination (178). This is in clear opposition to the previous assumptions of insignificant contribution of the ISI uncertainty to overall INR MU. Notably, EQA data when used yielded significantly lower uncertainty

at the lower INR range (6.3%), though it increased beyond INR 2.8, aligning with bottom-up estimates. EQA data alone however, is not recommended to be used as the determinant of MU.

Table 21: MU considerations in published studies of MU in the INR.

PT – Prothrombin Time, INR – International Normalised Ratio, ISI – International Sensitivity Index, MCS – Monte Carlo Simulation, MU – MU, EQA – External Quality Assurance

Parameter	Details
ISI Calibration	Local calibration using orthogonal regression, PT values logarithmically transformed
Precision Profiling	Relationship between PT values and CV, based on five replicate measurements over five days
Simulation	100,000 pseudo-PT datasets per thromboplastin in MCS
Relative Uncertainties (Plasma)	9.7% to 12.1%
Relative Uncertainties (ISI Calibrate)	11.6% to 12.0%
Discrepancies	Found between manufacturer-claimed ISI values and MCS-derived values
INR Uncertainty	Increases with INR value, below 20% within 1.0–4.5 range
INR MU (Local ISI Calibration)	15.8% to 26.4% (INR 1.03 to 5.40)
INR MU (Direct INR)	10.9% to 12.0% (INR 1.03 to 5.40)
ISI Assignment Contribution	~50% of total INR MU
Local PT Variability	More significant to direct INR determination
EQA Data	Lower uncertainty at lower INR range (6.3%), increased beyond INR 2.8
Recommendation	EQA data alone not recommended as determinant of MU

Furthermore, the method used for INR MU determination, will yield different results, specifically for INR, the study found Top-down MU values ranging from 6.3% to 13.5%, while bottom-up estimates

using ISI calibration ranged from 10.9% to 26.4%, with uncertainty increasing at higher INR values. While both methods show rising uncertainty with increasing INR, the bottom-up approach consistently produced higher values.

2.6 The Top-down method

The Top-down approach has been officially recognised with the publication of ISO/TS 20914:2019 (169). This methodology focuses on identifying sources of variation using data readily available to clinical laboratories, including internal quality control (IQC), external quality assurance (EQA), and manufacturer-supplied calibration and traceability information (152, 156, 179). It does not solely represent IQC performance nor is it defined exclusively by the coefficient of variation (CV) (152). Instead, it encompasses all sources of uncertainty throughout the entire traceability chain. Consequently, reference laboratories, manufacturers, and end users are required to collaborate in establishing the uncertainty across the entire calibration hierarchy (Figure 13) (118, 122, 146, 156, 180-182).

As shown in Figure 12, below, the Top-down method integrates all components of MU into a single model using laboratory-available data. Although sometimes viewed as a convenient alternative, the Top-down approach has demonstrated comparable performance to the bottom-up method in other disciplines. For example, Lee *et al.* reported no significant difference in glucose MU estimates between the two methods—0.18 mmol/L (bottom-up) vs. 0.17 mmol/L (Top-down) at low concentrations, and 0.34 mmol/L vs. 0.36 mmol/L at higher concentrations (183).

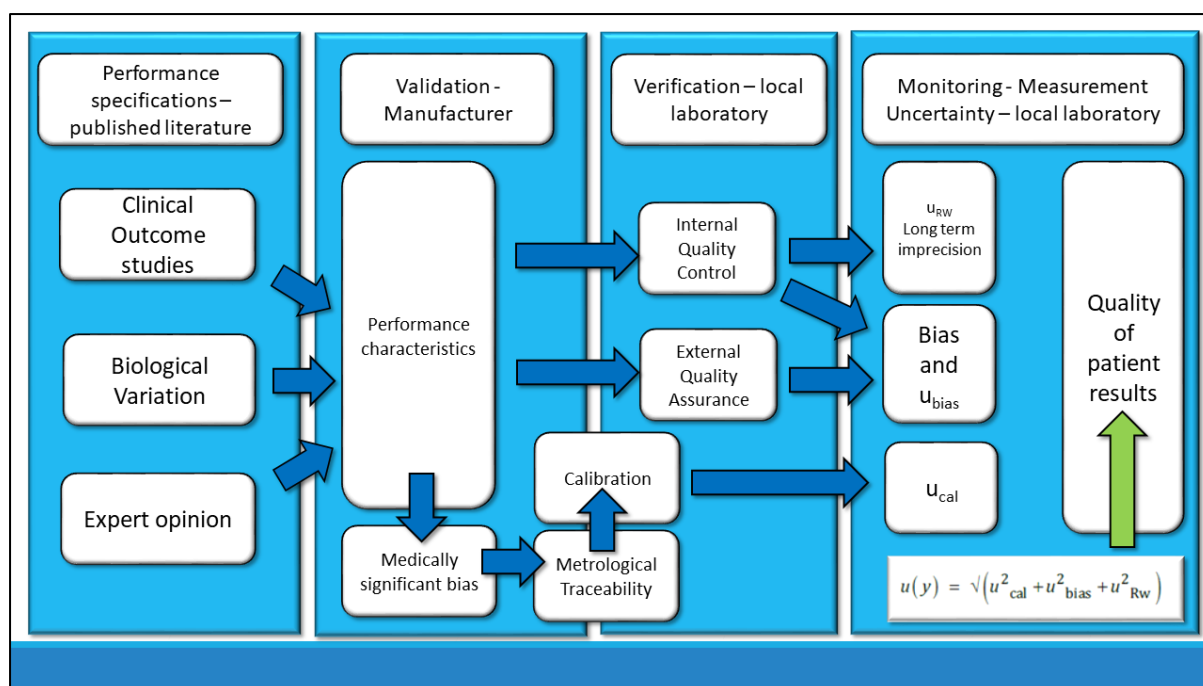


Figure 12: Linking performance specifications to measurement uncertainty.

Schematic overview of the top-down method of MU estimation as outlined in ISO/TS 20914:2019. The figure shows how uncertainty sources such as calibrator value uncertainty (u_{cal}), bias correction uncertainty (u_{bias}), and long-term imprecision (u_{RW}) are combined using the root-sum-of-squares method to generate the combined uncertainty for the final patient result ($u(y)$).

The ISO technical specification document promotes a Top-down approach to MU estimation, prioritising feasibility over mathematical complexity. It combines IQC data with manufacturer-supplied information on calibrator uncertainty. The main components are, long-term imprecision (u_{RW}), uncertainty of calibrator-assigned values (u_{cal}) and bias correction uncertainty (u_{bias}), if applicable.

These are treated as standard uncertainties and combined using the root-sum-of-squares method (Equation 2). The resulting combined standard uncertainty is expanded using a coverage factor (typically $k = 2$) to reflect a 95% confidence interval (ISO, 2019).

$$u(y) = \sqrt{u_{cal}^2 + u_{RW}^2 + u_{bias}^2}$$

Equation 2: Calculation of MU according to guidance provided in ISO/TS 20914:2019.

u(y) = uncertainty in the final patient result, u_{cal} = uncertainty of calibration (to include uncertainty of the reference method/material as well as manufacturers assigned values), u_{bias} = uncertainty of bias when present and corrected for (if applicable), u_{RW} – local laboratory variation.

2.6.1 Intermediate imprecision

An early comparison of MU methods by Favaloro in 2006 showed that assay precision varies across the measurement range (184). This pattern was also observed by Lim *et al.* in assays for fibrinogen, antithrombin, and protein C, where expanded uncertainty increased with concentration (143).

These studies highlight the challenges of consistently applying a MU approach to haemostasis assays, particularly where non-linear and reagent-dependent behaviours are observed. The ISO/TS 20914:2019 standard provides strategies to address these challenges, including evaluating imprecision across multiple lot numbers, different analyser platforms, and within laboratory networks (161). These variations are accounted for in the reproducibility component of MU and must be carefully considered during method implementation and ongoing monitoring. For practical purposes, the standard deviation of IQC performed over an extended period (generally six months), assuming normal distribution of data, can provide u_{RW} as a standard uncertainty derived from the calculated standard deviation of the IQC at each QC level (169).

2.6.2 Calibration uncertainty and Metrological Traceability

Traceability for haemostasis assays involves complex procedures, in part due to the difficulty in fully characterising MU throughout the calibration hierarchy. Plasma international standards, when available, support traceability and facilitate comparability between laboratories and manufacturer calibrations (146, 148, 180, 185). These standards form the basis of the calibration hierarchy, and

their associated uncertainties contribute to reference uncertainty (u_{ref}) and calibration uncertainty (u_{cal}) values in the MU model. When standards are assigned via consensus values, the resulting uncertainty is propagated through manufacturer calibrations and termed u_{ref} . In the Top-down MU model described by ISO/TS 20914:2019, the u_{cal} component includes both this reference uncertainty and the calibration uncertainty from manufacturers as u_{cal} .

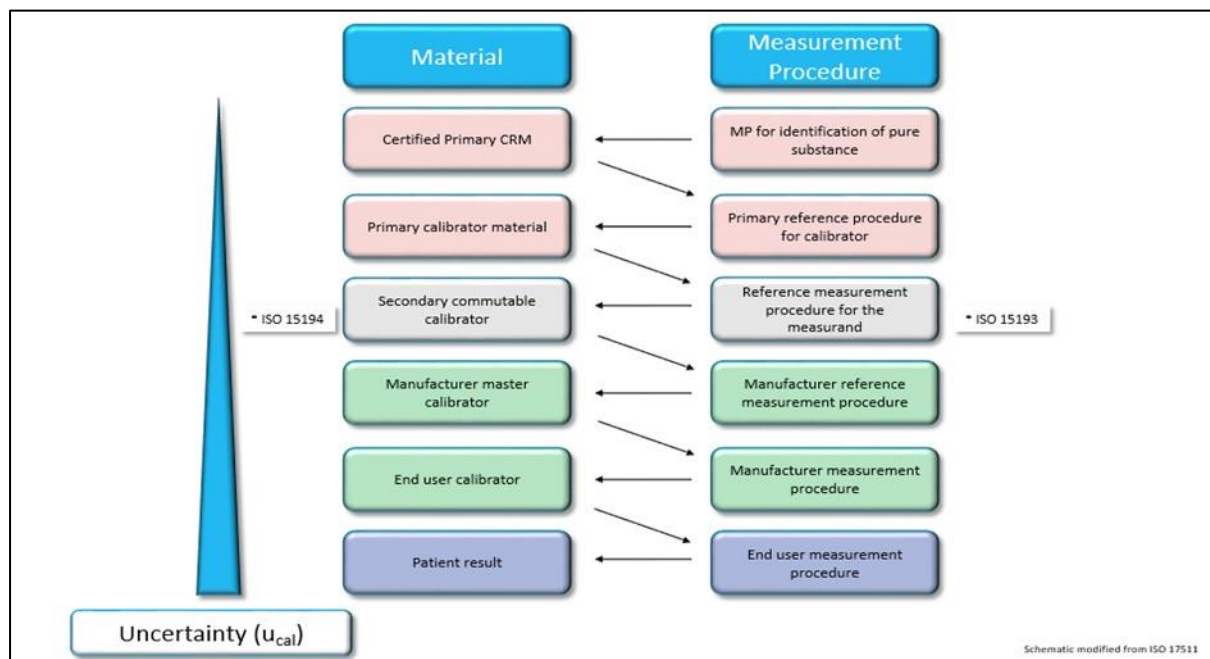


Figure 13: Measurement uncertainty and the calibration hierarchy.

Calibration hierarchy from reference measurement procedures and certified reference materials to uncertainty propagated to final patient results. Image adapted from ISO 17511:2020, produced by S MacDonald.

Each step taken away from the top of the calibration hierarchy introduces increased uncertainty. The assignment of values in each successive step is influenced by the process employed, whether it be from international standard providers or calibrator manufacturers. Additionally, calibrations of various measurement systems add further uncertainty along the traceability chain (Figure 13). By the time a calibrator, with its assigned uncertainty, is used to calibrate an end-user measurement

procedure, this accumulated uncertainty from the traceability chain becomes significant and directly impacts patient results.

Understanding the uncertainty associated with end-user calibrators is crucial for providing the u_{cal} estimate in

Equation 2. Without knowledge of the uncertainty of the assigned calibration value, MU cannot be accurately determined. This requires knowledge of the uncertainty of the assigned value for reference materials (u_{ref}) according to international standards, which supports manufacturers in providing this information. Stability and homogeneity assessments are commonly reported in international standards and often determine the uncertainty of assigned values. Zhang *et al.* evaluated Factor VIII and Factor IX reference standards, identifying long-term stability uncertainty (u_{lts}) of 0.5–5.1% for Factor VIII and 1.3–4.4% for Factor IX using a one-stage assay (186).

Characterisation uncertainties (u_{char}) ranged from 0.9–2.4% for Factor VIII and 1.1–2.4% for Factor IX. Combined uncertainties across multiple lots showed significant variation, with uncertainties for Factor VIII at (85±13)%, (36.0±3.4)%, and (20.5±2.3%), and for Factor IX at (102±13)%, (47.8±6.9)%, and (29.3±3.8%). The magnitude of these uncertainties presents a challenge for end users, particularly when performance specifications are designed based on clinical utility. Additionally, local performance of calibrations, whether by serial or direct calibration, and the number of calibrator measurements made (187), can influence uncertainty and the movement from stored to real time calibrations all may impact MU.

2.6.3 Bias and Long-Term Performance from EQA Data

The final term in the ISO procedure is u_{bias} . It refers to the standard uncertainty associated with any correction applied for systematic error, rather than quantifying bias itself as in the total error approach. This term accounts for uncertainty in determining that bias and must be included in the combined MU estimate if bias is corrected (151, 154, 188).

Estimating u_{bias} typically involves:

Identifying a reference value (e.g. CRM, international standard, or peer group mean).

Calculating the difference between observed and reference values across replicates.

Determining the mean bias and its associated standard uncertainty (e.g., standard deviation of bias across runs).

2.6.3.1 Including the standard uncertainty of the bias estimate in the overall MU model as u_{bias} .

Meijer *et al.* proposed a linear regression-based method to evaluate long-term performance in antithrombin assays utilising only external quality assurance (EQA) data (189). This method is aligned with the total error model by accounting for both random and systematic error. Qin *et al.* expanded on this approach within a cohort of 142 laboratories participating in the Beijing Centre for Clinical Laboratories (BCCL) EQA scheme (190). Their analysis indicated significant variation in MU related to the manufacturer of the coagulometer system.

Nevertheless, there are limitations in estimating MU based solely on EQA data. Qin *et al.* cautioned that such reliance might lead to an overestimation of MU unless appropriately managed (190). Lim *et al.* corroborated these findings by comparing EQA-derived estimates with those obtained using international standards where MU values were consistently lower (143). This indicates that while EQA provides a practical source of data, it may exaggerate uncertainty due to variability in consensus values and the absence of direct traceability. However, it can identify biases that require correction, thereby introducing the bias component into

Equation 2.

2.6.3.2 Bias Correction in Haemostasis Assays

Even when bias appears to be zero, statistical uncertainty remains due to the spread of values around the target. Bias is considered statistically significant if its absolute value exceeds twice the standard uncertainty associated with its estimation ($|\text{bias}| > 2 \times u_{\text{bias}}$) (151, 154, 188). Only under these

circumstances should a correction be considered, and the associated uncertainty (u_{bias}) should be propagated into the MU estimate. This threshold corresponds to a 95% confidence level and ensures that the observed bias is unlikely to result from random variation alone. When it is applied, the correction should only be used if the bias has been confirmed through a validated study using an international standard. Bias identified through External Quality Assessment (EQA) or IQC peer comparisons must first be verified using a traceable reference method before any correction is applied to patient results.

Although rare in haemostasis, one well-documented exception is the B-domain truncated recombinant Factor VIII product, Afstyl^a (CSL Behring, Germany). Both manufacturer recommendations and post-implementation field studies advise applying a correction factor of 2 when testing Afstyl^a using a one-stage clotting assay (191). This adjustment accounts for known assay-specific underestimation of activity. If correction is applied, the uncertainty of the correction factor must be incorporated into the total MU estimate. Although other EHL Factor VIII products also show discrepancies between one-stage and chromogenic assays, formal bias correction is generally not recommended due to variability between reagents and platforms (192).

2.7 The Top-down method in haemostasis assays

As early as 2006, Favaloro compared two models assessing random variation with or without systematic error; however, no consensus was reached regarding the most appropriate method (184). It was not until the publication of ISO/TS 20914:2019 that a structured framework was introduced. Additional methods, including the Nordtest framework, have also been employed to quantify uncertainty in haemostasis assays. There are in fact, many different formulas that could be employed (193).

Using long-term IQC data and comparisons of assays to international standards or EQA performance, Lim *et al.* studied MU in common specialist haemostasis assays employing the Nordtest method

(143). They observed a concentration-dependent increase in relative expanded uncertainty for INR, fibrinogen, antithrombin, and protein C, while APTT and Protein S exhibited a decrease in uncertainty across their respective measurement ranges.

A direct comparison of Factors V, VIII, and X revealed method-dependent differences in expanded MU between ACLTOP and STA Compact. Imprecision and bias were higher using the STA Compact than the ACLTOP 750 CTS (143). The lowest expanded uncertainties were reported for Factor X (3.3% ACLTOP, 6.7% STA Compact), followed by Factor VIII (7.7% ACLTOP, 11.9% STA Compact) and Factor V (5.7% ACLTOP, 11.5% STA Compact). Notably, all factor assays demonstrated significant bias compared to the CRM, except for ACLTOP Factor VIII.

Table 22: Comparative Overview of MU Approaches in Haemostasis Assays.

The table summarises various methods for characterising and estimating MU, including their theoretical foundations, practical applications, advantages, and limitations. Primary sources of variation, calibration uncertainties, and empirical constants are discussed, referencing specific standards and studies.

Approach	Description	Strengths	Weaknesses	Key Components	References
Bottom-Up	Formal metrological process, defines mathematical model and estimates input contributions	Theoretically complete characterisation, treatment of derived quantities	Mathematical complexity, limited uptake in clinical labs	Partial derivatives, sensitivity coefficients, empirical constants	JCGM:100, Frenkel and Farrance, 2018, Rigo-Bonnin and Canalias, 2020
Top-down	Captures routine sources of variation using readily available data	Feasibility, comparable performance to bottom-up	Less rigorous than bottom-up	IQC, EQA, calibrator uncertainty, bias correction	ISO/TS 20914:2019, Lee <i>et al.</i> , 2015

For thrombophilia assays, direct platform comparisons were not performed, but MU estimates were 10.4% (Antithrombin, ACLTOP), 7.1% (Protein C), and 14.5% (Protein S, STA Compact), with significant bias in Protein C and Protein S (143).

2.8 Using Monte Carlo Simulation for haemostasis MU

Due to the specialised nature of many haemostasis assays, laboratories often face challenges with low sample volumes. In situations where IQC is performed prior to batched analysis, data may be insufficient to provide a robust characterisation of intermediate imprecision using the Top-down method. This poses a challenge to the application of ISO/TS 20914:2019, as realistic MU estimates may not be achievable using the Top-down framework alone (144). Additionally, a recognised limitation of the Top-down approach in factor assays—particularly those based on activated partial thromboplastin time (APTT) for factors VIII, IX, XI, and XII—is that IQC may not fully capture imprecision in the final patient result when assays rely on single dilution points. The determination of parallelism, and the final result reported may also be different depending on the platform used, and local preferences (187). While traditional protocols typically involve multiple dilutions with extrapolated results, the necessity of such practices has been debated (194).

Lim *et al.* demonstrated that MCS could accurately characterize MU in calibrated assays for FV, FVIII, and FX, offering values closely aligned with Top-down estimates (150). Similarly, Cui *et al.* employed a bootstrap simulation method to determine MU for platelet counts, successfully integrating it with the Nordtest Top-down framework, thereby further validating simulation-based approaches in haematological settings (195). Guidance on performing the calculations in Microsoft Excel[®] have been published (196). To align with the JCGM document describing use of Monte Carlo methods (197).

In Lim's 2021 study, precision profiling and MCS were utilised to model the complex calibration processes in modern automated coagulation testing (150). Two automated coagulation systems—

ACL TOP 750 CTS and STA Compact—were used for factor assays, with MU estimates generated through MCS by simulating 10,000 calibration curves using quality control data and applying random variability in clotting times to model assay performance across the measurement range.

The factor activities were simulated at normal levels comparable to measured IQC levels, yielding simulated MUs of 3.5% (FV), 9.3% (VIII), and 2.8% (FX) for ACL TOP, and 8.5% (FV), 18.2% (FVII), and 6.5% (FX) for STA Compact. Results demonstrated that MCS and Top-down methods produced broadly comparable MU estimates, except for FVIII on the STA Compact system, where higher variability was observed at low factor activity levels. MU increased significantly at lower concentrations, particularly for FVIII, sometimes exceeding published limits, indicating potential interpretative limitations in this range.

Clearly, the calibration method impacts MU derived in this manner, and the different number of calibration points between analytical platforms may provide insight into observed differences, warranting consideration in future studies. More significant is the choice of models inherent to the test that restricts the Monte Carlo modelling approach. The difference between linear and polynomial regression will affect curve fitting, and although more complex curves may better fit the data, it is uncertain whether this characteristic is advantageous for determining MU. A recognized limitation of this study is the lack of uncertainty information in the calibrator assigned value, which hampers the application of the ISO/TS 20914 method.

2.9 Discussion

The Top-down method has provided a structured framework for MU estimation, simplifying the approach for clinical laboratories. However, much of the published literature has applied a uniform methodology across entire disciplines, treating all assays within a field identically. While this facilitates comparability in publications, it introduces systematic bias, as individual assays may require distinct approaches to ensure accurate uncertainty characterization.

This issue is particularly evident in INR uncertainty determination, where multiple methodologies with varying levels of complexity have been applied (143, 161, 166, 198, 199). The ISO/TS 20914:2019 document itself has been challenged for its approach to INR MU determination (161). Grouping assays by sample turnover or workload volume often leads to oversimplification, neglecting the unique complexities of each test.

As more MU estimation methods become available through ISO, CLSI, and JCGM, a tailored approach—selecting the most appropriate method based on the specific uncertainty contributors of each assay—is preferable to a one-size-fits-all framework. Lim et al illustrated this in Figure 1 (143), where the choice between Top-down and Bottom-Up approaches is delayed until significant uncertainty sources are identified and characterised. The two approaches are not mutually exclusive; in fact, simulation-based tools such as Monte Carlo modelling can complement the Top-down method, especially when evaluating complex or derived measurands.

Many haemostasis assays yield categorical results (e.g., positive vs. negative or detected vs. not detected). Although these are not traditionally quantitative, the underlying numerical values determine categorization and are therefore the point in the measurement range where MU has its greatest impact. Examples of such assays include Factor Inhibitor screens, mixing studies, or Heparin-Induced Thrombocytopenia (HIT) screening assays. For results near the cut-off, MU must be carefully considered, as both the measured result and the cut-off itself carry inherent uncertainty.

Future publications should focus on refining these methodologies rather than seeking a simplistic, generalized approach. Whether increased research into MU will drive the adoption of more tailored uncertainty estimation methods remains to be seen.

2.9.1 Defining Performance Specifications in Haemostasis

Performance specifications for MU must align with the clinical use intended for an assay. ISO 15189:2022 (Section 7.2.4) underscores that MU should be evaluated and maintained in accordance

with its clinical application. Diagnostic assays typically emphasize accuracy and bias control; for instance, Factor VIII activity measurements must reliably differentiate the severity of haemophilia. Conversely, monitoring assays, such as INR, require consistent precision to ensure safe treatment adjustments. Unlike clinical chemistry, haemostasis lacks consensus on acceptable MU targets due to high biological variability, diverse methodologies, and limited outcome-based data, making it challenging to define universal performance expectations.

In longitudinal patient monitoring, it is crucial to distinguish true clinical changes from analytical variation. MU alone does not capture expected physiological variability; it is complemented by biological variation, which accounts for within- and between-subject variability. This combined approach helps determine whether changes in serial results are clinically meaningful or fall within expected variability.

Key applications include monitoring oral anticoagulant therapy, unfractionated heparin, replacement therapy pharmacokinetics, and disease progression in coagulopathies. In some models, BV can be treated as a Type B uncertainty within the MU budget.

Despite its value, BV data in haemostasis remain limited. Random variation and assay complexity—especially in pharmacokinetic contexts—further complicate interpretation. Unlike clinical biochemistry, few robust BV-derived performance specifications exist for haemostasis, limiting their generalisability across assays and patient populations.

2.10 Conclusions

MU is crucial for the interpretation and clinical application of haemostasis assays. Its implementation, however, is complicated by the complexity of test design, the limited availability of reference materials, and the lack of assay-specific performance specifications. ISO/TS 20914:2019 offers a framework for routine MU application, but its top-down approach requires careful adaptation for specialist assays where traceability and calibration pathways are less defined.

Evidence from INR and other coagulation tests suggests that both bottom-up and simulation-based methods are important, particularly in addressing derived quantities and environments with limited data. Laboratories need to consider assay-specific sources of variability instead of adopting a uniform approach to MU estimation. Additional research, interdisciplinary collaboration, and clearer guidance from international bodies are necessary to support standardisation and develop robust APS in haemostasis.

Rationale for the Thesis

Limitations of Current Approaches to MU

The concept of MU is increasingly recognised as essential in medical laboratories, and guidance such as ISO/TS 20914:2019 and CLSI EP29-A has provided structured, top-down approaches for its estimation. These methods typically use IQC and manufacturer-provided calibration data to estimate uncertainty without the need for complex modelling.

However, these approaches rely on several assumptions that may not hold true in practice. For example, IQC data are usually treated as if they only reflect random variation. This overlooks the potential influence of short-term systematic effects, which may be present within the same dataset. Additionally, there is little consideration of how IQC results may be correlated over time, and these patterns are rarely accounted for when calculating uncertainty. These issues raise questions about the reliability of current MU estimates, especially in tests that are clinically sensitive and analytically complex.

Gaps in APS in Haemostasis

In parallel with limitations in MU estimation, the definition of APS in haemostasis remains underdeveloped. While the Milan consensus provides three models for deriving APS—based on clinical outcomes, biological variation, or state-of-the-art performance—these are rarely applied in practice for coagulation assays. For many haemostasis tests, particularly specialist ones, there is a lack of available biological variation data, and clinical outcome studies are scarce.

ADAMTS13 activity assays, which are essential for diagnosing TTP, are a clear example of this gap. Despite their clinical importance, there is no consistent method for assigning APS or linking performance expectations to medical decision limits. Without appropriate APS, uncertainty estimates have limited value in guiding assay verification or result interpretation.

Why ADAMTS13 Activity is an Appropriate Exemplar

This thesis uses ADAMTS13 activity assays as a detailed case study for exploring issues related to MU, APS, and assay verification in haemostasis. These assays are highly relevant for several reasons. First, their results directly influence urgent clinical decisions in the diagnosis and management of TTP. Second, multiple assay methods are available, including newer rapid systems that are increasingly used in emergency care. Third, these assays are often performed at relatively low volumes and have complex analytical characteristics, making them representative of many challenges seen across specialist haemostasis testing.

Because of these features, ADAMTS13 assays provide a suitable framework for evaluating the limitations of current uncertainty estimation methods and for testing new approaches to performance specification and verification.

Scope and Direction of the Thesis

This thesis addresses key challenges in how MU is estimated, reported, and used in haemostasis laboratories. It critically examines existing top-down methods, explores the application of APS, and proposes new approaches for integrating MU into result interpretation.

To support this, the thesis introduces methodological tools such as statistical simulation, calibration curve modelling, and likelihood ratio-based interpretation. It also considers how uncertainty can be contextualised for individual patients through the use of personalised reference intervals, particularly in long-term monitoring of TTP.

Statement of Originality and Contribution to Knowledge

This thesis makes original contributions to the application and interpretation of MU in medical laboratory science, particularly within specialist haemostasis testing. It introduces:

A novel application of time series analysis to IQC data for the purpose of distinguishing random from systematic sources of variability—an approach not previously applied in the context of MU estimation.

The development of simulation-based methods for modelling calibration curve behaviour and assessing MU across the full analytical range, rather than at fixed quality control levels.

An expanded framework for integrating MU into clinical result interpretation, including the use of likelihood ratios to contextualise uncertainty and the implementation of patient-specific reference intervals for longitudinal monitoring.

These contributions are applied to ADAMTS13 activity assays as an exemplar, but the methodologies proposed are intended to be transferable to other specialist tests where MU is not currently well understood or applied.

To the best of the author's knowledge, these approaches represent new applications and methodological developments within the fields of laboratory haemostasis and clinical quality assurance.

Thesis Structure

This thesis is organised into a series of chapters that build progressively from foundational principles to applied methodologies and clinical interpretation, using ADAMTS13 activity assays as a consistent exemplar throughout.

Chapter 1 introduces the diagnosis of TTP, analytical performance and MU in the context of ADAMTS13 testing and outlines its clinical and regulatory relevance.

Chapter 2 presents a comprehensive literature review, drawing on a published invited review article, and summarises current understanding, guidance, and practice related to MU in haemostasis.

The rationale for the research is presented, identifying limitations in current MU estimation methods, gaps in APS, and the need for improved clinical interpretability.

Thesis aims presents the intended investigations and hypotheses associated with the thesis

Chapter 3 outlines the methodological approaches used to address them.

Chapter 4 applies different models for deriving APS, including outcome-based, biological variation, and state-of-the-art approaches, specifically for ADAMTS13 testing.

Chapter 5 critically evaluates current verification practices in the context of MU

Chapter 6 evaluates the application of top-down MU methods, examining their assumptions and limitations when applied to ADAMTS13 assays and critiquing the standard approaches in preliminary verifications and in duplicate testing

Chapter 7 introduces time series analysis of IQC data to better understand the components of assay variability and assess the implications for MU estimation.

Chapter 8 presents a simulation-based approach to modelling calibration curve behaviour and estimating uncertainty across the analytical range and the impact of master calibrations on the understanding of bias correction.

Chapter 9 explores the clinical interpretation of results, proposing a framework for integrating MU with likelihood ratios and development of a novel analytical performance specification directly linked to clinical utility to support decision-making in TTP monitoring.

The final chapter (Chapter 10) synthesises the key findings, discusses their implications for laboratory practice and clinical care, and outlines opportunities for further research.

Thesis Aims

The assessment of ADAMTS13 activity plays a central role in the diagnosis and management of TTP and other related thrombotic or autoimmune disorders. However, significant variability exists in how these assays are designed, standardised, and interpreted across laboratories and clinical settings.

This thesis explores the analytical and clinical implications of this variability, with a focus on the limitations of current methods for defining performance specifications and calculating measurement uncertainty. This work aims to improve how assay performance is understood, quantified, and communicated. The following objectives define the specific aims of the thesis:

1. To evaluate current harmonisation challenges in ADAMTS13 activity testing by identifying key sources of variation between assays. This includes highlighting inconsistencies in measurand definition, calibration, and result interpretation that may impact downstream studies of analytical and clinical performance.
2. To critically assess existing approaches to analytical performance specification for ADAMTS13 activity assays and to question the validity of applying uniform performance expectations across different clinical applications. The aim is to demonstrate that performance requirements should reflect the specific clinical use case, and that in some situations, result-level interpretation may benefit from personalised performance thresholds.
3. To investigate how measurement uncertainty impacts the assessment of clinical and analytical agreement in method comparison studies, particularly when assays are used for diverse or evolving clinical indications. By integrating measurement uncertainty into these studies through the use of error grids, the goal is to provide a more clinically meaningful and analytically sound framework for evaluating assay performance.

4. To challenge current standard approaches to measurement uncertainty estimation, with a particular focus on the limitations of top-down methods. The aim is to explore time series analysis of internal quality control data as a novel approach to understanding the structure of assay variability and to assess its use in quantifying uncertainty more accurately and informatively.
5. To introduce and validate a new simulation-based framework for calculating measurement uncertainty and setting performance specifications in one-at-a-time ELISA methods, using Monte Carlo methods. This section aims to support practical and tailored uncertainty estimates in a format that reflects the specific nature of ELISA testing workflows.
6. To examine the contribution of bias arising from master calibration and local verification processes, and to evaluate whether current correction methods effectively reduce clinically meaningful bias. This investigation seeks to improve understanding of how bias influences assay performance and uncertainty calculations.
7. To apply the findings across previous sections to a clinically focused framework, including the development of novel tools for communicating uncertainty to clinicians. This includes investigating whether likelihood ratios can be used to measure acceptable clinical performance, enhancing the clinical utility and safety of ADAMTS13 activity testing.

3 Thesis Methods

3.1 Introduction to Methods

This chapter outlines the general laboratory procedures, data sources, and statistical methodologies that underpin the work presented in this thesis. The studies contained within this thesis explore various aspects of laboratory method performance, MU, and the clinical interpretation of ADAMTS13 activity testing in the context of TTP.

The methods described here are common to multiple chapters and provide the technical framework that supports the experimental work, simulations, and data analyses performed. This includes detailed descriptions of sample selection, laboratory processes, data governance, and statistical approaches to MU, method comparison, time series analysis, and simulation-based modelling.

Study-specific methodologies, including adaptations or additional approaches required to answer specific research questions, are described in the respective results chapters where appropriate.

Where novel methodological frameworks were developed—for example, in the application of MU to the development of an ADAMTS13-specific error grid—these are introduced in this chapter with further detailed expansion in the relevant chapters.

This approach avoids unnecessary duplication while ensuring transparency and reproducibility in how the data were generated, analysed, and interpreted.

3.2 Laboratory Processes

3.2.1 Sample Sources and Selection

All samples used in this thesis were surplus to routine clinical care and were collected exclusively from the Specialist Haemostasis Laboratory at Cambridge University Hospitals NHS Foundation Trust (CUHNHSFT). No prospective sampling was undertaken. Samples were retained following completion of diagnostic testing and were otherwise destined for disposal. This approach was applied uniformly across all aspects of the thesis, including studies related to ADAMTS13 activity assays, method comparisons, IQC evaluations, and time series analysis of routine coagulation assays.

Sample selection for each study was determined by the specific research objective. For studies evaluating ADAMTS13 activity assays, including both initial method verification and subsequent post-implementation comparisons, samples were selected either from cases under investigation for TMA or from patients with a confirmed diagnosis of TTP. In the initial method verification phases, samples were intentionally selected to represent a range of ADAMTS13 activity levels, particularly focusing on clinically relevant thresholds for diagnosis and remission monitoring. In contrast, the post-implementation comparison phase employed an unselected, consecutive series of samples referred for TMA investigation.

Samples with insufficient plasma volume were excluded from further testing, as were those affected by pre-analytical issues such as clotting, haemolysis, or contamination where identified.

For all studies involving routine coagulation assays, including analyses of IQC performance, time series decomposition, and third-party versus manufacturer-provided quality control materials, data were derived from testing performed within the same Specialist Haemostasis Laboratory with additional data from Diagnostic Haematology and Blood Sciences departments.

Ethical approval for the use of historical patient data and surplus clinical samples was obtained from the Research and Development (R&D) departments of both Cambridge University Hospitals NHS

Foundation Trust and the University of Hertfordshire. Formal patient consent was waived in accordance with institutional policies, given that all samples were anonymised and surplus to diagnostic requirements. No additional blood samples were collected specifically for the purposes of this research.

3.2.2 Sample Preparation and Storage

All sample processing was performed in accordance with the Specialist Haemostasis Laboratory's standard operating procedures (SOPs) in accordance with national guidance (200), and ISO 15189:2022 requirements. Platelet-poor plasma (PPP) was prepared by centrifugation at 4,000 rpm (equivalent to approximately $2,860 \times g$) for 15 minutes at room temperature using a Thermo Scientific Megafuge 8 centrifuge (Thermo Fisher Scientific, Horsham, UK). Double spinning was implemented as necessary according to the above guidance.

Following centrifugation, plasma was carefully separated from the cellular components using sterile Pasteur pipettes (Alpha Laboratories, Eastleigh, UK) and aliquoted into 2 mL cryovials (Eppendorf, Stevenage, UK). Samples were immediately frozen and stored at -80°C (PhCBI Twin guard, PHC, Birkenhead, UK).

All storage was continuously temperature-monitored using the Hanwell Environmental Monitoring System (EMS) (Ellab Monitoring Solutions, Ellesmere Port, Cheshire, UK). No storage failures or temperature excursions occurred during the period covered by this research.

For urgent ADAMTS13 activity requests, samples were processed and tested immediately following centrifugation without freezing. For all other applications—including method comparison studies, calibration modelling, simulation studies, and MU evaluations—samples underwent a single freeze-thaw cycle. Thawing was performed at 37°C for 10 minutes using a calibrated, temperature-monitored heat block (Labnet International Inc., Edison, USA). Following analysis, all thawed samples were discarded and were not refrozen.

These procedures ensured that pre-analytical variability was minimised and aligned with best practices for haemostasis laboratory testing.

3.3 Assay Platforms and Data Sources

This thesis incorporates data derived from a combination of ADAMTS13 activity assays, routine haemostasis tests, IQC data, and external quality assurance (EQA) performance. All data were generated within the Specialist Haemostasis Laboratory at Cambridge University Hospitals NHS Foundation Trust and the Diagnostic Haematology lab at the same institution.

3.3.1 ADAMTS13 Activity Assays

Three analytical platforms for ADAMTS13 activity measurement were used during the study period:

Technozym® ADAMTS13 ELISA (Technoclone, Vienna, Austria), performed using either the FluoScan Ascent (Alpha Laboratories, Eastleigh, UK) or the FLUOstar Omega (BMG LabTech, Aylesbury, UK) microplate readers. The ELISA method was in routine use from 23rd September 2009 to the time of writing.

HemosIL AcuStar® ADAMTS13 Activity (Werfen, Bedford, USA), run on the fully automated AcuStar® analyser. Data from this method were collected from 1st September 2019 to the time of writing.

Technofluor® ADAMTS13 Activity (Technoclone, Vienna, Austria), performed on the Ceveron® s100 analyser. This method was evaluated over a shorter period, with data collected between 1st June 2020 and 1st September 2020, for comparative method validation purposes.

For each platform, comprehensive datasets were extracted, including sample results, IQC data, and calibration data where applicable. This allowed assessment of assay performance, method comparison studies, and simulation modelling incorporating analytical variability and MU.

3.3.2 Routine Haemostasis Assays

For analyses related to IQC variability, time series decomposition, and comparisons between manufacturer-supplied and third-party IQC materials, data were extracted from routine haemostasis assays performed exclusively on ACL-TOP 750 CTS analysers (Werfen, Bedford, USA). This included assays for:

Activated Partial Thromboplastin Time (APTT)

Prothrombin Time (PT)

QFA-Fibrinogen (Clauss method)

Thrombin Time (TT)

Quality control materials were supplied by:

Werfen (manufacturer-provided IQC)

Pathways Diagnostics (Pathqas, third-party IQC)

3.3.3 IQC Data

IQC data were collated for all ADAMTS13 assays and routine coagulation assays described above. This included both manufacturer-provided and third-party QC materials, where applicable. Data were used for analyses of measurement precision, time series modelling, and the derivation of within-laboratory uncertainty estimates.

3.3.4 External Quality Assurance (EQA) Data

EQA data were obtained from two external providers:

- United Kingdom National External Quality Assurance Scheme (UKNEQAS) for Blood Coagulation and Haematology (Sheffield, UK).

- External quality Control of diagnostic Assays and Tests with a focus on Thrombosis and Haemostasis (ECAT Foundation, The Netherlands).

Specifically, EQA data from all ADAMTS13 activity scheme submissions between 26th January 2015 and 25th September 2023 were collated for bias estimation, method performance evaluation, and incorporation into MU modelling.

3.3.5 Data Types and Utilisation

The following data types were utilised across the various analyses presented in this thesis:

- Patient result data: used for method comparisons, calibration evaluation, uncertainty modelling, and clinical risk simulations, de identified following extraction, except where indicated in specific sections, where hospital number only was used for data stratification and modelling after recoding for deidentification.
- IQC data: used for precision analysis, time series decomposition, and derivation of within-laboratory standard uncertainty.
- Calibration data: extracted from consecutive ELISA assay runs and consecutive AcuStar[®] lot numbers as indicated in the relevant sections to inform curve fitting, variability in calibration parameters, and uncertainty propagation.
- External Quality Assurance (EQA) data: used for assessment of inter-laboratory bias and as an input to MU calculations.

All data were generated within the same laboratory environment, with consistent quality control and calibration procedures applied throughout the study periods described.

3.4 Data Management and Governance

All data used in this thesis were managed in accordance with the data protection and information governance requirements of Cambridge University Hospitals NHS Foundation Trust (CUHNHSFT), the University of Hertfordshire, and applicable legislation, including the UK Data Protection Act (2018) and the General Data Protection Regulation (GDPR). Data management practices were also compliant with the laboratory's obligations under ISO 15189:2022.

3.4.1 Data Storage and Security

Following extraction from laboratory information systems, analyser platforms, and quality control databases, all data were initially stored on a secure, access-restricted network server within the CUHNHSFT infrastructure. Access to this storage location was limited exclusively to the author and governed by the Trust's Information Technology and Information Governance policies.

3.4.2 Anonymisation

All patient-identifiable information was removed prior to any data analysis. This included removal of patient names, hospital numbers, dates of birth, and any other direct identifiers. Anonymisation was performed manually at the point of data extraction or immediately following, prior to transfer to any statistical or analytical environments. Only anonymised datasets were used for statistical modelling, quality assessment, method comparison, and simulation studies.

3.4.3 Data Transfer and Analysis Environments

Data analysis was conducted using a combination of secure environments:

- CUHNHSFT network drive: for initial data extraction, pre-processing, and storage of raw files.
- A personal, encrypted laptop: used for statistical analysis performed in R.
- Cloud-based environments (Google Colab): utilised exclusively for Python-based simulation studies. Data transfer to cloud environments occurred only after full anonymisation.

At no point were identifiable patient data stored or processed outside the secure CUHNHSFT network.

3.4.4 Compliance

All data handling was undertaken in full compliance with:

- CUHNHSFT Information Governance policies.
- University of Hertfordshire Research Governance policies.
- ISO 15189:2022 standards relating to data integrity and security.
- The UK Data Protection Act (2018) and the General Data Protection Regulation (GDPR).

Ethical approvals obtained for this research, as described previously, incorporated explicit consideration of data handling, anonymisation, and retention.

3.4.5 Data Retention

In accordance with institutional policies and the conditions of ethical approval, anonymised datasets will be retained securely following completion of this thesis to support future research and ongoing quality improvement projects within the Specialist Haemostasis Laboratory.

3.5 General Statistical Methods

A range of statistical methods was employed across the studies presented in this thesis to support analysis of method performance, MU, IQC performance, and method comparison. The selection of statistical methods was driven by the nature of the data, the analytical question, and the availability of appropriate models in line with international guidance, including ISO/TS 20914:2019, the GUM (JCGM 100:2008), and relevant Clinical and Laboratory Standards Institute (CLSI) guidelines.

All statistical analysis was undertaken using one or more of the following environments:

- Microsoft Excel® (Microsoft Corporation, Redmond, WA) – for data pre-processing, descriptive statistics, preliminary uncertainty calculations, and Monte Carlo simulations for selected analyses.
- R statistical programming environment (R Core Team, Vienna, Austria) – for regression modelling, time series decomposition, resampling methods, uncertainty propagation, and diagnostic performance evaluation. Specific packages used are detailed in the relevant methodological subsections.
- Python (Google Colab platform) – used exclusively for simulation-based assessment of MU in correlation models and likelihood ratio analysis. Data used in this environment were fully anonymised prior to transfer.

3.5.1 Descriptive Statistics and Normality Testing

Descriptive statistics were calculated to summarise central tendency (mean, median) and variability (standard deviation, interquartile range) for all datasets. Assessment of data distribution was performed using a combination of:

- Shapiro-Wilk test – for small to medium-sized datasets.
- Anderson-Darling test – particularly for larger datasets or where distribution tails were of interest.
- Visual inspection of histograms and Q-Q plots – to support statistical tests and guide the selection of appropriate parametric or non-parametric methods.

3.5.2 MU Modelling

MU was calculated using methods appropriate to the nature of the assay and the available data:

- The ISO/TS 20914:2019 Top-Down Approach was the primary framework employed where IQC data were available to characterise within-laboratory imprecision and where manufacturers provided calibration uncertainty (u_{cal}).
- Where the top-down approach was not applicable—particularly where measurement models required propagation of uncertainty—calculations followed the GUM (JCGM 100:2008).
- For models involving complex uncertainty propagation (e.g. in calibration curve modelling or correlation studies), Monte Carlo simulation was used in line with JCGM 101:2008 guidance. This approach allowed uncertainty distributions to be simulated directly by random sampling from input parameter distributions.

The choice of MU estimation method was determined by the data structure, the availability of uncertainty components (u_{RW} , u_{cal} , u_{bias}), and the complexity of the measurement model. The specific application of each method is described in the corresponding experimental chapters.

3.5.3 Regression and Method Comparison

Several regression and method comparison techniques were employed to evaluate agreement between analytical methods:

- OLS regression – applied where assumptions of constant variance and error-free reference measurements were met.
- Deming regression – used where both comparative methods carried MU, allowing for errors in both axes.
- Passing-Bablok regression – a non-parametric regression method robust to outliers and used for method comparisons without assumptions of normality.
- Bland-Altman analysis – used to evaluate bias and agreement across the measurement range, including plots of mean difference against average values.

All regression models incorporated assessment of parameter uncertainty using resampling techniques, including:

- Bootstrap resampling – providing non-parametric estimates of confidence intervals by repeated random sampling with replacement.
- Jackknife resampling – an alternative approach for assessing parameter stability, particularly for smaller datasets.

3.5.4 Time Series Analysis

Time series analysis was applied to IQC datasets to investigate patterns of variability over time and to assess the random versus systematic components of measurement variability:

- Classical additive decomposition was employed to separate IQC data into trend, seasonal, and residual (random) components.
- Assessment focused on descriptive characterisation rather than predictive modelling or forecasting.

Time series data was assessed using:

- Runs tests for randomness – to evaluate whether sequences of IQC data deviated from randomness.
- Autocorrelation analysis – to assess dependencies between sequential IQC measurements.
- Augmented Dickey-Fuller (ADF) test – used to assess stationarity, although differencing was not applied as the focus remained descriptive.

All analysis was performed in the statistical programming environment R (R Team, Austria) using packages available on the CRAN network as indicated in the specific methodological subsections.

3.5.5 Simulation and Resampling Methods

Monte Carlo simulation was employed extensively where analytical solutions were not feasible or where propagation of uncertainty through non-linear models was required. Simulations were typically run for 10,000 iterations, sampling from distributions defined by input parameter uncertainties (e.g. standard deviations from IQC data, calibration parameter variability).

Both bootstrap and jackknife resampling methods were used routinely for:

- Estimation of regression parameter uncertainty.
- Evaluation of confidence intervals in method comparison studies

These were implemented in both R and Python as indicated in the relevant sections

3.5.6 Diagnostic Accuracy and ROC Analysis

Receiver operating characteristic (ROC) curve analysis was used in R to assess diagnostic performance where applicable. This included:

- Calculation of area under the curve (AUC) to quantify diagnostic accuracy.
- Derivation of likelihood ratios (positive and negative) to support clinical interpretation of diagnostic thresholds.

3.6 APS Calculator for Simulating Clinical Risk and APS

To evaluate how MU affects clinical risk and APS, the APS Calculator tool (accessible at <https://hikmetc-apscalculator.streamlit.app/>) was employed.

The APS Calculator was used to simulate the probability of correct classification at key clinical decision limits for ADAMTS13 activity assays, specifically:

- 10 IU/dL for initial diagnosis of TTP.
- 20 IU/dL to assess relapse risk in patients with previously confirmed TTP.
- The reference range boundary (~50 IU/dL) relevant to pregnancy planning or remission monitoring.

Monte Carlo simulation, set to 10,000 iterations, was performed for each scenario, incorporating the known MU for each assay. The calculator generated estimates of:

- The probability that a patient result falls within acceptable APS limits (minimum, desirable, optimal).
- The probability of misclassification based on the defined decision thresholds.

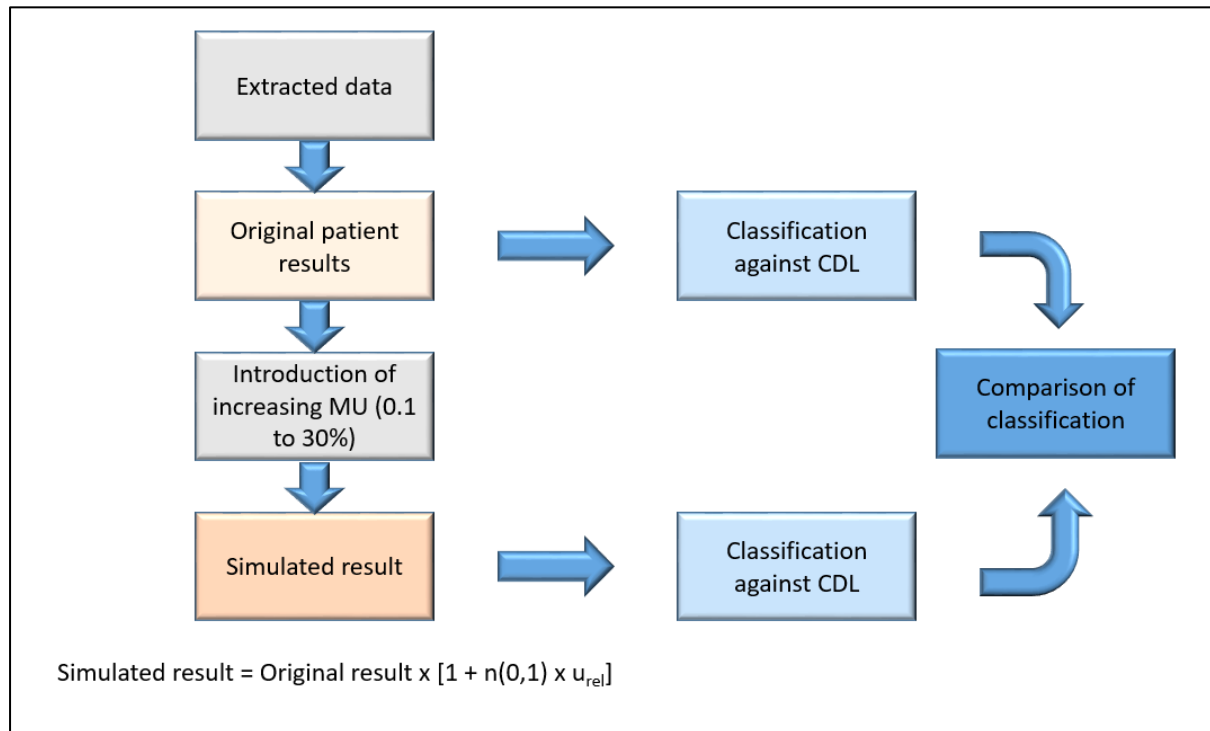


Figure 14: Simulating MU in patient results.

Schematic representation of the Monte Carlo Simulation process according to Çubukçu et al, 2023. CDL = chosen Clinical Decision Limits, MU – Measurement Uncertainty. Image created by S MacDonald, 2024.

Default agreement thresholds of 90%, 95%, and 99% were used to represent minimum, desirable, and optimal APS, respectively, in line with standard models in the literature.

This approach allowed direct visualisation of how uncertainty affects diagnostic safety margins and informed whether each assay achieved fitness for purpose in clinical practice.

3.7 Biological Variation Data Source

Biological variation data for ADAMTS13 activity were sourced from the EFLM biological variation database (<https://biologicalvariation.eu/>). Data from healthy male and female volunteers (median age: 30 years) indicated intra-individual variation (CV_i) of 12.73% and inter-individual variation (CV_g) of 9.63%.

Four different methods for determining BV based APS were used according to calculations shown in methods section 3.11.

3.8 Biological Variation APS calculations

The conventional linear model for setting APS (Equation 3) from combined BV was used as a baseline for comparison as this is the most commonly applied model (110).

$$CV_a = 0.25 * (CV_i^2 + CV_g^2)^{1/2} + 1.65 * (0.5 * CV_i)$$

Equation 3: The Fraser model for analytical variation.

This model defines the acceptable analytical coefficient of variation (CV_a) incorporating intra individual (CV_i) and inter individual (CV_g) biological variation (From reference (110)).

Used primarily for diagnosis, Gowans model (111) was used stating the maximum permissible error to be 0.6 x combined biological variation as below:

$$CV_a = 0.6 * (CV_i^2 + CV_g^2)$$

Equation 4: The Gowans model for analytical variation.

This models defines the acceptable analytical coefficient of variation (CV_a) incorporating intra individual (CV_i) and inter individual (CV_g) biological variation (From (111))

Finally, to reflect the updated models of Fraser (113), minimum, desirable and optimal specifications were applied as below:

optimum ($CV_a \leq 0.25 * CV_i$)

desirable ($CV_a \leq 0.5 * CV_i$)

minimum ($CV_a \leq 0.75 * CV_i$)

Equation 5: The updated model of Fraser.

Optimum, desirable and minimum acceptable analytical coefficient of variation (CV_a) incorporating intra individual (CV_i) and inter individual (CV_g) biological variation

Once biological variation data is known, and CV_a is calculated using the methods above, maximum allowable MU and Total Allowable Error are calculated according to the equations below:

$$MAU < 2 * 0.5 * CV_i$$

Equation 6: Maximum Allowable Measurement Uncertainty (MAU) from Biological Variation

$$TA_E < 1.65 * (0.5 * CV_i) + 0.25 * \sqrt{CV_i^2 + CV_g^2}$$

Equation 7: Total Allowable Error (TA_E) calculated from Biological Variation

Discussion between the two paradigms of specification limits is a heated debate that has gone on for many years (103, 201, 202), and remains unresolved with criticisms of mathematical validity (102) and handling of bias taking centre stage (203).

3.9 Development of the ADAMTS13 Error Grid

An error grid framework was developed based on the principles outlined in CLSI EP27-A to evaluate the clinical impact of discrepancies between the Technozym® ELISA and HemosIL AcuStar® ADAMTS13 assays.

Zone boundaries were defined using a combination of:

- Literature-based thresholds, including key clinical decision limits such as 10 IU/dL (diagnostic threshold for TTP), 20 IU/dL (associated with relapse risk), and ~50 IU/dL (reference interval lower limit).
- Local clinical consensus, involving consultation with haematologists, laboratory scientists, and specialist nursing staff experienced in TTP management.

The measurement range spanned 0 to 150 IU/dL, standardised to IU/dL for compatibility between the two methods.

The error grid incorporated MU in its design by defining guard bands around zone boundaries, thereby accounting for analytical variability. This enabled assessment of whether observed differences between methods were clinically significant versus analytically explainable.

The error grid was applied exclusively to the comparison of the ELISA and AcuStar® methods. It was not extended to the Technofluor® Ceveron platform due to the limited evaluation period and smaller dataset for that method.

3.10 Reference Interval Determination Using the refineR Algorithm

Reference intervals for ADAMTS13 activity were derived using the refineR algorithm (204), an indirect statistical method that estimates reference intervals from routine patient data without the need for formally defined healthy populations.

The refineR package, implemented in the R statistical environment (R Team, Austria), models the non-pathological distribution within a dataset that contains both pathological and non-pathological results. This is achieved through optimisation of a parametric distribution model based on maximum likelihood estimation and outlier-resistant fitting procedures.

In this thesis, refineR was applied to ADAMTS13 activity data generated from routine clinical requests. This allowed the derivation of a reference interval that reflects the laboratory's specific patient population and analytical characteristics. This approach aligns with ISO 15189:2022 requirements for reference interval verification and provides an evidence-based alternative to relying on manufacturer-supplied reference values, or locally derived reference populations not representative of patient requests.

Detailed methods, including the distribution fitting, goodness-of-fit assessment, and derived reference limits, are provided in Chapter 4.4.

3.10.1 The refineR algorithm

3.10.1.1 Model fitting and derivation of reference intervals

The refineR algorithm offers an innovative approach to estimating reference intervals using real-world data (RWD) (Bahr *et al.*, 2025). Unlike traditional direct methods that rely on predefined healthy cohorts, refineR analyses routine laboratory measurements. The algorithm assumes that most RWD results are non-pathological and focuses on identifying the main non-pathological distribution. To address skewness often present in such data, refineR applies a Box–Cox transformation—a power transformation technique that stabilises variance and makes the data more

normally distributed. This transformation enables more accurate modelling by improving symmetry and reducing the influence of outliers. The algorithm then optimises parameters for skewness (λ), mean (μ), and standard deviation (σ) using a multi-level grid search to minimise a negative log-likelihood cost function.

RefineR is implemented as an open-source R package on CRAN to process data, estimate reference intervals, and visualise results. This makes refineR a practical alternative to the direct method, particularly for laboratories working with complex or limited datasets.

3.10.1.2 *Pathological distributions*

The implementation of refineR to identify pathological distributions within RWD begins with preprocessing the dataset to separate pathological results from non-pathological ones. The algorithm assumes that most results are non-pathological, focusing on the main peak of the distribution, which represents this majority. It uses a Box–Cox transformation to accommodate skewness in the data, allowing for accurate modelling of both Gaussian and non-Gaussian distributions (Ammer *et al.*, 2021).

Pathological distributions, which deviate from the identified non-pathological range, are then highlighted. These distributions may overlap partially with the non-pathological range but are characterised by distinct density peaks outside the main distribution. By isolating these pathological results, refineR enables a clear delineation between normal and abnormal results, providing valuable insights into the presence and characteristics of underlying pathological conditions within the RWD.

3.11 Personalised Reference Intervals

3.11.1 Analytical Methods

The Homeostatic Set Point (HSP) for each individual was calculated as the mean of their previous test results. If the number of available measurements was insufficient for a reliable mean, the median was used instead to minimise the influence of potential outliers.

Analytical variation (SD_A) was estimated using data from routine quality control processes conducted within the laboratory. The total variation (TV_{set}) was calculated by combining analytical variation with biological variation, which was assessed at two levels:

Within-subject biological variation (SDI): This was sourced from the EFLM Biological Variation Database.

Within-person biological variation (SDP): This was derived from the individual's own historical test results, when available, to provide a personalised measure of variation.

The inclusion of both SDI and SDP in the calculation of total variation ensures that the estimate captures both general biological trends and individual-specific variability.

3.11.2 Statistical Methods

All calculations were performed in Microsoft Excel 365 (Microsoft, USA). Two statistical models based on prediction intervals were employed to calculate the personalised reference intervals (prRI).

3.11.2.1 Parametric Approach:

Patient individualised mean

Defined as the mean of the patient results at steady state, with no evidence of outliers and no detectable trend. This defines the HSP of an individual patient

$$Mean = \frac{\sum_{i=1}^n x_i}{n}$$

Equation 8: Derivation of the mean of the patient results

Total variation of the result (V_{TV}) combines analytical variability (CV_a) and biological variation (CV_b)

$$V_{TV} = k \times \sqrt{CV_a^2 + CV_b^2}$$

Where:

k = coverage factor for 95% probability

CV_a = Analytical Coefficient of Variation

CV_b = Biological Coefficient of Variation

Equation 9: Total variation of the result.

Variation incorporating analytical variability and biological variation

ariability of the homeostatic set point (V_{HSP}) determines the amount of variability, in an individual, around their own individual HSP

$$V_{HSP} = k \times \sqrt{\frac{CV_a^2 + CV_b^2}{n}}$$

Where:

k = coverage factor for 95% probability

CV_a = Analytical Coefficient of Variation

CV_b = Biological Coefficient of Variation

n = number of samples

Equation 10: Variability of the homeostatic set point

TV_{set} is the total variability of the homeostatic set point accounting for variability in the set point itself, as well as the total variability of the result

$$TV_{set} = k \times \sqrt{\frac{n+1}{n}} (CV_i^2 + CV_A^2)$$

Equation 11: Variability of the set point, including the set point

Personalised reference interval

$$prRI = HSP \pm k \times \sqrt{\frac{n+1}{n}} (CV_i^2 + CV_A^2)$$

Equation 12: Personalised reference interval

3.11.2.2 Non-Parametric Approach:

When the number of observations was small or the data distribution was unknown, a non-parametric method was used to estimate the prediction interval.

The prRI limits were then calculated using the formula: $prRI = HSP \pm TV_{set}$

This provided the upper and lower limits of the prRI, allowing for a more personalized interpretation of laboratory test results.

3.11.2.3 Reference Range Index (RR_i)

Used as a measure of reliability of the population reference range the RR_i should ideally be ≥ 0.9 and is derived from:

$$RR_i = \frac{prRI (UL - LL)}{popRI (UL - LL)}$$

Equation 13: reference Range Index (RR_i).

UL – upper limit, LL – lower limit, prRI – personal reference interval, popRI – population reference interval

3.12 Reference Change Value Calculations

This study applied four different statistical models for calculating the RCV to assess whether serial changes in ADAMTS13 activity represented statistically significant variation beyond expected analytical and biological noise. Each model incorporates varying assumptions about within-subject variability and result interpretation.

3.12.1 Common RCV Method

The common RCV formula is based on the combination of analytical imprecision (CV_a) and within-subject biological variation (CV_i), using the formula:

$$RCV = Z \times \sqrt{2} \times \sqrt{CV_i^2 + CV_a^2}$$

Equation 14: The common RCV method.

This method is used for determining significant differences between results. RCV == Reference Change Value, z = coverage factor (usually taken as 1.96 for 95% confidence), CV_i = intra individual biological variation, CV_g – inter individual biological variation.

where Z corresponds to the z-score for the desired probability (typically 1.96 for 95% confidence). CV_i is the intra individual biological variation sourced from the EFLM database, CV_a is the analytical CV from routine laboratory processing. This method uses the mean of two consecutive results as the reference value (homeostatic set point) and produces symmetrical thresholds for both increases and decreases.

3.12.2 Soletormos Method

The Soletormos model follows the same formulaic structure as the common RCV method above (Equation 14) but applies it specifically to log-transformed data, assuming multiplicative variability rather than additive. This method improves robustness when data distributions are skewed or

heteroscedastic. The threshold is then back-transformed to yield percentage change limits, maintaining symmetry but accommodating lognormality in the data.

3.12.3 Fraser Updated Model

The Fraser model, also derived from the classical RCV equation, refines the estimate of the homeostatic set point by incorporating longitudinal patient data. Instead of comparing only two values, it uses a rolling average of prior results to determine the patient's biological baseline. This aims to increase precision in determining whether a change is truly abnormal, especially in chronic disease monitoring where intra-individual trends are relevant.

3.12.4 Fokkema Method (EFLM Database)

The Fokkema model introduces asymmetrical RCV limits, recognising that biological variability may differ for increases versus decreases. Based on log-transformed ratios, the Fokkema model calculates RCV using:

$$RCV (increase) = e^{z \times \sqrt{2} \times \sqrt{CV_I^2 + CV_A^2}} - 1$$

Equation 15: RCV Increase according to the Fokkema model

$$RCV (decrease) = 1 - e^{-z \times \sqrt{2} \times \sqrt{CV_I^2 + CV_A^2}}$$

Equation 16: RCV decrease according to the Fokkema model

This yields directional thresholds, in this study corresponding to +66.2% for increases and –39.8% for decreases, as obtained from the EFLM biological variation database.

Each method was applied uniformly across all eligible patient data, using validated estimates for CV_a and CV_i. Absolute differences between sequential ADAMTS13 activity results were compared against

each model's respective RCV threshold to determine the proportion of results exceeding the criterion for significant change.

3.13 Ethical and Regulatory Approvals

All research conducted within this thesis was undertaken in compliance with the ethical, regulatory, and governance frameworks of Cambridge University Hospitals NHS Foundation Trust (CUHNHSFT) and the University of Hertfordshire, as well as relevant national legislation, including the UK Data Protection Act (2018) and the General Data Protection Regulation (GDPR).

Ethical approval was obtained from both:

The CUHNHSFT Research and Development (R&D) Department, and

The University of Hertfordshire Ethics Board (approval granted in 2017).

These approvals covered all components of the research, including:

The use of surplus patient samples originally collected for routine clinical care.

The retrospective analysis of patient data, IQC records, calibration data, and external quality assurance (EQA) performance.

All statistical modelling, method comparison studies, uncertainty evaluations, and quality control analyses presented within this thesis.

Patient consent was formally waived in accordance with institutional policies and the terms of ethical approval, given that no additional samples were collected for research purposes and all data were fully anonymised prior to analysis. Only surplus clinical samples that would otherwise have been discarded were used.

The research was conducted in full accordance with the principles of Good Clinical Practice (GCP), the Declaration of Helsinki (2013 revision), and the requirements of ISO 15189:2022, specifically with respect to data protection, confidentiality, and the responsible use of patient-derived material.

4 Chapter 4 - Clinical Thresholds and clinical

Interpretation

4.1 Introduction

APS are fundamental to defining acceptable limits for MU in clinical laboratory assays (114, 122, 145, 149, 151, 156). The Milan consensus criteria provide the established framework for determining APS, based on three models (96). Model 1, which assesses the impact of assay performance on clinical outcomes directly (Model 1a) or via simulation (Model 1b) (86, 100). Model 2, which uses biological variation (BV) data (205); and Model 3, which applies a 'state-of-the-art' approach derived from the performance of best available methods (120). The most established application of clinical outcome based APS in haemostasis is for the INR (175). Despite the critical clinical role of ADAMTS13, no APS have been defined using the Milan framework and supported by clinical and biological reasoning.

This chapter aims to address this gap by challenging the traditional approach of a single APS model being applied universally to ADAMTS13, or any clinical assay. Instead, it presents a clinically adaptive framework wherein different APS models are applied depending on the specific clinical utility of the assay. ADAMTS13 activity testing presents unique challenges across three distinct contexts: diagnosis of TTP at guideline defined clinical decision limits (3, 33), monitoring of patients during remission or relapse according to expert clinical practice (93) , and evaluation of treatment efficacy in patients receiving replacement ADAMTS13 therapy based on clinical trials (206, 207). Each scenario requires distinct considerations in defining analytically acceptable performance, requiring tailored APS models to support clinically meaningful interpretation.

The first section of this chapter focuses on defining allowable uncertainty at clinical decision limits using Model 1 of the Milan framework. Model 1b is adopted, using simulation-based approaches to evaluate how MU propagates into patient misclassification risk at critical diagnostic thresholds.

Previous work extending the simulation approach proposed by Farrance and Frenkel (170) and implemented using Microsoft Excel was further developed using the R statistical environment (196). Due to increasing computational complexity, a web-based solution using the Streamlit platform (100) was adopted, enabling *in silico* modelling of misclassification risk due to MU for ADAMTS13 activity assays.

The second section addresses APS for monitoring patient movement towards decision limits, applying Model 2 of the Milan criteria. Here, the limitations of population-based biological variation data are considered, particularly given the restricted BV datasets available for ADAMTS13 activity (112) (108) Drawing on learnings from other measurands (208); (104, 205, 209) the variability within published BV data and its impact on APS are critically evaluated. Furthermore, recognising that population-level BV-derived APS may inadequately reflect individual patient trajectories, this section proposes a novel application of personalised reference intervals (prRI) and reference change values (RCV), tailored to individual patient data rather than relying solely on population metrics. Methods including the refineR algorithm (204) are applied to local ADAMTS13 activity datasets, aiming to construct a practical framework for laboratories to define patient-specific analytical performance specifications based on real-world data.

Finally, the third section explores APS for treatment monitoring, a clinical application often overlooked in traditional frameworks. With the advent of recombinant ADAMTS13 therapy (210); (211); (212); (206) new challenges arise in defining analytically acceptable performance when monitoring therapeutic dosing and response. Analogous to therapeutic drug monitoring, this section applies concepts from pharmacokinetic and therapeutic monitoring frameworks (213); (214) to define APS based on treatment-related clinical endpoints.

4.2 Aims

This chapter aims to define clinically meaningful APS for ADAMTS13 activity assays across diagnostic, monitoring, and treatment contexts. It seeks to determine allowable MU at diagnostic decision limits

using simulation-based outcome modelling; establish monitoring APS through biological variation, reference change values, and personalised reference intervals; and explore APS requirements for treatment monitoring using recombinant ADAMTS13. By evaluating these distinct clinical scenarios, the chapter aims to develop a flexible, patient-centred APS framework, challenging the conventional application of single-model performance specifications.

4.3 Defining Allowable Uncertainty at Clinical Decision Limits

To evaluate how MU affects diagnostic decision-making for ADAMTS13 activity assays, this section applies Monte Carlo simulation modelling to derive APS at key clinical thresholds. The analysis quantifies the risk of patient misclassification across two platforms, supporting context-specific performance specifications.

4.3.1 Methods

4.3.1.1 Data Source and Extraction

Two Hundred consecutive ADAMTS13 activity results were retrospectively extracted for both the Technozym® ELISA and AcuStar® platforms from laboratory middleware (Modulab Gold®, Werfen, USA). Samples represented local requesting patterns during the period 01/02/2021 to 01/02/2024, and were de identified at the time of extraction. Only the first result for each identifier was used. The results spanned the entire measurement range, with lower reportable limits of 0.3 % for AcuStar® and 1.0 IU/dL for Technozym® ELISA (value of the lowest calibrator). Numeric results were required for simulation, therefore values of <0.2% on AcuStar® were imputed as 0.2% and retained for analysis as these results usually represent iTTP first time presentation samples .

4.3.1.2 Analysis Design

Ten Thousand Monte Carlo simulations were performed on each of the extracted results using the APS calculator available at: <https://hikmetc-apscalculator.streamlit.app/> (100). MU was added to each baseline result in 0.1% increments from 0.1% to 30% (see Figure 14). A new, simulated patient

results was derived. Original results and results with added MU were compared to clinical decision limits (defined as below). Misclassification rates (false positives/negatives) based on comparisons with original measured values were calculated at agreement thresholds of 90%, 95%, and 99%, corresponding to error (disease misclassification) rates of 10%, 5%, and 1%, respectively.

4.3.1.3 Clinical Decision Limits

Simulations were weighted to specific clinical scenarios based on expert clinical judgement and guideline recommendations according to Table 23:

Table 23: Decision limits applied to the APS Calculator.

TTP – Thrombotic Thrombocytopenic Purpura

Decision Limit	Clinical Scenario	References
< 10 IU/dL	Diagnosis of TTP	(1-3)
10–20 IU/dL	Sometimes referred to as the equivocal range. Traditionally complex results where interpretation is heavily reliant on clinical pre assessment	(1-3, 93)
20–60 IU/dL	Remission and relapse monitoring	(1-3, 93)
> 60 IU/dL	Return to reference range	(1-3, 93, 201)

4.3.1.4 Limits of agreement

Agreement thresholds are user defined in the APS calculator and define the magnitude of misclassification considered clinically tolerable. For this study levels were set at 90% (minimum), 95% (desirable) and 99% (optimal).

4.3.2 Results

4.3.2.1 Simulation Findings

The MAU for Technozym® ELISA and AcuStar® for diagnosis at 10 IU/dL with 1% risk of classification was 19.5% and 12.5%, respectively (Table 24). This represents optimal APS criteria. Optimal performance at additional clinical decision limits are also summarised in Table 24.

Table 24: Performance Specifications for Technozym® ELISA and HemosIL AcuStar®.

Based on using clinical decision limits to guide optimal performance specifications. MAU – Maximum Allowable Uncertainty

Clinical Scenario	Technozym® ELISA (Optimal MAU %)	AcuStar® (Optimal MAU %)
Diagnosis (<10 IU/dL)	19.5	12.5
Borderline (10–20 IU/dL)	5.3	3.7
Remission (20–50 IU/dL)	1.0	1.9
Reference Range (>50 IU/dL)	1.4	2.2

4.3.2.2 Model results

In the examples below, simulations are applied for different decision limits (sublevel) indicated in the top left corner of the figure (<10, 10-20, 21-60 and >60 IU/dL). Each graph represents summarised data from all MU simulations (y-axis, 0.1-30% in 0.1% increments). For example, for the <10 IU/dL decision limit (Figure 16) and addition of 30% MU, the sensitivity (purple) is reduced to <95% and the specificity (orange/red) is 98.5% when compared to the original result as ground truth. This increases the false positive rate to 1.5% and false negative rate to >5%. When MU is reduced to 10% these rates drop to 0.5% false positive and 1% false negatives indicating better analytical performance improves clinical outcome by reducing misclassification.

Acceptable limits are defined when the coloured (blue/green/red) line crosses the dashed vertical lines representing acceptable levels of agreement. These are colour coded as blue for optimum, green for desirable and red for minimum levels of agreement. The MU level at which the respective

line crosses the limit of acceptable agreement defines the MU level that must not be exceeded for that criterion – read from the y-axis. For example, in Figure 16, the blue colour (optimum) is present until the line crosses the dashed line corresponding to 99% agreement – this gives the MAU of 19.5% and the line turns green. Further increasing simulated MU causes the line to continue left (decreasing agreement) until it crosses the 95% line at which point it turns red. Importantly, if the line does not continue left far enough to cross the limits of agreement line – the maximum MU added (in this experiment 30%) does not cause misclassification in excess of that limit.

4.3.2.2.1 APS based on overall agreement

Overall agreement (Figure 15) represents diagnostic, monitoring, and prognostic applications, with all clinical uses considered of equal importance (and therefore no single clinical decision limit is weighted for). For each analytical method—ELISA and AcuStar[®] — three tiers of MU acceptability were specified: minimum, desirable, and optimum. The minimum acceptable MU was 13.8% for ELISA and 11.6% for AcuStar[®] (Figure 15, figure not shown for AcuStar[®], results in Table 24, simulations in Appendix B). The desirable level tightened these limits to 7.0% and 5.8% respectively, while the optimum acceptable MU reflected the highest level of stringency, defined as 1.0% for ELISA and 1.7% for AcuStar.

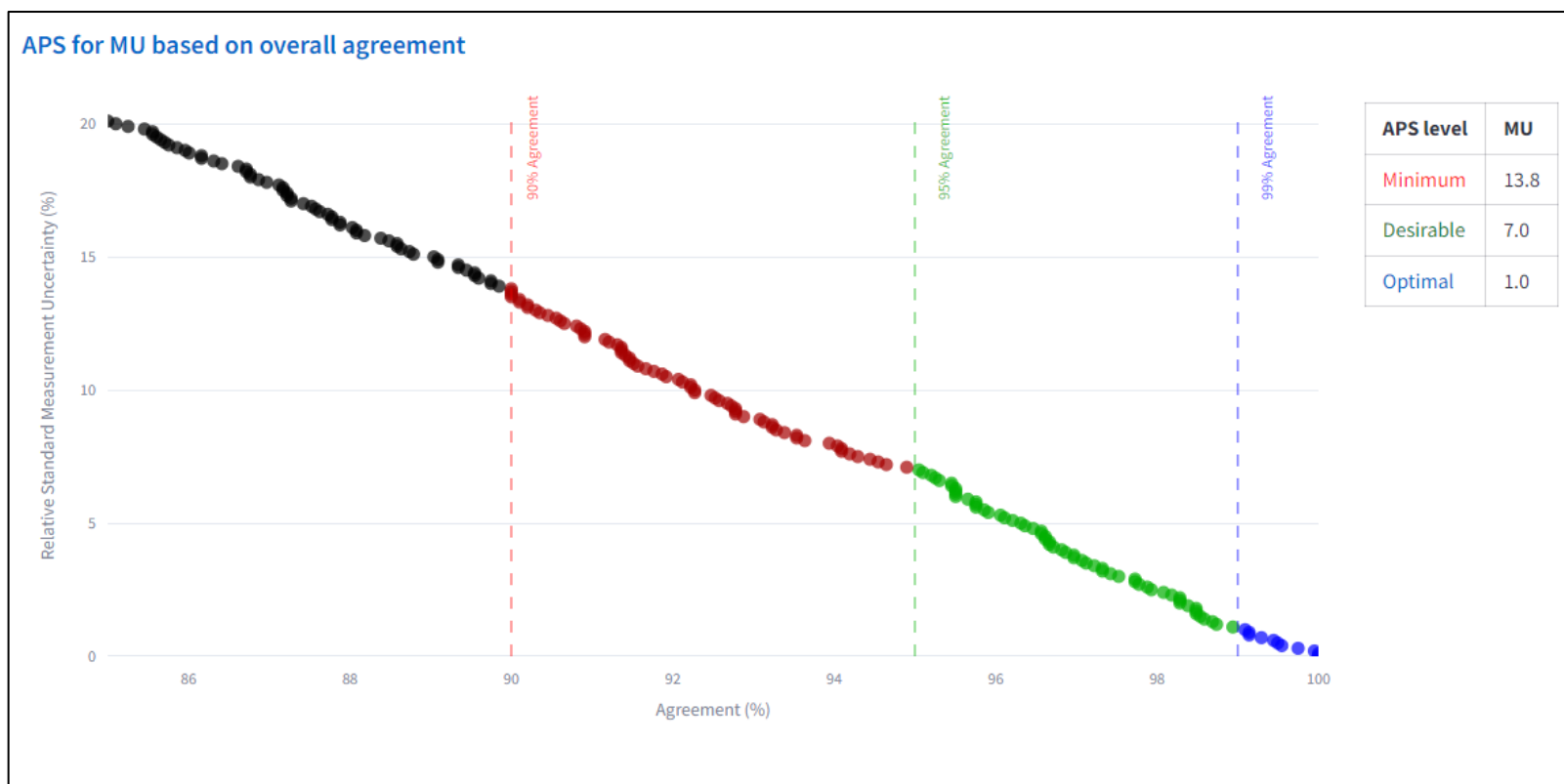


Figure 15: Analytical Performance Specifications for overall agreement.

Simulated APS (APS) limits based on overall agreement across the measurement range for ADAMTS13 activity using Technozym ELISA ®. The x-axis shows the extent to which metrics overall agreement decreased as MU added to patient results increased (y-axis). Dashed lines represent agreement limits pre-determined prior to the simulation, in the case here, 90%, 95% and 99% agreement.

In practical terms, however, only the minimum thresholds were considered achievable by both methods under routine laboratory conditions, with unacceptable misclassification rates of 10%. This approach, however, does not weight performance to where it is clinically most important, further investigations focused on specific areas of the measurement range to ensure this was accounted for.

4.3.2.2.2 APS maximizing agreement at the diagnostic threshold of 10 IU/dL

For TTP diagnosis at ADAMTS13 activity less than 10 IU/dL, optimum MAU was 19.5% for ELISA (Figure 16), reducing misclassification rate to <1%. The same optimal MAU was derived to be lower for AcuStar at 12.5% (Table 24).

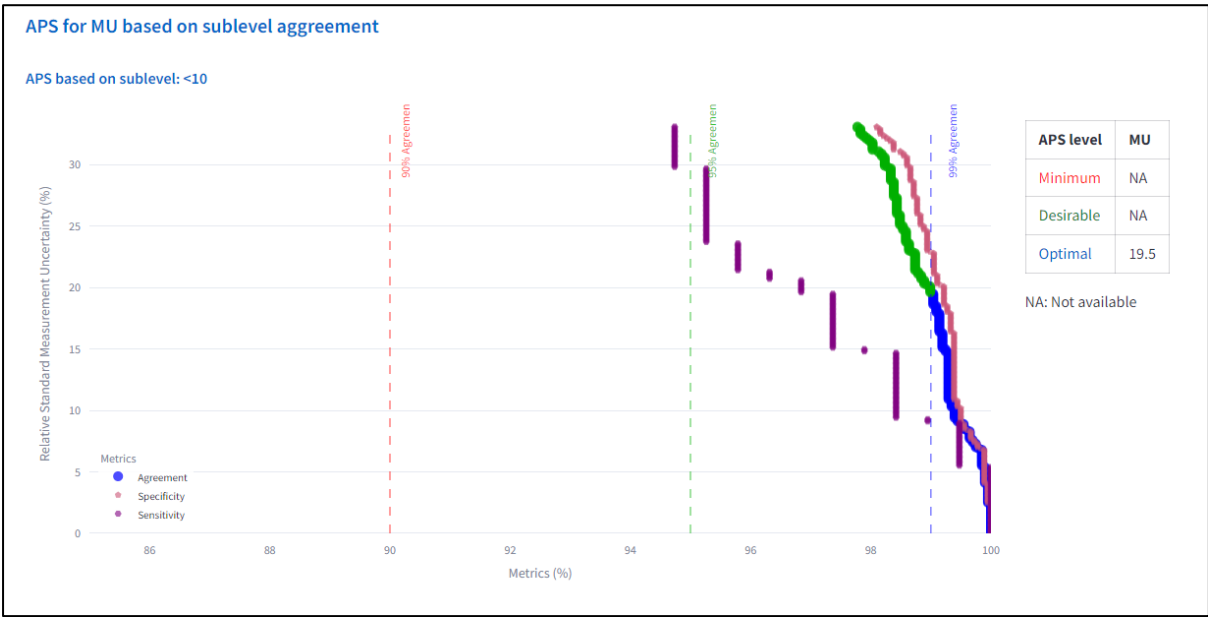


Figure 16: Analytical Performance Specifications for agreement at 10 IU/dL.

Simulated APS (APS) limits based on the diagnostic threshold for TTP(TTP) of <10 IU/dLADAMTS13 activity using Technozym® ELISA The x-axis shows the extent to which overall agreement (Blue/green/red line), sensitivity (purple line) and specificity (pink line) decrease as MU added to patient results increases (y-axis). Dashed lines represent agreement limits pre-determined prior to the simulation, in the case here, 90%, 95% and 99% agreement. In this case exceeding 5% misclassification (equivalent to 95% agreement) is not reached so only optimal MU is provided, ensuring between 1-5% misclassification if relative MU remains below 19.5%.

Specificity was largely maintained, only dropping to 98% even in the presence of significant (>30%) MU (Figure 16). Sensitivity was more markedly impacted with sensitivity dropping to 95% at 30% MU, but remaining high (98%) for MU up to 15%.

4.3.2.2.3 APS maximising agreement in the equivocal range of 10-20 IU/dL

The equivocal range of 10–20 IU/dL represents a borderline region used to support initial diagnosis, assess treatment efficacy (rarely), and monitor for early relapse (93). In this critical decision range, desirable and optimum MU thresholds were defined, with no minimum threshold specified as misclassification never exceeded 10% irrespective of MU added (Figure 17). Desirable MU was 27.7% for ELISA and 19.8% for AcuStar, while the more stringent optimum thresholds were 5.3% and 3.7%, respectively (Table 24, Table 25), recognising these would not be achievable with current analytical platforms.

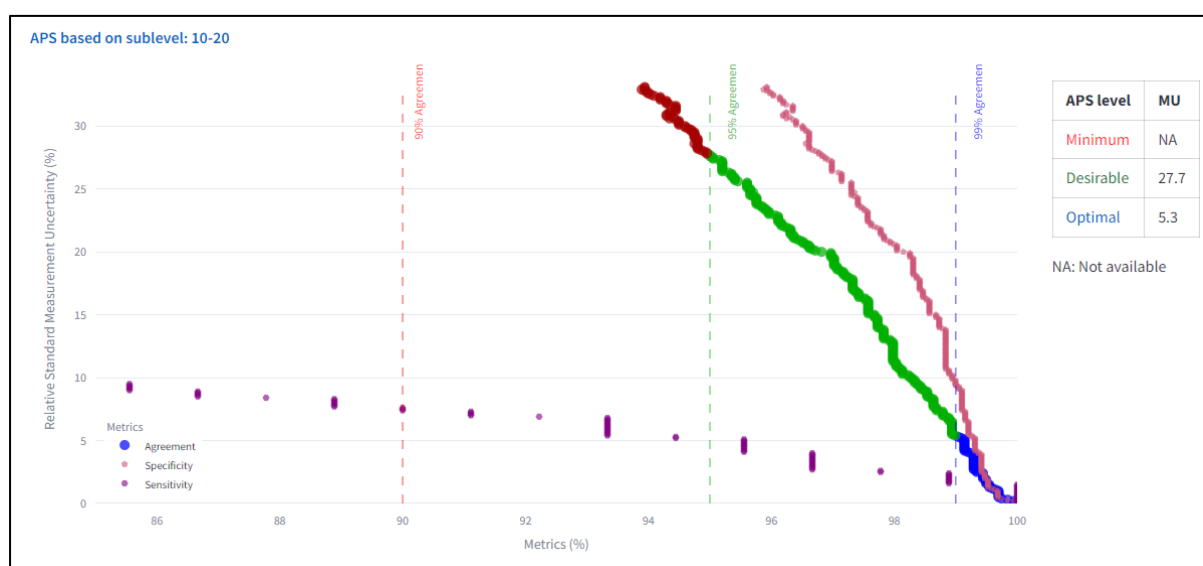


Figure 17: Analytical Performance Specifications for agreement at 10-20 IU/dL.

Simulated APS (APS) limits based on the equivocal range of 10-20 IU/dL used for first time diagnosis and treatment/relapse monitoring using Technozym® ELISA. The x-axis shows the extent to which overall agreement (Blue/green/red line), sensitivity (purple line) and specificity (pink line) decrease as MU added to patient results increases (y-axis). Dashed lines represent agreement limits pre-determined prior to the simulation, in the case here, 90%, 95% and 99% agreement. In this case exceeding 10% misclassification (equivalent to 90% agreement) is not reached so only desirable and optimal MU is provided, ensuring between 1-10% misclassification if relative MU remains below 5.3% (1% misclassification) or 27.7% (10% misclassification).

As with the overall agreement analysis (Figure 16) specificity was largely maintained only dropping to 96% even in the presence of significant (>30%) MU (Figure 17). Sensitivity was impacted with even very low levels of MU added and steadily declined as MU increased. With only 10% MU present, at the 10-20 IU/dL clinical decision range, sensitivity had dropped from near 100% to <86%. The drop in sensitivity was less marked using the AcuStar® platform, where 10% MU reduced the sensitivity from

100 to 90%. Interestingly the sensitivity in the AcuStar remained >96% if the equivocal range was changed to 5-20 IU/dL from 10-20 IU/dL (See Appendix B). The same was not seen for ELISA if the cutoff was changed. This implies 10 IU/dL is not an optimally sensitive cut-off for the AcuStar® and that a lower cut-off may be more appropriate for this platform. This is a concept studied in greater depth in chapters 5.4 and 9.3.

4.3.2.2.4 APS maximising agreement for the relapse range of 20-60 IU/dL

Within the 20–60 IU/dL range, ADAMTS13 activity is primarily used for assessing treatment response (traditionally immunosuppression but potentially replacement therapy in the near future) and monitoring for disease recurrence (93). All three levels of acceptable MU were defined for both methods. The minimum thresholds were 14.7% (ELISA) and 13.0% (AcuStar®), while desirable MU was significantly lower at 7.1% and 6.5%, respectively (Table 24, Table 25). The optimum MU was most stringent at 1.0% for ELISA and 1.9% for AcuStar ®. Both platforms were expected to meet the minimum thresholds in routine clinical practice, but again with misclassification risk raised to 10%.

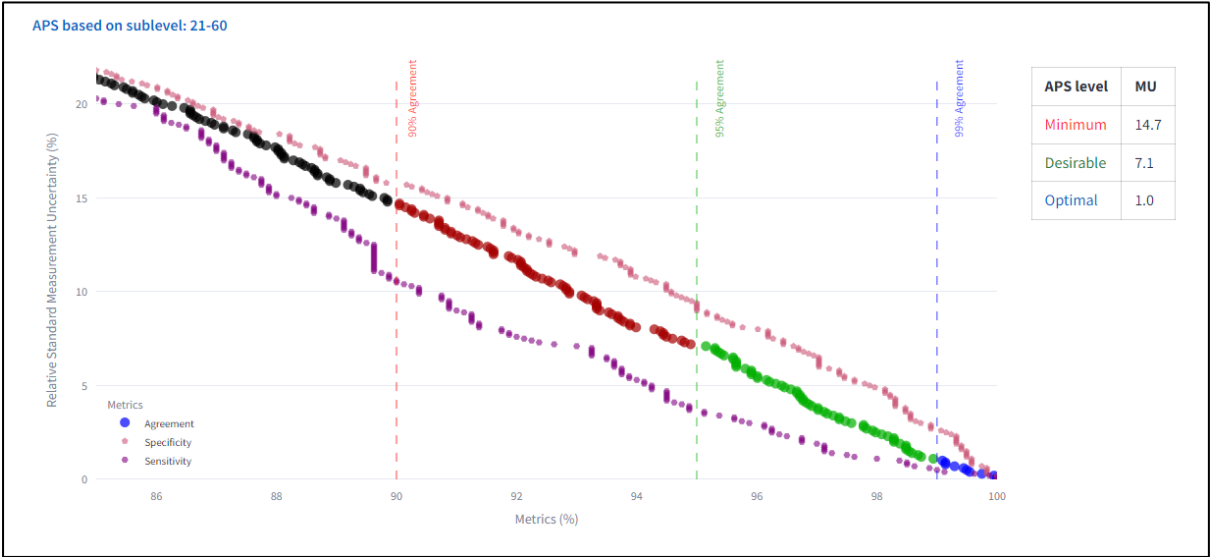


Figure 18: Analytical Performance Specifications for agreement at 21-60 IU/dL.

Simulated APS (APS) limits based on the relapse range of 20-60 IU/dL used for treatment/relapse monitoring using Technozym ELISA. The x-axis shows the extent to which overall agreement (Blue/green/red line), sensitivity (purple line) and specificity (pink line) decrease as MU added to patient results increases (y-axis). Dashed lines represent agreement limits pre-determined prior to the simulation, in the case here, 90%, 95% and 99% agreement. In this case misclassification rates >10% are achieved so minimum, desirable and optimal MU is provided.

When optimising agreement between 20-60 IU/dL, significant impact on sensitivity and specificity are observed (Figure 18). Both decrease consistently across the range of MU added with high sensitivity and specificity at low MU dropping to less than 90% with 15% MU added (specificity) and 10% MU (sensitivity) added. Despite the marked drop in performance, it is unlikely that this range would be prioritised in clinical practice.

4.3.2.2.5 APS maximizing agreement in the reference range

When ADAMTS13 activity exceeds 60 IU/dL, it is typically interpreted as a return to the reference range. Here, the analytical performance expectations reflect the reduced clinical risk associated with higher concentrations. Minimum MU thresholds were 17.0% for ELISA and 17.4% for AcuStar, while desirable MU levels were reduced to 8.1% and 7.9%, respectively. Optimum MU targets were defined as 1.4% for ELISA and 2.2% for AcuStar (Table 24, Table 25), clearly unachievable with current methods.

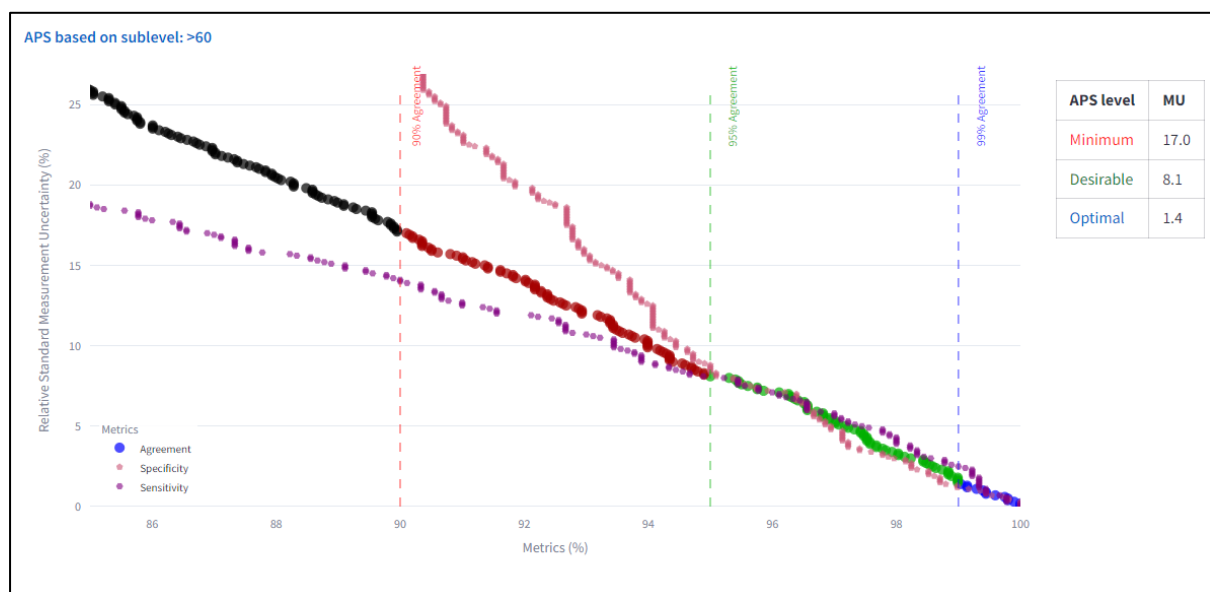


Figure 19: Analytical Performance Specifications for agreement at >60 IU/dL.

Simulated APS (APS) limits based on the reference range (>60 IU/dL) using Technozym ELISA. The x-axis shows the extent to which overall agreement (Blue/green/red line), sensitivity (purple line) and specificity (pink line) decrease as MU added to patient results increases (y-axis). Dashed lines represent agreement limits pre-determined prior to the simulation, in the case here, 90%, 95% and 99% agreement. In this case misclassification rates >10% are achieved so minimum, desirable and optimal MU is provided.

Sensitivity again shows a constant decrease with increasing MU; a pattern also reflected in specificity up to 10% MU added to the measured result (Figure 19). However, the influence on the specificity reduced at MU >10%, with a comparatively slower decline. The significance of this is in reality minimal as this range would not be prioritised for APS setting.

Table 25: ADAMTS13 analytical performance specifications by Model 1b.

Comparison of Technozym ELISA® and HemosIL AcuStar® derived APS from consecutive first time presentation ?TTP requests. IU/dL – International units per decilitre, MU – Measurement Uncertainty. Blank cells (-) are specifications that were not defined by the simulations as thresholds were not triggered

ADAMTS13 activity (IU/dL) investigated	Clinical Scenario where decision limit is applicable	Minimum Acceptable MU (%)		Desirable Acceptable MU (%)		Optimum Acceptable MU (%)	
		ELISA	AcuStar	ELISA	AcuStar	ELISA	AcuStar
Overall	Applicable for diagnosis, monitoring and prognosis with equal weight	13.8	11.6	7.0	5.8	1.0	1.7
<10	Diagnosis of TTP according to ISTH and BCSH Guidelines	-	-	-	-	19.5	12.5
10-20	Borderline for diagnosis, treatment efficacy and monitoring for relapse	-	-	27.7	19.8	5.3	3.7
20-60	Treatment efficacy and monitoring for relapse	14.7	13	7.1	6.5	1.0	1.9
>60	Return to within reference range	17	17.4	8.1	7.9	1.4	2.2

For overall agreement optimal MAU is very low (ELISA – 1.0%, AcuStar[®] - 1.7%, Table 25). This exceeds the capability of any analytical platform available currently. It is also unlikely to ever be achievable considering the uncertainty of the calibration hierarchy, in addition to imprecision performance observed in current laboratory practice. Even minimum acceptable goals (ELISA 13.8%, AcuStar[®] 11.6%), based on the Model 1b simulation are not likely to be routinely achievable.

Using a single decision limit, the impact of MU is reduced. The most clinically significant decision level is 10 IU/dL. For this decision level, irrespective of the amount of MU added to original results, misclassification rates did not exceed 5%. It was found that the optimal MAU for the ELISA and AcuStar[®] was 19.5% or 12.5% respectively (Table 24). In practice the AcuStar[®] may not achieve MAU because the CV is commonly 7-8% (discussed in chapter 6) and calibration uncertainty is 10% (information provided by manufacturer, Werfen, USA). The ELISA limits are at a level that is achievable based on current performance. For the AcuStar[®], if 5% misclassification rate is clinically tolerated, MAU should be <20%. This introduces a method specific change in risk profile, based on a single MU level being applied across two methods, highlighting additional complexity if MU is quoted by assay without taking method into account.

4.4 Biological Variation and Individual Interpretation

Section 4.3 suggested the MAU at the guideline defined clinical decision limit of 10 IU/dL to be 20% for both methods, with platform specific risk of misclassifications being a limitation. Other clinical decision limits were less well characterised due to drop in sensitivity and specificity with increasing MU. They do however, have secondary clinical importance as limits compared to the diagnostic 10 IU/dL cut off. These decision limits take on primary importance as patients transition through remission and relapse (93). Here, interpreting changes in ADAMTS13 activity becomes critical. To characterise this, biological variation is an often-recommended approach to quantify the likelihood of a clinically significant change in sequential results (110, 113). BV derived APS limits are often difficult to achieve with commonly used methods (section 1.5.5) and are intuitively not representative of the clinical situation where recognition of change is most needed. Application of population based, and personalised reference intervals and their impact on reference changes values is investigated here for the first time in the context of TTP remission and relapse.

4.4.1 Methods

4.4.1.1 *Sample data extraction*

Consecutive ADAMTS13 activity results on samples for investigation of TTP or MAHA were extracted from the laboratory middleware system (n=1637), Modulab Gold[®] (Werfen, Bedford, USA). The data covered the period from 2 October 2018 to 27 September 2024 and included results from both HemosIL AcuStar[®] and Technozym[®] ELISA. Following extraction, the data was deidentified except for hospital number, stored and processed in Microsoft Excel[®] 2016 (Microsoft Corporation, USA) to facilitate further analysis and preparation for statistical evaluation.

The results spanned the entire measurement range, with lower limits defined by the analytical methods: <0.2 IU/dL for AcuStar[®] and <1.0 IU/dL for Technozym[®] ELISA. To enable numerical assessment, results reported as below the measurement range (<1.0 IU/dL for ELISA and <0.2 IU/dL

for AcuStar[®]) were coded as 1.0 and 0.2, respectively. Similarly, results exceeding the measurement range (e.g., >150 IU/dL) were coded as the upper limit of the range. This coding approach was validated and did not affect algorithmic performance.

4.4.1.2 *Population based reference intervals*

4.4.1.2.1 *Data preparation*

For each analytical method (ELISA and AcuStar[®]), two datasets were created: one containing all results (AcuStar n = 631, ELISA n = 1007) and another limited to first-time presentation samples only (AcuStar n=416, ELISA n = 298). For hospital numbers with multiple results, only the first result was retained. The first sample received was used to construct the first-time presentation datasets, resulting in four datasets in total (two per method).

4.4.1.2.2 *Data Analysis*

The data was summarised in Microsoft Excel 2016. The data was then imported into the statistical environment R (R Team, Austria) for advanced analysis. The R package refineR (204) was used to process and prepare the data for statistical modelling and interpretation, according to the recommended practices from the package repository (https://cran.r-project.org/web/packages/refineR/vignettes/refineR_package.html) summarised in section 3.10.

4.4.1.3 *Personalised reference intervals (prRI)*

4.4.1.3.1 *Data Preparation*

From the Technozym[®] ELISA cohort, 25 patients had more than 20 measurements. Two patients with the highest numbers of samples (73 and 70) were excluded due to clinical but not ADAMTS13 remission following a full course of caplacizumab as has been reported in the literature (215). These patients exhibited prolonged low ADAMTS13 activity for over 12 months and required close monitoring, making them unsuitable for inclusion. The next five patients were selected for the analysis, with sample counts of 61, 51, 51, 43, and 40. These patients were long-term monitoring cases with recurrent relapses. Samples collected during evident relapse or while patients were

receiving immunosuppression were excluded in accordance with recommended approaches (216) to ensure the integrity of the steady-state analysis.

In the HemosIL AcuStar[®] cohort, repeat monitoring was less common due to the method's higher costs, later introduction, and preference for maintaining consistent testing methods for patients already monitored using ELISA. Despite this, five patients were identified with more than 10 samples, and two were selected for the pilot study. These patients had 12 and 10 samples, respectively.

The imbalance in cohort sizes reflects the laboratory's workflow. The AcuStar[®] assay is predominantly used for first-time presentations when results are urgently required, whereas the ELISA method is used for routine batch testing during follow-up monitoring.

4.4.1.3.2 Data Analysis

Serial results for each selected patient were summarised to identify outliers, which could indicate potential relapse, and to locate samples demonstrating ADAMTS13 recovery. All samples included were originally measured during routine monitoring while patients were in a healthy clinical state.

The data were divided into two categories: the full dataset, used for back-testing the utility of personalised reference intervals (prRI), and a subset containing only results from steady-state conditions. Median and interquartile ranges (IQR) of steady-state results were calculated to define the homeostatic set point (HSP) for each individual (section 3.11).

Linear regression was performed on each patient's cumulative data to identify significant trends that might suggest the absence of a steady state. No significant trends were detected, as evidenced by slope values near 0, confirming the stability of the selected samples for all seven patients (5 by ELISA, 2 by AcuStar[®]).

4.4.1.4 Reference Change Values

4.4.1.4.1 Data preparation

Adults (≥ 16 years old) undergoing longitudinal monitoring for previously diagnosed iTTP were selected from the dataset above (section 4.4.1.1). Only patients where five or more assays were performed were included in the final dataset. Within this cohort, 25 patients were monitored by ELISA and 16 by AcuStar[®] that met the inclusion criteria.

4.4.1.4.2 Data Analysis

This study assessed whether different RCV models reliably detect clinically significant changes in ADAMTS13 activity. Four RCV methods from Lund *et al.* (2016)—Common RCV, Soletormos, Fraser and Fokkema models—were evaluated, each differing in how they estimate homeostatic set points and set limits for significant change (section 3.12).

Within-subject biological variation was taken from the EFLM database (eflm.database.eu), while analytical variation was derived from IQC data. Simulation techniques were used to evaluate false-positive rates, and the sensitivity and specificity of each method were assessed for detecting significant declines in ADAMTS13 activity before clinical relapse.

4.4.1.4.3 RCV limits for ADAMTS13 activity

For the purposes of setting RCV limits in this study, two different approaches were applied in alignment with those referenced in the respective studies. Firstly, the Common RCV method, along with the Solétormos and Fraser models, applied a fixed threshold for detecting significant changes. This threshold was set at 44.3%, meaning that a difference of $\geq 44.3\%$ between successive results was required to trigger a significant change, regardless of the absolute value of the result.

The Fokkema method, as implemented in the EFLM Biological Variation Database, utilises asymmetrical limits due to the nature of log-normal distributions. In this model, the RCV threshold

for a decrease in ADAMTS13 activity was set at -39.8%, whereas the threshold for an increase was +66.2%.

4.4.2 Results

4.4.2.1 *Population based reference intervals*

The traditional use of CV_i and CV_g in APS modelling, while mathematically robust, often lacks alignment with the characteristics of actual patient populations from routine requesting patterns. In particular, the assumption that published CV_g values accurately reflect the population under test may introduce misestimation of performance requirements. This motivates exploration of an alternative approach: the use of empirically derived population reference intervals to reinterpret CV_g .

4.4.2.1.1 Technozym[®] ELISA Results all samples

The histogram (grey shading, Figure 20) displays the distribution of values measured in the total dataset ($n=1007$), while the green curve represents the estimated non-pathological distribution. The derived reference interval spans 55.8 to 110 IU/dL, with a median of 82.7 IU/dL—closely aligning with the locally established range from healthy volunteers. Notably, all data do not follow a typical Gaussian shape; instead, they resemble a uniform distribution, reflecting the heterogeneity of the sample population. The refineR algorithm delineates this central reference range with vertical dashed green lines and models the main distribution of non-pathological values effectively despite the complex input data. The CV% from the SD of the non-pathological group is 13.6% compared to a

CVg from BV studies of 9.63%.

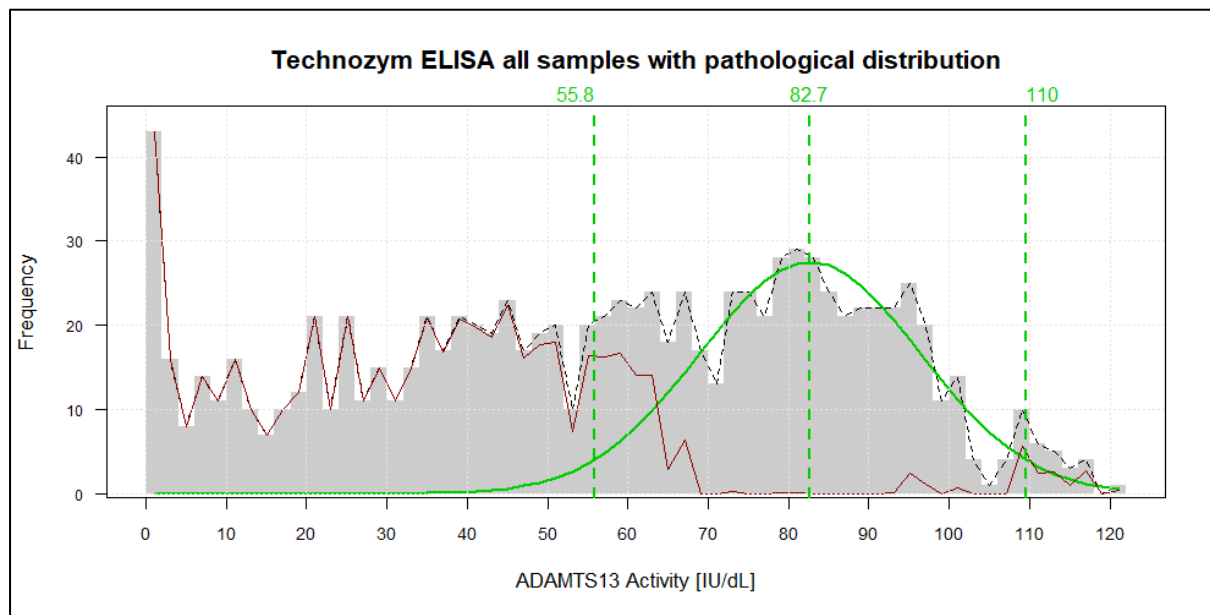


Figure 20: Population specific reference intervals from local data – Technozym ® ELISA

Graphical output of the refineR algorithm applied to ADAMTS13 activity data measured by Technozym ® ELISA including all results. The density plot shows the distribution of real-world laboratory results (grey/black dashed curve), with the estimated non-pathological distribution overlaid (green curve). The central 95% reference interval derived from the main distribution is indicated by vertical dashed lines (green). Pathological distributions, where identifiable, are visualised as distinct peaks outside the main reference interval shown in red.

Three distinct pathological distributions are evident (shown in red, Figure 20). The most prominent is a sharp, high-frequency peak at 0–5 IU/dL, corresponding to samples from patients with confirmed iTTP. A secondary, less pronounced distribution appears at the upper end of the range (90–120 IU/dL) but contains relatively few samples, more samples being in the final, uniform distribution between 5–70 IU/dL. These features indicate that inclusion of both diagnostic and remission-phase samples has introduced multimodal characteristics into the dataset, contributing to its non-normality. Nonetheless, refineR successfully isolates the main reference distribution from pathological outliers, demonstrating its utility in generating data-driven reference ranges from this local population of results.

4.4.2.1.2 HemoSIL[®] AcuStar Results all samples

For the AcuStar[®] results, the green density curve (Figure 21) represents a much wider non-pathological distribution compared to ELISA, bounded by the 95% reference interval (23.7–102%) indicated with green dashed lines. The median of the estimated reference range is 57.4%, the CV being 19.6%. Compared to the ELISA method, the AcuStar[®] reference interval spans a wider range and includes lower values, suggesting a method-specific definition of reference ADAMTS13 activity results. Notably, the non-pathological distribution appears asymmetrical, reflecting the dataset's heterogeneous composition.

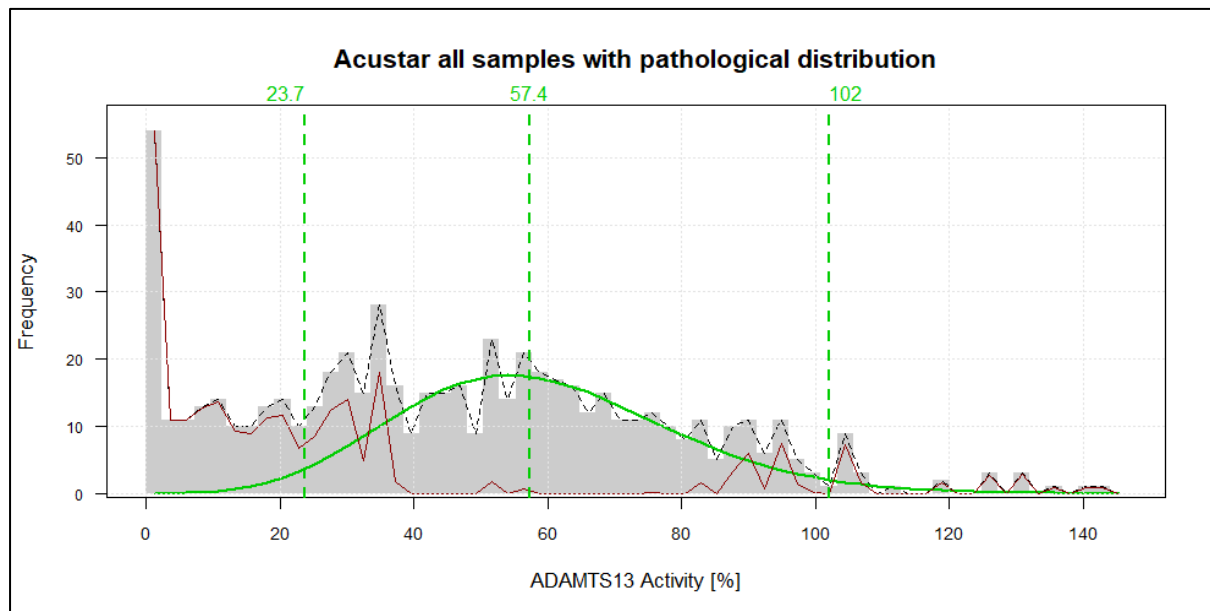


Figure 21: Population specific reference intervals from local data – AcuStar[®]

Graphical output of the refineR algorithm applied to ADAMTS13 activity data measured by HemoSIL AcuStar[®] including all results. The density plot shows the distribution of real-world laboratory results (grey/black dashed curve), with the estimated non-pathological distribution overlaid (green curve). The central 95% reference interval derived from the main distribution is indicated by vertical dashed lines (green). Pathological distributions, where identifiable, are visualised as distinct peaks outside the main reference interval shown in red lines.

As expected, a distinct and prominent peak is seen below 2% (red lines, far left, Figure 21), corresponding to samples taken during acute iTTP episodes. Additional pathological values are observed throughout the 2–40% range, although the overall spread is narrower than that observed in the ELISA distribution. The presence of both high and low ADAMTS13 values outside the reference interval, and the relatively uniform distribution within the 0–40% range, suggests a mixed population of samples, likely due to the inclusion of both initial diagnostic and follow-up monitoring specimens. As with ELISA, this multimodal profile complicates traditional reference range estimation but is well-resolved by the *refineR* algorithm, which distinguishes the principal non-pathological distribution from clinically relevant outliers.

4.4.2.1.3 Technozym[®] ELISA first time presentation samples

The calculated reference interval ranges from 36.3 to 82.3 IU/dL (Figure 22), with a median value of 59.3 IU/dL—demonstrating a notable leftward shift compared to the range derived from the full dataset seen in Figure 20. This narrowing and displacement likely results from excluding monitoring samples, which tend to cluster at higher activity levels, allowing a clearer delineation of diagnostic results. Notably sample results in the range between 50-70 IU/dL are not identified as being pathological in this range with essentially all results contributing to the reference range and not pathological distributions.

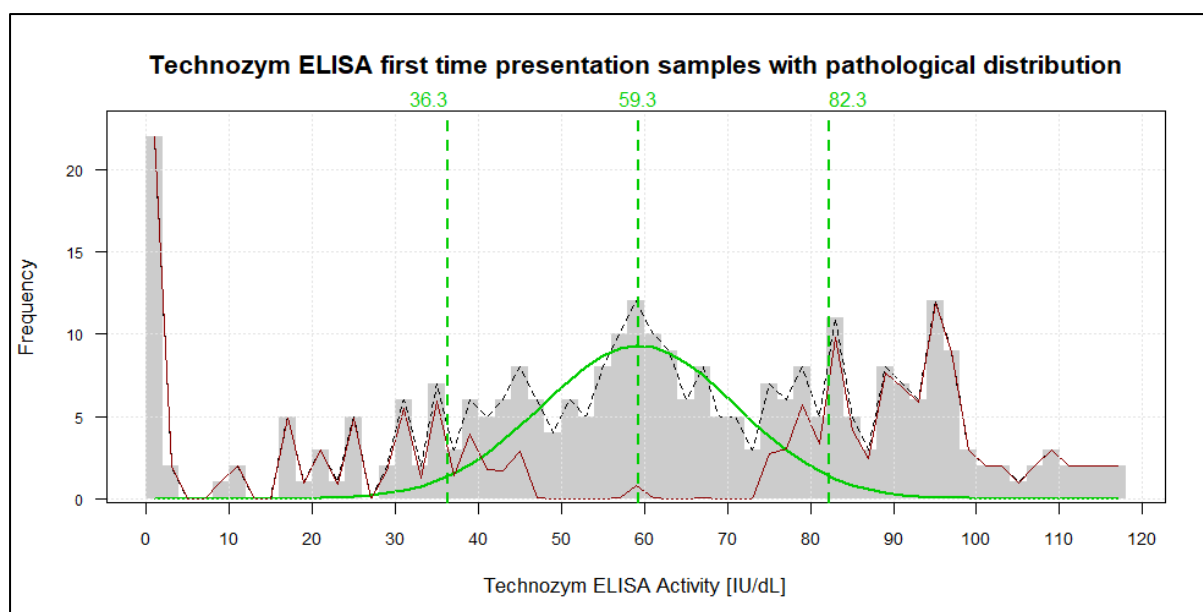


Figure 22: Reference intervals from local data – Technozym[®] ELISA, first presentation.

Graphical output of the refineR algorithm applied to ADAMTS13 activity data measured by Technozym[®] ELISA[®] including first time presentation results. The density plot shows the distribution of real-world laboratory results (grey/black dashed curve), with the estimated non-pathological distribution overlaid (green curve). The central 95% reference interval derived from the main distribution is indicated by vertical dashed lines (green). Pathological distributions, where identifiable, are visualised as distinct peaks outside the main reference interval shown in red lines.

Four distinct distributions are evident in the data: one central non-pathological distribution within the reference range, and three pathological components outside of it. Most striking is the sharply defined peak below 5 IU/dL, corresponding to confirmed cases of TTP, which is now isolated from the broader spread of low-level results observed in the full dataset (Figure 20). Two additional pathological bands appear between approximately 10–40 IU/dL and above 90 IU/dL, suggesting clinically distinct subgroups with mildly reduced or atypically elevated ADAMTS13 levels. The improved separation achieved by analysing first-presentation data alone highlights the potential of this approach for refining diagnostic thresholds and enhancing the interpretability of results in the context of acute disease presentation

4.4.2.1.4 HemoSIL[®] AcuStar first time presentation samples

Compared to the full dataset in Figure 21, the reference range in Figure 23, below, demonstrates only a modest shift, reflecting relative consistency in the AcuStar[®] method even when monitoring samples are excluded. However, as with the ELISA method, this stratified analysis allows more distinct pathological patterns to emerge.

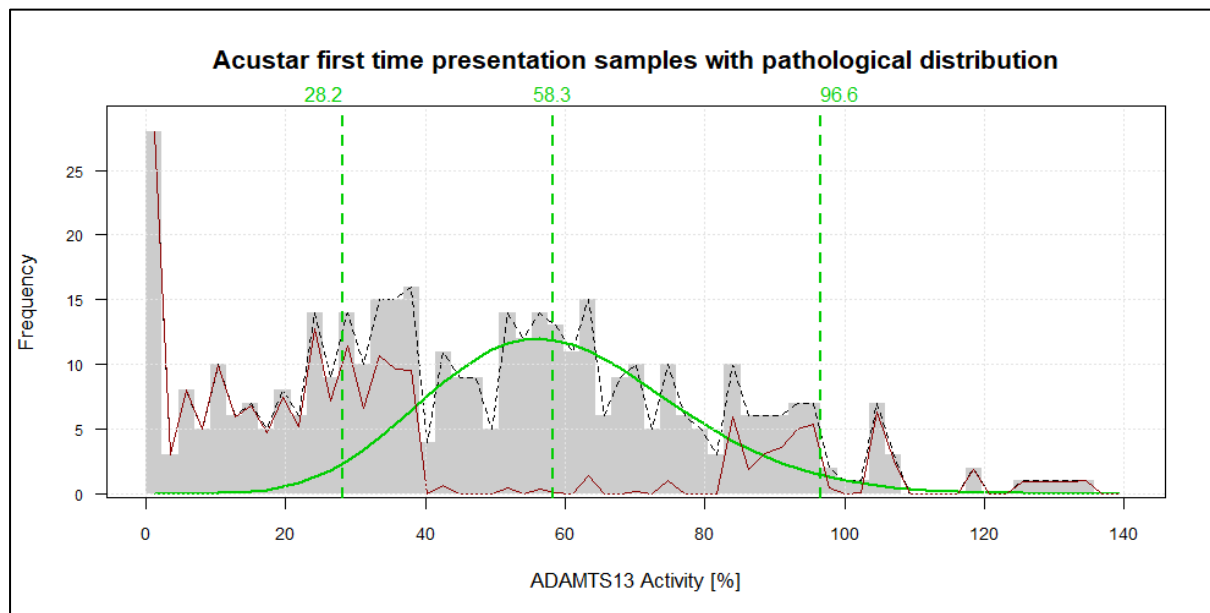


Figure 23: Population specific reference intervals from local data – AcuStar[®], first presentation

Graphical output of the refineR algorithm applied to ADAMTS13 activity data measured by HemoSIL AcuStar[®] including first time presentation results only. The density plot shows the distribution of real-world laboratory results (grey/black dashed curve), with the estimated non-pathological distribution overlaid (green curve). The central 95% reference interval derived from the main distribution is indicated by vertical dashed lines (green). Pathological distributions, where identifiable, are visualised as distinct peaks outside the main reference interval shown in red lines.

Figure 23 reveals a sharply defined peak in the <10% range, consistent with acute iTTP cases, alongside a broader pathological distribution between 10–40%, suggesting a cohort with mild to moderate ADAMTS13 deficiency. These features were also observed with the ELISA method, but was significantly more marked in the AcuStar[®] with more first time presentation samples appearing to be abnormal, but not severely so. This reflects observations in routine clinical practice (authors experience). The improved separation of distributions highlights the refineR algorithm's capacity to

extract clinically meaningful categories from heterogeneous datasets, including the ability to distinguish between diagnostic and monitoring samples. This approach may support improved interpretation of borderline results and stratification of risk, potentially informing future updates to clinical decision limits based on method-specific and patient-phase-specific patterns.

4.4.2.1.5 Summary of ranges described

Table 26 summarises the findings for the reference intervals and pathological distributions for ADAMTS13 activity derived using the refineR algorithm across two analytical methods. This was applied to both complete datasets and first-time presentation samples. In the full dataset, the ELISA-derived reference range was 55.8–110 IU/dL, with a notable pathological peak below 2 IU/dL and a broader uniform distribution of values below 70 IU/dL. For AcuStar[®], the reference range was wider (23.7–102%), but showed a similarly clear pathological peak below 2%, with a broader pathological range extending from 2–40%, suggesting a wider spread of low-activity results across the method.

Table 26: ADAMTS13 reference intervals from population sample data.

Summary of ADAMTS13 activity reference intervals and pathological distributions derived using the refineR algorithm for Technozym® ELISA and HemoSIL® AcuStar methods.

Dataset	Analytical Method	Patient data derived reference range	Pathological reference range (s)
All samples	Technozym® ELISA	55.8-110	Uniform distribution <70 IU/dL, clear peak < 2IU/dL
	HemoSIL® AcuStar	23.7-102	Uniform distribution 2-40 %, clear peak < 2%
First time presentation samples	Technozym® ELISA	36.3-82.3	Distribution > 70IU/dL Small distribution 5-40 IU/dL Peak < 2 IU/dL
	HemoSIL® AcuStar	28.2-96.6	Clear distributions > 80% and 2-40%. No pathological distribution within reference range, peak < 2%

In the first-time presentation subset, the reference intervals narrowed for both methods, with ELISA yielding 36.3–82.3 IU/dL and AcuStar® 28.2–96.6%. Importantly, both methods demonstrated improved separation of pathological distributions, with ELISA identifying distinct peaks above 70 IU/dL, a small pathological cluster from 5–40 IU/dL, and a strong peak below 2 IU/dL. AcuStar® similarly showed separation between a high-activity distribution above 80%, a distinct low-activity band from 2–40%, and a pathological peak below 2%, without any pathological overlap within the estimated reference interval. These findings demonstrate the value of separating first-time presentation data to more clearly delineate pathological and non-pathological activity profiles.

4.4.2.1.6 Contextualising APS with Observed Patient Distributions

From the intervals derived above, approximate coefficients of variation (CVs) were calculated to represent the spread of values observed in routine, non-pathological testing populations.

In the all-sample dataset, the Technozym® ELISA produced an estimated CV of 16.4%, corresponding to an allowable imprecision (CV_a) of 8.2% at the 50% level, and 12.3% at the 75% level. For the same dataset, the AcuStar® assay yielded a markedly higher CV of 31.5%, resulting in corresponding allowable CV_a thresholds of 15.8% and 23.6%, respectively.

Table 27: RefineR-Derived Coefficients of Variation.

CV – Coefficient of Variation, CVI – Intra individual biological variation, CV_a – Analytical Coefficient of variation

Source	Method	Dataset	CV Input (%)	Fraser Desirable CV _a (%)	Fraser Minimum CV _a (%)
Published CVI (used as CVI)			12.7	6.4	9.5
RefineR CV (population-level CV)	Technozym ELISA	All Samples	16.4	8.2	12.3
RefineR CV (population-level CV)	HemoSIL AcuStar	All Samples	31.5	15.8	23.6
RefineR CV (population-level CV)	Technozym ELISA	First-time Presenters	19.4	9.7	14.6
RefineR CV (population-level CV)	HemoSIL AcuStar	First-time Presenters	27.4	13.7	20.6

In the subset of first-time presenters, narrower distributions were observed. The Technozym® ELISA returned a CV of 19.4%, with 50% and 75% CV_a thresholds of 9.7% and 14.6%. The AcuStar method in this group showed a CV of 27.4%, yielding allowable CV_a values of 13.7% and 20.6% under the same proportional constraints.

For comparison, a published CVI of 12.7% was used to calculate benchmark thresholds based on the Fraser model. These yielded allowable imprecision values of 6.4% (50% of CVI) and 9.5% (75% of CVI).

The complete set of calculated thresholds is presented in Table 27, showing how allowable CVa values shift in relation to the source and magnitude of the underlying variability estimate, as well as between assay platforms and patient populations.

4.4.2.2 *Personalised Intervals (prRI)*

While population-based reference intervals offer an alternative representation of between-subject variability based on local patient testing, they do not address the within-subject variation (CV_i) that underpins individualised interpretation of laboratory results. Likewise, biological variation estimates in health do not accurately represent patients in remission for iTTP. Therefore, the next section considers personalised biological variation — particularly through personalised reference intervals (prRI) — as a method to more accurately represent clinical changes in patients during remission for iTTP. This method is gaining popularity in the clinical biochemistry field (217-219), but has not yet been considered for haemostasis applications.

4.4.2.2.1 *Dataset summary*

Among the selected patients (section 4.4.1.3.1), differences in sample numbers and apparent HSPs were observed, as detailed in Table 28. After removing outliers and excluding results from patients undergoing immunosuppressive therapy, the number of samples available for analysis ranged from 4 to 19 per patient. This variability in sample numbers allowed for a comparison of parametric (section 3.11.2.1) and non-parametric (section 3.11.2.2) approaches to calculating personalised reference intervals.

Variability around each patient's HSP was found to be patient-specific and appeared unrelated to the absolute value of the HSP.

Table 28: Longitudinal data summary for monitored patients.

ELISA (x) represents each patient for which monitoring was performed using the Technozym[®] ELISA. AcuStar (x) represent the two patients for which routine monitoring was performed on the HemoSIL AcuStar[®]. N = number of samples, SD = Standard Deviation, IQR – interquartile range, df = degrees of freedom (n-1), t-value – critical t value derived from t table to represent probability of 95% for respective degrees of freedom, CV – coefficient of variation.

All units expressed as IU/dL for ELISA, % for AcuStar)	ELISA (1) (IU/dL)	ELISA (2) (IU/dL)	ELISA (3) (IU/dL)	ELISA (4) (IU/dL)	ELISA (5) (IU/dL)	AcuStar (1) (%)	AcuStar (2) (%)
n	4	19	19	19	11	8	4
Mean ADAMTS13 activity	94.2	68.7	53.6	22.4	57.8	59.7	69.9
SD	9.1	17.7	11.6	9.4	12.8	18.3	7.9
CV%	9.7	25.8	21.6	42.0	22.1	30.7	11.3
Median	92.9	65.7	50.4	19.3	51	53.4	66.9
IQR	88.4 - 98.7	55.0 - 86.4	43.4 - 63.6	15.1 - 31.6	46.9 - 67.6	45.3 - 70.7	65.9 - 71.0
df	3	18	18	18	10	7	3
t value	3.18	2.10	2.10	2.10	2.23	2.37	3.18

ELISA (1) yielded the highest mean value of 94.2 with relatively low variability (IQR - 88.4 to 98.7 IU/dL), indicating consistently elevated results (Table 28). In contrast, ELISA (4) reported the lowest mean of 22.4 and a wider interquartile range (15.1 to 31.6 IU/dL), reflecting greater spread among low values. The remaining ELISA runs showed intermediate means between 53.6 and 68.7, with differing degrees of variability. AcuStar[®] assays exhibited tighter clustering, especially AcuStar (2), which had a narrow interquartile range (65.9–71.0 %).

4.4.2.2.2 Technozym[®] ELISA prRI

Personalised reference ranges for each patient, as determined by the Technozym[®] ELISA, showed notable variability (Table 29). ELISA (1) was the only individual whose personalised reference interval (58.1–130.3 IU/dL) was comparable to the range established in local normal donors, or that shown in sections (4.4.2.1.1 and 4.4.2.1.3) using patient data for reference intervals. The remaining four patients had narrower reference ranges than the normal population. They also showed a lower end of their personal reference range, within the intermediate pathological area of patient derived population reference ranges from section 4.4.2.1.

ELISA (2) had a marginally lower reference range, with a lower boundary of 44.6 IU/dL. Patients 3 and 5 exhibited similar ranges, reinforcing the tendency for reduced ADAMTS13 activity levels in TTP patients in remission.

Patient 4 had the narrowest range, reflecting stable but persistently low ADAMTS13 activity levels during remission. This patient's activity remained abnormally low compared to normal reference values, yet no clinical signs of relapse were observed. Despite receiving multiple immunosuppressive treatments over several years, ADAMTS13 activity did not approach the normal reference range.

Table 29: Personalised reference intervals (prRI) for ADAMTS13 activity.

Intervals are defined by upper (UL) and lower limit (LL) using the biological variation method and personal variability method. popRI – population (local laboratory) reference interval

Biological variation method	ELISA (1) (IU/dL)	ELISA (2) (IU/dL)	ELISA (3) (IU/dL)	ELISA (4) (IU/dL)	ELISA (5) (IU/dL)	AcuStar (1) (%)	AcuStar (2) (%)
prRI LL	58.1	44.6	34.8	14.6	37.1	36.2	41.0
prRI UL	130.3	92.8	72.4	30.3	78.5	83.1	98.9
SD (CV%)	18.1 (19.1)	12.1 (17.6)	9.4 (17.5)	3.9 (17.4)	10.35 (17.9)	11.7 (19.6)	14.5 (20.7)
Personal variability method							
prRI LL	61.9	30.6	28.7	2.3	27.9	13.8	42.0
prRI UL	126.5	106.8	78.5	42.6	87.7	105.6	97.9
	16	19.05	12.45	10.1	14.95	22.95	13.97
popRI LL	60.6						
popRI UL	130.6						

4.4.2.2.3 HemoSIL ® AcuStar prRI

The patterns observed in Technozym ® ELISA results were also reflected in the AcuStar ® method, although data were available for only two patients. Both personalised reference ranges had a lower limit below the established normal reference interval but, as seen with ELISA, these ranges were narrower and therefore more representative of each patient's stable state. This study was limited to only two patients, as routine monitoring is typically conducted using a single analytical method. In most cases, ELISA is preferred because rapid results are not required, whereas AcuStar® is generally used in situations requiring urgent testing.

With a tailored reference range for individual patients, personalised interpretation of changes in results, during remission is possible, where the baseline expected level of patients can be taken into account, before instigating clinical intervention. This is possible by simple comparison of latest result

to previous result, and prRI. However, commonly statistical methods are employed to identify clinically significant changes, accounting for biological and analytical variability. To assess whether reference change values provide more meaningful interpretations compared to personalised reference intervals, studies of RCV by different statistical methods were performed.

4.4.2.3 *Reference Change Values*

While references ranges often guide clinical decision making, in patients with previously diagnosed iTTP early recognition of dropping ADAMTS13 activity can facilitate early immunosuppressive intervention irrespective of the reference range (1, 93). Common practice is that an ADAMTS13 level of < 20 IU/dL is used as the trigger point for immunosuppression, usually with Rituximab (1, 3, 93). However, detecting a trend toward this threshold, before it is reached, could provide clinical benefit. Reference change values provide a statistical measure, based on biological variation, analytical variation and the patients current and last result(s) to determine whether a clinically significant change has occurred (113, 220, 221) but are not without their challenges (221). There are numerous different statistical approaches to implementing the RCV concept (222, 223), here a comparison of their performance in longitudinally monitored iTTP patients is investigated.

4.4.2.3.1 *Frequency of Significant Changes Detected Across RCV Methods*

Figure 24 presents the percentage of measured values across all successive samples for 26 patients monitored using the Technozym ELISA, indicating the proportion of results that exceeded the RCV threshold according to four different calculation methods and two limit-determination approaches (section 3.12).

The proportion of flagged significant changes varied substantially between the different RCV models (Figure 24). The Common RCV method was the most conservative, detecting 14% of samples as significant, whereas the Solétormos method flagged the highest proportion of changes (29%). The Fraser (23%) and Fokkema (27%) methods exhibited intermediate detection rates.

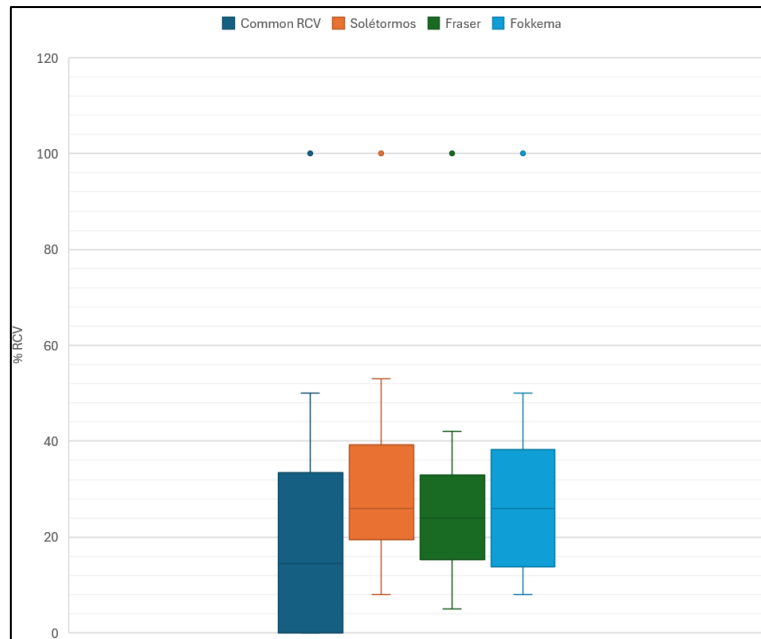


Figure 24: Proportion of samples giving positive RCV change.

When considering individual patients, variability in flagged changes was notable. For example, patient O was flagged by Solétormos in 50% of samples, compared to 0% by the Common RCV method and 33% by Fraser (Appendix C). Patient L was flagged 44% of the time by Solétormos and Fokkema, while the Common RCV method detected no changes. Patient W was flagged 100% of the time by all methods, demonstrating that all RCV models function consistently when changes are extreme. These results indicate that the choice of RCV model significantly impacts the likelihood of interpreting a significant change, with Solétormos being the most sensitive and Common RCV being the most conservative.

A key determinant of the differences observed between methods lies in their calculation approaches. The Common RCV, Solétormos, and Fraser methods apply a percentage-based threshold, whereas Fokkema utilises a ratio-based model. Notably, the Fokkema model frequently detected increases from low baselines, likely due to its asymmetrical limits (-39.8% for decreases and +66.2% for increases). This pattern was observed particularly in patients undergoing immunosuppressive therapy, where low ADAMTS13 levels increased significantly.

4.4.2.3.2 Sensitivity of RCV Methods at ADAMTS13 Levels Below and Above 20 IU/dL

Stratifying results by the 20 IU/dL clinical intervention threshold as suggested by Cuker *et al* (93) further illustrates method-specific differences in sensitivity. When the current ADAMTS13 result was below 20 IU/dL, the Solétormos method flagged 48% of cases as significant, far exceeding the detection rates of the Common RCV (20%), Fraser (17%), and Fokkema (43%) models (Table 30). However, when the current result was above 20 IU/dL, detection rates dropped significantly for all methods, with Solétormos reducing to 23%, suggesting its higher sensitivity at lower ADAMTS13 levels.

Table 30: High probability relapse reference change values.

RCV method detection rates on data stratified on the decision threshold for high probability of relapse (20 IU/dL)

	Common RCV	Solétormos	Fraser	Fokkema
Current result < 20 IU/dL	20%	48%	17%	43%
Current result > 20 IU/dL	13%	23%	26%	21%
Next result < 20 IU/dL	15%	25%	25%	24%
Next result > 20 (predictor)	14%	41%	19%	36%

Similarly, when using the next ADAMTS13 result as an outcome measure, 25% of cases with a next result below 20 IU/dL were flagged as significant by Solétormos, compared to 15% (Common RCV), 25% (Fraser), and 24% (Fokkema). When used as a predictor of future ADAMTS13 status, Solétormos detected 41% of cases where the next result exceeded 20 IU/dL, compared to 14% (Common RCV), 19% (Fraser), and 36% (Fokkema). Overall, Solétormos exhibited high sensitivity at low ADAMTS13 levels, identifying the highest proportion of changes when the current result was <20 IU/dL (48%). Fraser and Fokkema behaved more conservatively, detecting 17% and 43% of cases, respectively, in

this scenario. Solétormos also had the highest rate of change detection (41%) when predicting the next result to be >20 IU/dL, reinforcing its predictive sensitivity, but also raising concerns about potential false positives. Fraser provided more consistent detection rates across different stratifications, suggesting a more balanced approach in steady-state monitoring.

It is clear, that the successful implementation of statistical measures for detecting significant change in sequential results is entirely method specific. From an iTTP perspective, using prRI and understanding of where patients naturally fluctuate in remission, is more akin to how clinicians and scientists interpret clinical results.

4.5 APS for Treatment Response

4.5.1 Methods

A review of published recommendations and summaries of product characteristics (SPC) for treatments for iTTP was conducted to identify commonly applied treatment regimens. Using these data, plasma half-life and dosing frequencies were incorporated into the APS model proposed by Cattaneo (213). Minimum performance specifications were derived by an additional 50% of the desirable goal being added to the desirable goal.

The design of the study allowed all potential clinical applications of each treatment including the following treatment regimens:

1. Prophylactic Administration of r-ADAMTS13
 - Dosing every other week.
 - Weekly dosing.
2. On-Demand Therapy
 - Day 1: 40 IU/kg body weight.
 - Day 2: 20 IU/kg body weight.
 - Day 3 and beyond: 15 IU/kg body weight until two days after the acute event resolves.

Equation 17, originally developed for clinical biochemistry drug monitoring by Fraser (1987), was applied to derive APS for r-ADAMTS13. Key parameters included the time interval between doses (T) and the average elimination half-life (t).

$$\text{Desirable performance specification} = \frac{1}{4} [(2T/t - 1)/(2T/t + 1)] \times 100$$

Equation 17: Maximum allowable uncertainty for drug therapies.

Desirable performance specification (maximum allowable MU) based on pharmacokinetic modelling of replacement therapy in immune TTP (iTTP). T = Time interval between doses, t = average elimination half-life. (formula from Cattaneo, 2023)

4.5.1.1 Analysis Design

To determine whether dosing frequency significantly impacts APS, simulations were conducted using different values of T. These simulations assessed the impact of varying dosing regimens on APS outcomes. All analysis was performed in Microsoft Excel[®] 2016 (Microsoft Corporation, USA)

4.5.1.2 Data Sources

Data for APS calculations were derived from publicly available information, including prescribing information for ADZYNMA[®] (ADAMTS13, recombinant-krhn, Takeda, London, UK). Both prophylactic and on-demand applications were considered. Data for Fc-Fusion r-ADAMTS13 were derived from (224).

4.5.2 Results

Allowable MU estimates varied according to frequency of administration (Figure 25). Interestingly MU limits for rADAMTS13 and Fc-fusion ADAMTS13 were very similar with a limit of 25% for desirable APS and minimum being between 35 and 40.

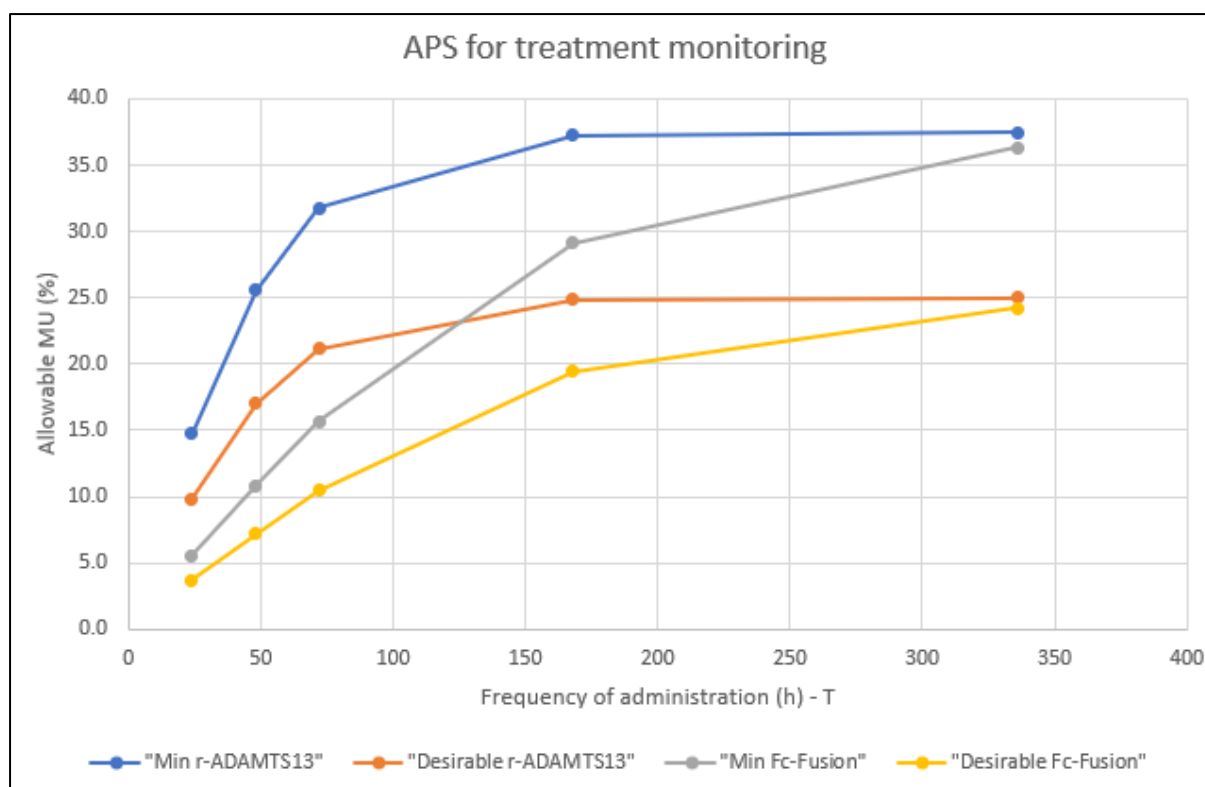


Figure 25: Analytical Performance Specifications for ADAMTS13 replacement therapy.

Profile of changing maximum allowable MU limits as a function of time between doses for recombinant ADAMTS13 (r-ADAMTS13) and Fc-fusion experimental ADAMTS13 (Fc-Fusion). Minimum and desirable MU

The profile of APS relationship with frequency of administration showed differences between both types of ADAMTS13 with r-ADAMTS having a steeper APS increase at low frequency doses compared to Fc-fusion (Figure 25). Allowable MU at frequent dosing times < 50 hours, fall below what is technically achievable and raises concerns about compliance with recommended MU levels in this clinical utility if such dose frequencies are required, for on demand treatment of acute episodes. Whilst this model when applied here reassures that, in the most part current ADAMTS13 assays are able to achieve APS to make them clinically useful in this setting, close scrutiny of dosing regimens, and discussion with clinicians for expected changes in results, in particular over short time frames would be advisable.

4.6 Chapter 4 Discussion

This chapter explored multiple, distinct yet complementary approaches to defining APS for ADAMTS13 activity assays across three principal domains: diagnosis at clinical decision limits, monitoring patient trends over time, and evaluating replacement therapy in treatment contexts. The central argument advanced is that no single APS model is universally appropriate; instead, context-specific performance criteria are necessary to support reliable and clinically meaningful interpretation of ADAMTS13 results.

4.6.1 Defining APS at Clinical Decision Limits

Simulation-based modelling using the Milan criteria Model 1b offered a pragmatic and clinically oriented method for deriving APS at decision thresholds. Through *in silico* simulations according to previously published frameworks (86, 100, 225, 226), platform-specific differences in maximum allowable uncertainty were identified, with the Technozym® ELISA tolerating a higher MU at the critical 10 IU/dL threshold than the HemosIL AcuStar® for the same misclassification risk. This highlights the need to align APS with local analytical performance and population characteristics, rather than adopting universal thresholds. Importantly, this work demonstrates that the tolerable MU must be defined not only by analytical capability but also by the clinical risk associated with misclassification at key diagnostic cut-offs (227) (228). In this context, the simulation model enables laboratories to explicitly balance analytical limitations against acceptable clinical risk, a paradigm not well supported by traditional models.

4.6.2 Monitoring and Reference Change Values

In the context of longitudinal monitoring, BV derived APS provided an initial framework, but several limitations were evident. For ADAMTS13, BV estimates remain sparse and are derived from only two studies (108) (112). Furthermore, the biological regulation of ADAMTS13—unlike many measurands—is not under classical homeostatic control, calling into question the suitability of BV-

based APS (205), (101), (120). As an alternative, the number of distinct categories (NDC) offers a more practically applicable framework for evaluating analytical resolution at key clinical thresholds (229), and was indicated by population reference intervals using the refineR algorithm.

When assessing reference change value (RCV) models, significant variability was observed depending on the calculation method. Ratio-based and asymmetrical models (230), (231) demonstrated higher sensitivity to changes at low ADAMTS13 levels, which may aid early detection of relapse but also increase the risk of false positives. In contrast, conservative models (110, 113) reduced false alerts but risked underestimating clinically significant changes. These findings underscore the importance of selecting an RCV model tailored to the specific clinical context and patient baseline, rather than relying on a single universal model (232).

4.6.3 Real-World Data and Personalised Monitoring

The refineR-derived population variability highlights the potential for considering local data in the development of APS. While the estimated CVs from patient distributions do not meet the requirements to be classified as within-subject biological variation (CV_i), they provide a pragmatic approximation of total observed variability under routine testing conditions. When allowable imprecision thresholds (CV_a) were modelled as proportions of these CVs, the values were consistently higher than those derived from the published CV_i , particularly for the HemoSIL® AcuStar method.

This highlights some differences between traditional biological variation-based APS and the realities of clinical laboratory practice. In this context, using local CVs offers a practical, context-specific view of what level of analytical imprecision may be tolerable without disrupting clinical interpretation.

It is important to note, however, that these RefineR-derived distributions are not equivalent to clinical reference intervals. ADAMTS13 interpretation typically relies on clinical decision limits (e.g. 10 IU/dL) rather than population-based norms. This is supported by the index of individuality (II), calculated from the Kilercik *et al.* biological variation data as 1.32 (108), indicating moderate utility of

reference intervals but reinforcing the need for more individualised or decision-limit-based interpretive strategies (233-235). Accordingly, the RefineR outputs are best viewed as descriptors of local population variability rather than tools for direct clinical interpretation. This approach may be helpful where formal BV data are absent, outdated, or poorly matched to local testing practice. It also provides a platform for comparing method-specific variability under local conditions, as demonstrated by the higher CVs observed for the AcuStar[®] method.

Personalised reference intervals (prRI) represent a logical extension of this locally focussed concept (216, 218). The prRI approach, based on an individual's historical data during steady-state remission, enables a more accurate determination of clinically significant change (236). While not yet widely implemented, this method offers considerable promise, especially in chronic conditions like iTTP where relapse detection depends on trend interpretation rather than absolute thresholds (217) (237). The primary barrier to broader adoption remains technical: laboratory information systems must evolve to support individualised interpretive models and accommodate more complex reporting structures.

4.6.4 APS for Treatment Monitoring

For replacement therapies, particularly recombinant ADAMTS13, APS must reflect pharmacokinetic principles rather than diagnostic decision limits (214). APS models applied here mirrored those used in therapeutic drug monitoring (214),(213), allowing performance targets to be linked to expected dose–response characteristics. Treatment type and dosing frequency influenced allowable uncertainty, with extended half-life (EHL) products demonstrating more stable pharmacokinetics (224). However, current assay performance often fails to meet the precision requirements implied by these pharmacokinetic targets, suggesting a mismatch between clinical need and analytical capability. Moreover, the WHO standard for ADAMTS13 activity may not be suitable for quantifying recombinant products (WHO/BS/2014.2246), reinforcing the need for updated standards and calibration strategies.

4.6.5 Conclusions and Implications

Taken together, these findings support the central argument that APS for ADAMTS13 activity assays must be tailored to the clinical scenario. Diagnostic, monitoring, and treatment applications each demand distinct performance criteria. While this approach challenges the conventional notion of applying a single APS model across all contexts, it offers a more clinically coherent framework for managing ADAMTS13 measurement uncertainty.

The studies presented here also demonstrate the feasibility of implementing such a framework using local laboratory data and readily available modelling tools. While broader validation is necessary before universal adoption can be recommended, the conceptual model offers a blueprint for integrating analytical performance with clinical decision-making in a patient-centred and evidence-informed manner.

5 Chapter 5 – Method Comparison of ADAMTS13

Assays: Linking Verification to Clinical Risk

5.1 Introduction

Method comparison is a fundamental process in laboratory medicine, designed to assess whether different analytical platforms provide interchangeable results. In ADAMTS13 activity testing, however, variability between assay methods directly threatens clinical decision-making, particularly around critical diagnostic thresholds for TTP. Despite this, current verification practices remain analytically focused, relying on correlation coefficients, regression methods, and Bland-Altman plots that fail to reflect the clinical consequences of inter-assay differences.

A structured review of published method comparison studies for the HemosIL AcuStar® assay illustrates the limitations of conventional statistical approaches. High correlation coefficients (238), (239), (123) and regression analyses are frequently interpreted as evidence of assay comparability, while Bland-Altman plots, when presented, reveal substantial systematic biases and wide limits of agreement (240). Yet, these analytical differences are rarely contextualised against diagnostic thresholds or evaluated for their potential to misclassify patients near clinical decision points (see Appendix I). Deshpande *et. al.* addressed this clinical gap through a robust meta-analysis of 19 studies, confirming the high diagnostic accuracy of the AcuStar® and reinforcing its integration into diagnostic pathways for suspected TTP (241). However, their outcome-focused analysis did not explore the analytical variability underlying inter-assay differences, nor the impact of measurement uncertainty (MU) on patient-level diagnostic interpretation.

This chapter addresses that methodological gap by reframing method comparison as an assessment of clinical interpretative risk, explicitly incorporating MU into the analysis for the first time in the context of ADAMTS13 assays. Likelihood ratio (LR) analysis and error grid modelling, are applied as complementary tools to evaluate how assay differences—and their associated uncertainties—propagate into interpretative variability and potential patient harm.

MU is positioned as the unifying concept throughout this chapter. Rather than treating MU as a technical characteristic reported separately from verification, it is embedded within all analyses. Using local verification data from the AcuStar® platform, the chapter introduces a novel, proof-of-concept approach wherein clinical decision limits are re-evaluated based on their impact on diagnostic likelihood ratios. By dynamically adjusting cut-offs to preserve a positive LR (LR^+) above a clinically meaningful threshold (typically ≥ 10), diagnostic performance can be standardised even as analytical platforms vary. This represents a paradigm shift: instead of adhering to historically defined fixed cut-offs, laboratories can aim to maintain interpretative performance—defined as preservation of LR^+ —thereby harmonising diagnostic output across different analytical methods. This platform-agnostic approach offers a route to reduce diagnostic uncertainty and supports a transition towards Bayesian, probability-based reporting frameworks introduced later in this thesis (Chapter 9).

A second key methodological advance in this chapter is the development of an ADAMTS13-specific error grid. Adapting the principles of CLSI EP27-A, originally designed for glucose monitoring, this proof-of-concept tool visualises the clinical risk associated with assay discrepancies by categorising errors into defined risk zones. For the first time, MU is explicitly represented within the error grid through the introduction of guard bands, delineating regions where uncertainty alone may shift patient classification. This novel enhancement allows for a more accurate appraisal of misclassification risk near clinical thresholds and reframes result variability in terms of potential patient harm. Beyond evaluating local data, the error grid is applied retrospectively to critique

published method comparison studies, providing an alternative perspective on conclusions drawn using traditional analytical metrics.

Collectively, this chapter presents a structured, clinically focused approach to method comparison for ADAMTS13 assays. MU serves as both a technical parameter and a clinical risk factor, integrated into verification processes rather than treated as a separate post-analytical consideration. By applying likelihood ratios and error grids as complementary tools, this chapter offers a proof-of-concept framework that extends method comparison beyond analytical agreement towards clinical interpretability and patient safety. While focused on ADAMTS13, the principles established here are generalisable to other quantitative assays and lay the foundation for a fully Bayesian diagnostic framework developed in subsequent chapters.

5.2 Aims

This chapter aims to evaluate the clinical interpretative impact of inter-assay variability in ADAMTS13 activity testing by incorporating MU into method comparison analysis. Likelihood ratio analysis is applied to assess how cut-off adjustments influence diagnostic performance, supporting harmonisation across analytical platforms. An ADAMTS13-specific error grid, incorporating MU guard bands, is developed to visualise misclassification risk.

5.3 Limitations of Standard Verification Methods

5.3.1 Methods

5.3.1.1 *Data and Samples Used*

A total of 50 samples were selected from the stored repository in the haemostasis laboratory at CUH to reflect real-world requesting patterns for ADAMTS13 testing. These samples were stratified into clinically relevant ADAMTS13 activity ranges: 0–10 IU/dL, 10–30 IU/dL, 30–50 IU/dL, and >50 IU/dL to align with published guidance (3, 33, 93). The dataset included known immune TTP and congenital TTP cases, as well as samples from patients with both high and low clinical suspicion of TTP, which were later confirmed to be either other thrombotic microangiopathies or entirely normal. Additionally, samples from individuals with no clinical suspicion of TTP or TMA were included as controls. This selection ensured that the dataset was representative of the diverse range of clinical scenarios encountered in routine laboratory practice.

5.3.1.2 *Statistical Methods*

5.3.1.3 *Correlation Analysis*

Both Pearson's and Spearman's correlation coefficients were used to compare the strength and direction of association between methods. Pearson's assesses linear relationships and Spearman's captures monotonic trends. This provided a more comprehensive evaluation of correlation irrespective of distributional assumptions or potential outliers across clinically relevant ADAMTS13 activity ranges. Correlation analysis was conducted in R using base R functions and the stats package (R Team, Austria). Monte Carlo analysis was performed in the Google Colab[®] environment (Google, USA) using the following packages *pandas*, *matplotlib*, *scipy.stats* (*spearmanr* and *pearsonr* functions).

5.3.1.4 *Monte Carlo Simulations for MU Impact*

Monte Carlo simulations were performed to assess the impact of MU on correlation and regression estimates in a probabilistic approach. Simulations were conducted in Python using the Google Colab environment (<https://colab.research.google.com/>). The analysis used the following Python packages:

- scipy for statistical functions, including random sampling and correlation calculations.
- scipy.stats for probability distributions and statistical testing.
- numpy for numerical computations.
- matplotlib for data visualisation.
- pandas for data manipulation and structuring.

A total of 10,000 simulations were run for each scenario, incorporating varying levels of MU (10% and 30%) to quantify its effect on correlation and regression estimates. These simulations provided a probabilistic assessment of the variability introduced by MU in method verification studies.

5.3.2 Results

5.3.2.1 *Correlation at clinical decision limits in method comparison studies*

Acknowledging the low sample number (n=50), stratified analysis was performed to provide insight into potential differences in association between numerical results within example clinical ranges.

ADAMTS13 activity results between 0-10 IU/dL (or % in the case of the AcuStar[®]) represented results at the diagnostic threshold, whereas 10-30 IU/dL represented equivocal and relapse ranges.

ADAMTS13 activity results between 30-50 IU/dL (or %) represented partial remission, with full remission contained within the > 50 IU/dL (or %) subgroup.

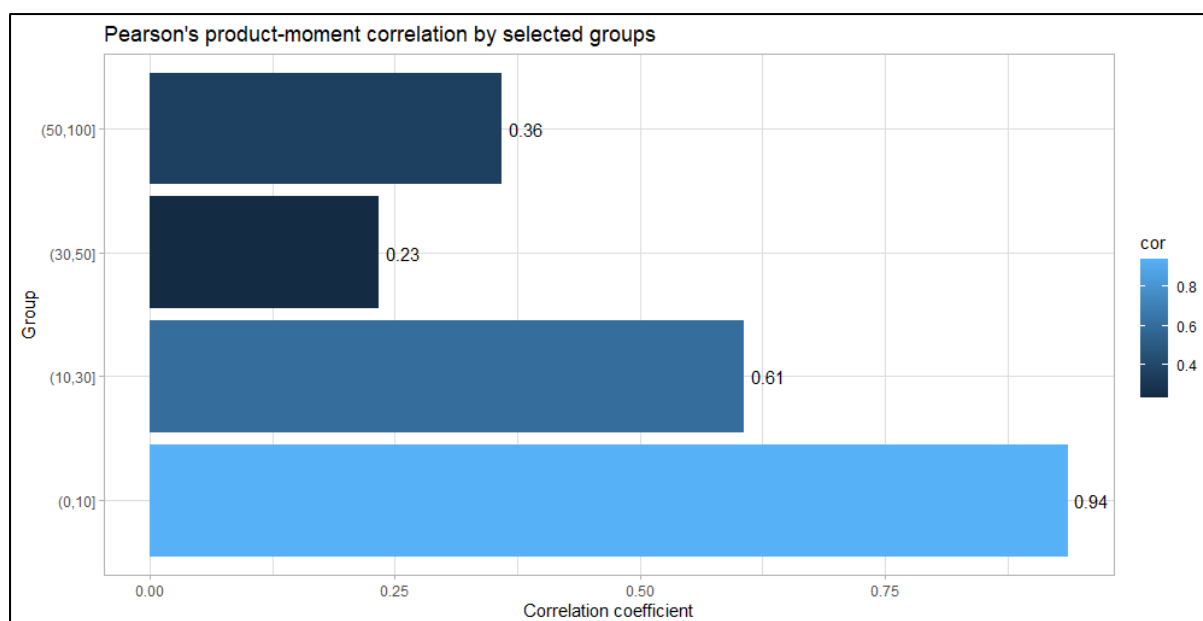


Figure 26: Stratified correlation of ADAMTS13 method comparison.

Pearsons correlation of a method comparison between Technozym[®] ELISA and HemosIL AcuStar[®], n=50. Data was stratified according to guideline based clinical decision limits relating to clinical outcome in patients with immune TTP. (1), (3); (93)

Pearson's correlation for the entire dataset was 0.902, but stratification by ADAMTS13 activity reveals significant variation across clinically relevant ranges. This stratification was necessary to evaluate correlation within the most critical decision-making areas, rather than relying on an overall correlation that may obscure clinically meaningful trends.

The highest correlation was observed in the <10 IU/dL range (Pearson's correlation 0.94, Figure 26), where ADAMTS13 activity is most relevant for diagnosing first-time TTP. This strong correlation is reassuring, as it suggests good agreement between the two methods in the range where clinical decisions about initiating plasma exchange, caplacizumab and immunosuppression are made. Notably, the correlation in this range was even higher than the overall dataset correlation, reinforcing the importance of stratified analysis.

However, correlation declines significantly as ADAMTS13 activity increases. In the 10–30 IU/dL range, correlation drops to a level that raises concerns about the interpretation of results in equivocal cases.

Given that patients in this range may be at risk of misclassification, there is potential for inappropriate clinical interventions if overall correlation alone is relied upon.

The decline is most pronounced in the 30–50 IU/dL range, where correlation falls to 0.23. This has implications for patients with a history of TTP in remission, particularly in scenarios where clinicians assess the safety of pregnancy planning based on ADAMTS13 levels (1, 93). Poor correlation in this range suggests that method agreement may be insufficient for reliably informing clinical decision-making.

Beyond 50 IU/dL, correlation remains low at 0.36, but the clinical impact is likely reduced, as these levels are typically associated with a lower risk of TTP recurrence and may not influence urgent treatment decisions. Nonetheless, the findings highlight the limitations of relying solely on correlation in method verification and reinforce the need for a broader statistical evaluation incorporating MU.

5.3.2.2 The impact of MU in correlation

Despite recommendations against its use (242), correlation coefficients are frequently cited as evidence of comparability in peer-reviewed publications (see Appendix I). However, correlation only describes the linear association between two analytical methods and does not assess their agreement. To date, the impact of MU on correlation interpretation has not been investigated.

This study hypothesises that incorporating MU into correlation analysis will significantly alter its interpretation. To test this, method comparison data from the CUH laboratory was analysed, assessing correlations between the Technozym® ELISA and AcuStar® methods in both the presence and absence of MU. Relative uncertainties of 10% and 30% were applied to each measured value and correlation coefficients were recalculated using 10,000 Monte Carlo simulations to assess their statistical significance. All data analysis was performed in Python within the Google Colab environment (<https://colab.research.google.com/>).

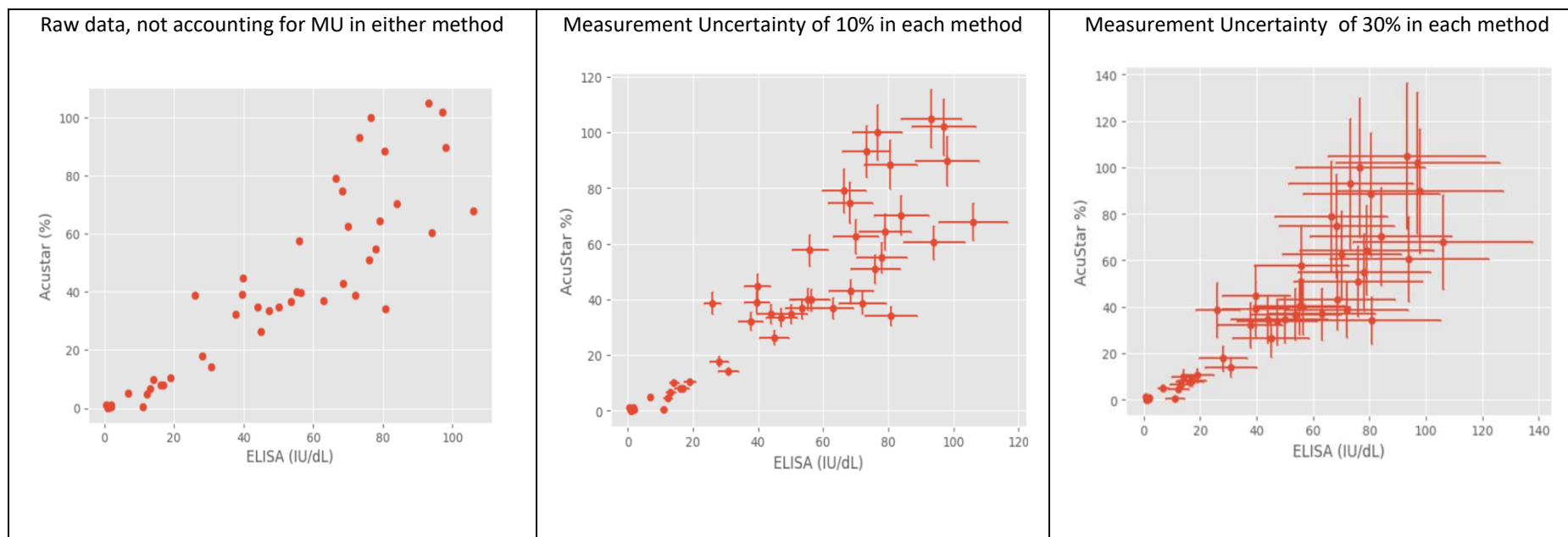


Figure 27: Pearson's correlation in ADAMTS13 activity method comparisons.

Scatter plots depicting the association of Technozym[®] ELISA and HemosIL AcuStar[®] during method comparison studies. (left) – Raw data, not accounting for MU in either method, (Middle) – Data points with added relative MU of 10% in each method, (Right) - Data points with added relative MU of 30% in each method. AcuStar[®] results are measured in % according to manufacturer's recommendations.

The method comparison between Technozym® ELISA and HemosIL AcuStar® showed a strong correlation at lower ADAMTS13 activity levels, particularly in the <10 IU/dL range, which is clinically critical for diagnosing TTP. However, correlation progressively weakened at higher concentrations (Figure 26, Figure 27), highlighting a loss of comparability between the methods.

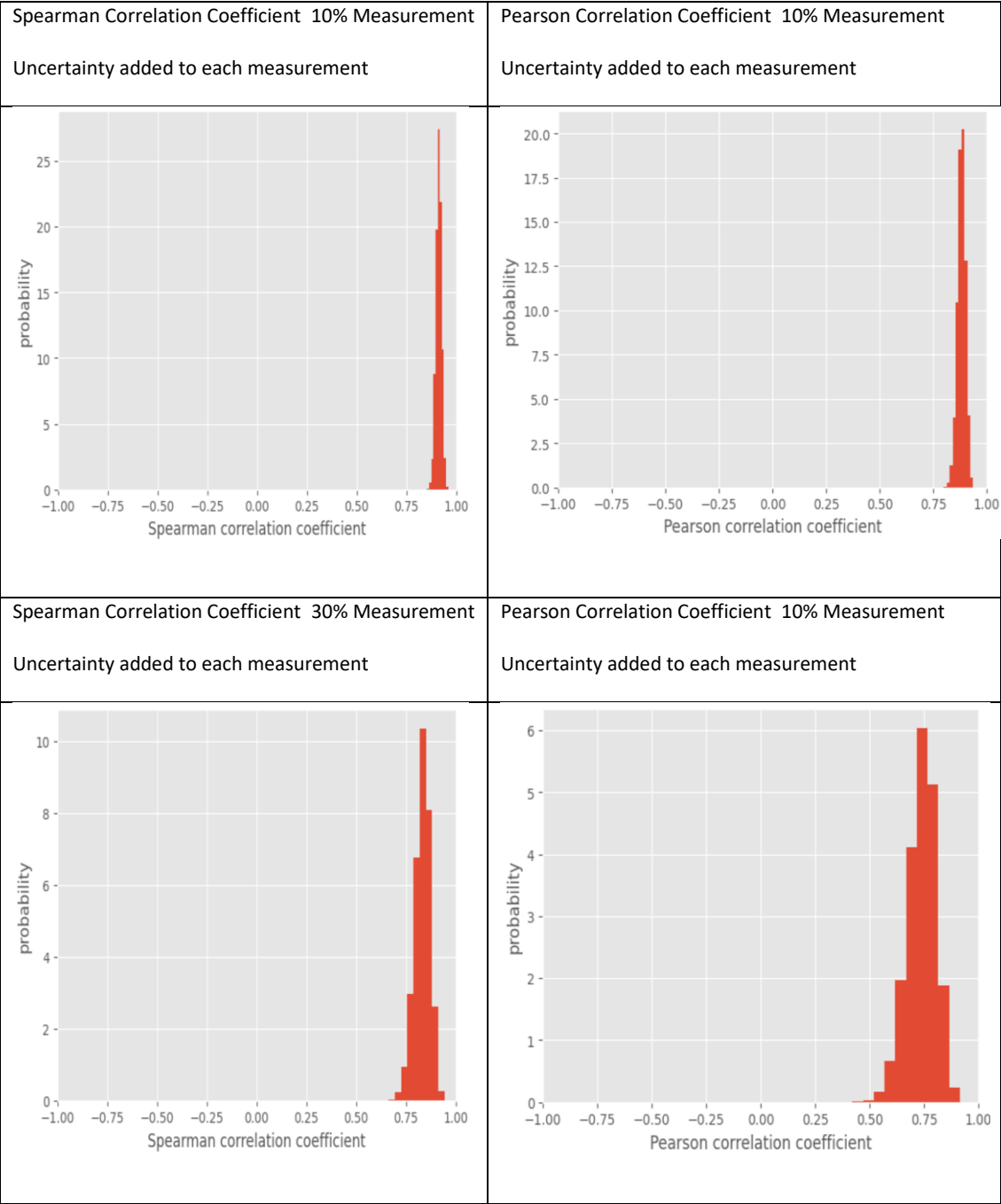


Figure 28: Correlation in ADAMTS13 method comparisons with measurement uncertainty.

Distribution of correlation coefficients (Spearman's left, Pearsons, right) realised after 10,000 Monte Carlo simulations of method comparison data values with 10% (Upper row) v 30% lower row added.

To assess the impact of MU on correlation, relative MU of 10% and 30% was added to each measured result. As seen in Figure 28, increasing MU significantly reduced the observed correlation and widened confidence intervals (

Figure 27). The error bars in

Figure 27 represent the uncertainty range of each measurement, assuming a Gaussian error distribution. Each data point, along with its uncertainty bounds, was used to generate 10,000 Monte Carlo simulations, producing a bivariate distribution centred on the original measurement value.

The introduction of MU systematically reduced the mean correlation coefficient, shifting it to the left in the correlation distribution plots (Figure 28). This effect was more pronounced in Pearson's correlation, which assumes a constant variance (homoscedasticity) and is therefore more sensitive to increasing measurement variability. Spearman's correlation, being rank-based, was less affected but still showed a decline, particularly at 30% MU (Figure 28, lower left).

Table 31: The impact of MU on correlation statistics.

Comparison of correlation statistics in the absence of MU or when adding 10% and 30% MU to Technozym[®] ELISA and HemosIL AcuStar[®]. MU = Measurement Uncertainty.

Correlation method	Metric	MU absent	10% MU	30% MU
Spearman's correlation	Mean	0.92	0.91	0.83
	SD		0.01	0.04
	Range		0.85-0.96	0.63-0.94
Pearsons Correlation	Mean	0.902	0.88	0.74
	SD		0.02	0.06

	Range		0.80-0.94	0.42-0.91

Table 31 summarises these effects. Without MU, Pearson's correlation for the dataset was 0.902.

When 10% MU was introduced, it fell to 0.88, and at 30% MU, it dropped to 0.74. Spearman's correlation showed a more moderate decline, from 0.92 (no MU) to 0.91 (10% MU) and 0.83 (30% MU). These results indicate that increasing uncertainty introduces greater variability in measured values, disproportionately affecting correlation coefficients.

The impact of this finding is significant for method verification. Laboratories often use a correlation threshold (e.g., 0.90 or 0.95) as an acceptance criterion for method comparability. However, this analysis demonstrates that MU alone can reduce correlation below this threshold, potentially leading to the rejection of otherwise clinically acceptable methods. This highlights the limitations of using correlation as the sole metric for verification and underscores the importance of incorporating other statistical measures, such as Bland-Altman analysis or regression-based agreement methods, to assess true method comparability.

5.3.2.3 *The impact of MU in regression*

Acceptance of method comparability using regression depends on obtaining parameter estimates (slope and intercept) within predefined acceptable limits (126, 243). However, the choice of regression method significantly influences these estimates and may lead to differing conclusions when parameter values vary between methods (244, 245). Deming regression is the preferred parametric approach in method comparison studies when MU is present, as it accounts for errors in both the independent and dependent variables (246). Unlike OLS regression, which assumes no measurement error in the independent variable, Deming regression adjusts for proportional errors by incorporating a variance ratio (λ), thereby providing more reliable slope estimates in the presence of MU.

If MU is sufficiently high, it can significantly impact the slope parameter, leading to underestimation of proportional error when using OLS regression (244, 247). In such cases, Deming regression mitigates bias by incorporating measurement variability into the model (248, 249).

To evaluate the impact of regression method selection, data from nine method comparison studies from a UK multicentre evaluation of AcuStar® implementation were analysed (250)). The variability in slope and intercept estimates was assessed using different regression techniques. Passing-Bablok regression, a non-parametric method that does not assume normality, was found to underestimate constant error compared to Deming regression, like the underestimation observed with OLS regression (Figure 29, left). Proportional bias was also underestimated using Passing-Bablok, though to a lesser extent than with OLS (Figure 29, right).

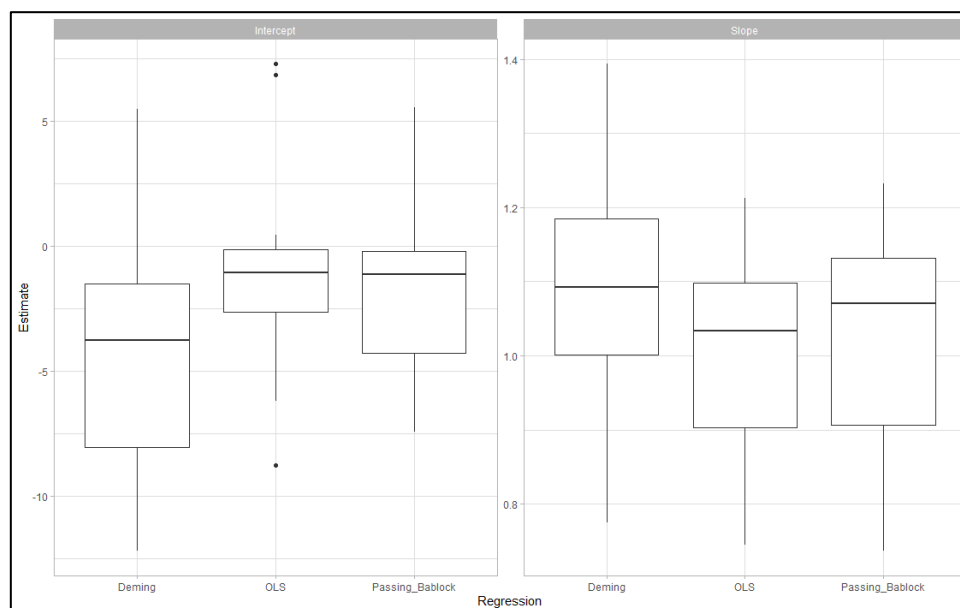


Figure 29: Regression in ADAMTS13 method comparisons with measurement uncertainty.

Comparison of regression parameter estimates using three different methods of regression. OLS – Ordinary Least Squares. Estimate is the statistical estimate of each of the slope and intercept parameters according to the heading of the figure.

Despite these differences, statistical significance was not reached for any of the comparisons between regression methods based on the Kruskal-Wallis test ($p=0.623$). This suggests that, within the variability observed in these datasets, the parameter differences were not large enough to be

statistically distinguishable at the predefined significance level ($p < 0.05$). However, while not statistically significant, the observed trends highlight important considerations when selecting regression methods for method comparison studies, particularly in cases where MU is a contributing factor.

Passing-Bablok underestimates constant error compared to Deming Regression similar to that seen with ordinary least squares (Figure 29, left). Proportional bias is underestimated using Passing-Bablok, compared to Deming regression but to a lesser extent than that seen in OLS (Figure 29, right).

5.4 Optimising clinical cut off from local verification data

5.4.1 Methods

5.4.1.1 Sample Selection and Study Design

5.4.1.2 Pre implementation – assay verification

Samples from the verification study of the HemosIL AcuStar[®] ADAMTS13 activity assay were reviewed (see section 5.3.1). Sample selection was intentionally structured to support analytical method comparison between the AcuStar[®] and Technozym[®] ELISA assays.

A total of 50 samples were retrospectively selected, comprising known positive and negative cases. Importantly, negative samples were deliberately chosen to include a proportion of results located close to the diagnostic decision threshold. This approach was designed to challenge the diagnostic classification of both assays across the full analytical measurement range, ensuring adequate representation of cases where diagnostic ambiguity is most likely to arise. While this selection strategy introduces spectrum bias and does not reflect the true prevalence or spectrum of disease in clinical practice, it was necessary to assess assay agreement under analytically demanding conditions. This design decision was made to provide a controlled assessment of assay performance, particularly in the borderline range where method comparison is most informative.

5.4.1.3 Post implementation – assay in clinical use

In contrast, the post-implementation dataset consisted of 63 consecutive, unselected patient samples submitted for routine testing, providing a real-world assessment of assay performance in the context of local requesting patterns and clinical use. Each sample was assayed on both platforms as a real time assessment of comparability in a patient cohort that represented the local population being tested.

For both datasets, samples were tested in parallel using the AcuStar[®] and Technozym[®] ELISA methods. Each sample was analysed using standard laboratory procedures according to

manufacturer instructions and local verification protocols. The primary aim was to assess diagnostic agreement between the two assays across a range of predefined cutoff values.

5.4.1.4 Cutoff Range Justification

To monitor the impact of changing the diagnostic threshold for TTP diagnosis, the range of cutoff thresholds evaluated in this study, spanning from 5 to 15 IU/dL, was selected to encompass the established diagnostic decision threshold of 10 IU/dL and to include both lower and higher values where diagnostic uncertainty is likely to be most pronounced. This range was chosen to evaluate how assay performance metrics, particularly LR^+ , sensitivity, specificity, and other diagnostic accuracy measures, varied across thresholds that surround the clinical decision point. Inclusion of cutoffs below and above the reference threshold enabled assessment of the effect of threshold variation on diagnostic performance, as well as to investigate the influence of MU within a clinically relevant interval.

5.4.1.5 Statistical Analysis

All statistical analyses were performed using the R statistical programming environment (R Team, Austria). Diagnostic accuracy metrics were calculated using a custom script using the built-in functions of the dplyr package (Wickham H, François R, Henry L, Müller K, Vaughan D (2025). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4.9000, , <https://github.com/tidyverse/dplyr>.). The script aggregated results by sample group and calculated sensitivity, specificity, PPV, NPV, false negative rate (FNR), false positive rate (FPR), false discovery rate (FDR), false omission rate (FOR), LR^+ , negative likelihood ratio (LR^-), predictive threshold (PT), test strength (TS), accuracy, balanced accuracy (BA), diagnostic odds ratio (DOR), and F1 score.

Each patient sample assayed on the AcuStar[®] platform was compared to the chosen clinical decision limit and classified as positive or negative. This result was then compared to the classification from the Technozym[®] ELISA method using 10 IU/dL. Changes in performance due to an increase or

decrease in false positives or negatives were reflected in the respective diagnostic accuracy measures.

5.4.1.6 Handling of Missing Data and Statistical Estimates

All analyses were conducted using complete data; no missing values were present in the dataset for the variables required to calculate diagnostic performance metrics. Positive and negative likelihood ratios (LR^+ and LR^-) were calculated based on observed sensitivity and specificity values at each cutoff threshold. Confidence intervals for LR^+ and LR^- were not calculated in this analysis, as the primary aim was to illustrate the impact of MU on point estimates of LR^+ derived from locally verified data. However, the variability introduced by analytical imprecision was explicitly represented through the incorporation of MU into the range of LR^+ values associated with individual test results.

5.4.2 Results

5.4.2.1 The impact on detecting TTP using ADAMTS13 activity on the AcuStar[®]

Changing cut off across the 5-15 % range had very little impact on the sensitivity of the AcuStar[®] method to accurately identify true TTP (or at least a result <10 IU/dL by Technozym[®] ELISA). It wasn't until the cut-off was dropped to 5% that both datasets deviated from 100% sensitivity (90% for pre implementation, 80% for post implementation, data not shown). Retaining 100% sensitivity across all diagnostic cut-offs from 5 to 15 IU/dL indicates that the test successfully identified every true positive case within the studied populations. This level of performance suggests that, regardless of where the threshold was set, the assay did not miss any individuals with the condition, eliminating false negatives entirely.

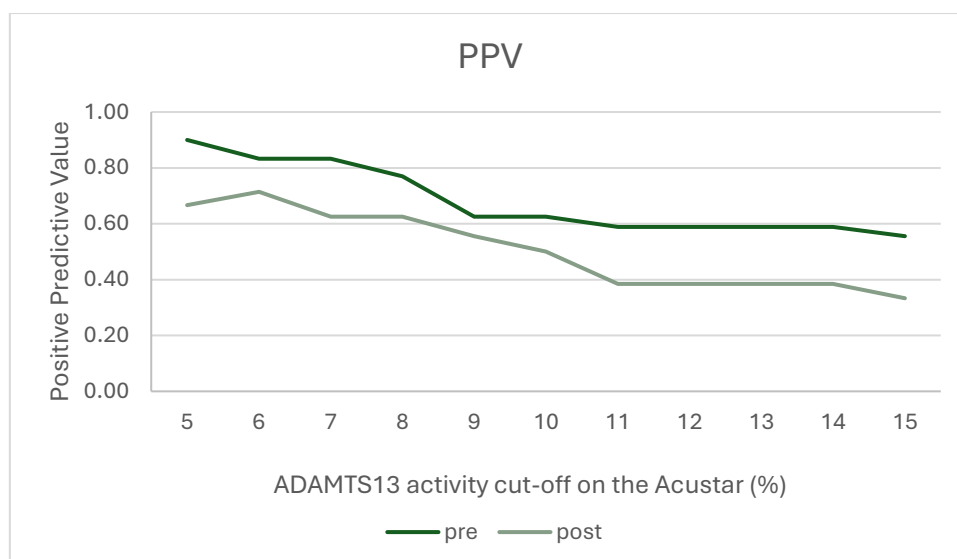


Figure 30: Cut-off determined positive predictive value.

Positive Predictive Value (PPV) across diagnostic cut off values from 5-15% on the HemoSIL[®] AcuStar

While sensitivity remained high, the PPV progressively declined as the diagnostic cut-off increased (Figure 30). This decline reflects the increasing number of false positives being captured at higher thresholds: as more patients without disease are classified as positive the proportion of true positives among all positive results decreases.

5.4.2.2 The impact on false positive detection

Specificity is highest at the lowest cut-offs in both datasets, with values of 0.98 (pre) and 0.97 (post) at cutoff 5 (Figure 31). As the cut-off increases, specificity declines progressively, indicating an increasing false positive rate.

In the pre-implementation dataset, specificity drops more sharply, falling to 0.80 by cutoff 15. In contrast, the post-implementation dataset maintains relatively higher and more stable specificity, particularly between cut-offs 6 and 10.

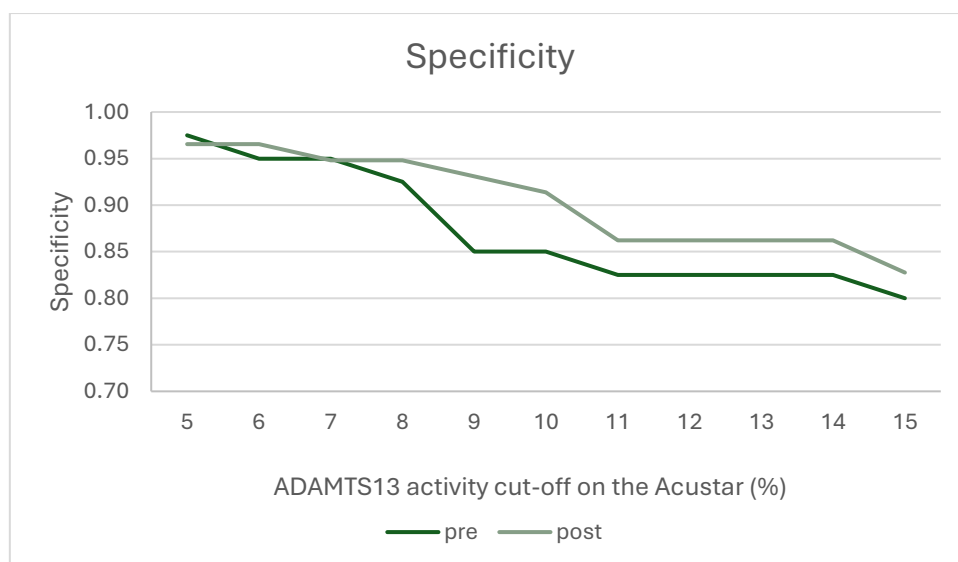


Figure 31: Cut-off determined specificity.

Specificity across diagnostic cut-offs (5–15 %) in pre- and post-implementation datasets.

As the cutoff increases, the false positive rate (FPR) rises progressively in both datasets, reflecting a steady decline in test specificity at higher thresholds (Figure 32).

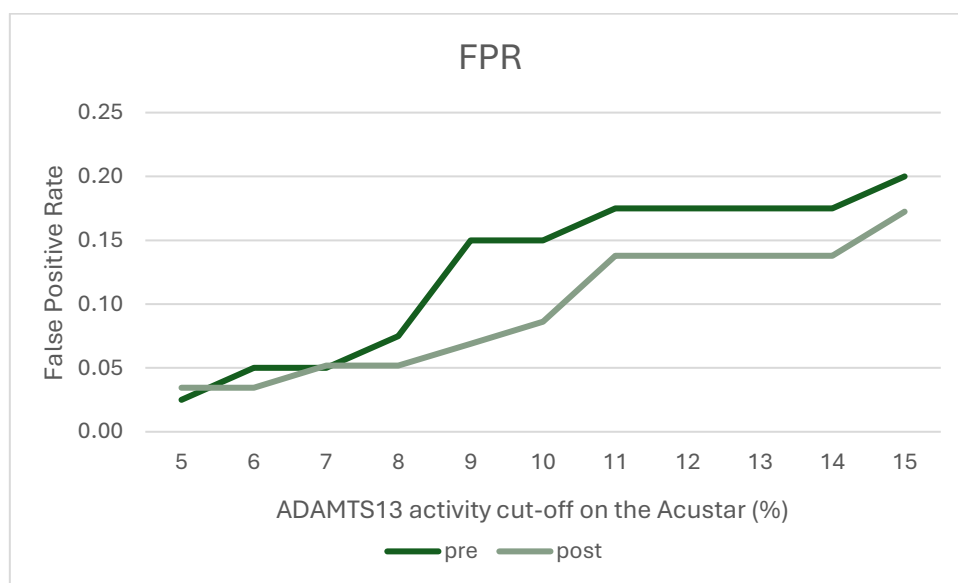


Figure 32: Cut-off determined false positive rate.

False positive rate (FPR) across diagnostic cut-offs (5–15 %) in pre- and post-implementation datasets.

In the pre-implementation data, the FPR remains ≤ 0.05 up to a cutoff of 7, after which it increases more noticeably, reaching 0.15 at cutoff 9 and plateauing at 0.18 between cutoffs 11 and 14, before rising slightly again to 0.20 at cutoff 15 (Figure 32). In contrast, the post-implementation FPR remains slightly lower across the range, peaking at 0.17 at cutoff 15.

5.4.2.3 The impact on likelihood ratio from changing cut-off

Table 32 summarises Likelihood ratios (positive and negative) between the AcuStar[®] and ELISA methods across the same range of cut-offs. Performance metrics were calculated at cutoff thresholds ranging from 5 to 15 IU/dL, with the reference cutoff (for the Technozym[®] ELISA) set at 10 IU/dL.

Table 32: Likelihood ratio of ADAMTS13 activity across varying decision limits.

Diagnostic performance metrics for the AcuStar assay compared to ELISA across varying cutoff thresholds. LR+ = Positive Likelihood Ratio, LR- = Negative Likelihood Ratio

Cutoff	LR ⁺ Pre	LR ⁻ Pre	LR ⁺ Post	LR ⁻ Post
5	36.00	0.10	23.20	0.21
6	20.00	0.00	29.00	0.00
7	20.00	0.00	19.33	0.00
8	13.33	0.00	19.33	0.00
9	6.67	0.00	14.50	0.00
10	6.67	0.00	11.60	0.00
11–14	5.71	0.00	7.25	0.00
15	5.00	0.00	5.80	0.00

At the lowest cutoff of 5 IU/dL, the pre-implementation data show the highest LR⁺ value at 36.00, indicating very strong diagnostic evidence when a positive result is obtained. The corresponding LR⁻ at this cutoff is 0.10, suggesting good but not perfect sensitivity (0.90, data not shown). In the post-implementation dataset, the LR⁺ is slightly lower at 23.20, and the LR⁻ is slightly higher at 0.21,

reflecting a slight reduction in diagnostic strength and sensitivity under real-world conditions (sensitivity – 0.80).

From cutoff 6 IU/dL onwards, both datasets exhibit an LR^- value of 0.00, indicating that all true positive cases were detected by the assay, with no false negatives. This suggests that the assay achieves perfect sensitivity at or above this threshold in both settings. Notably, the LR^+ values remain high at cutoffs of 6 and 7 in both datasets, with values of 20.00 in the pre-implementation data and a peak of 29.00 in the post-implementation data at cutoff 6. These values reflect strong diagnostic evidence when test results are positive at these thresholds.

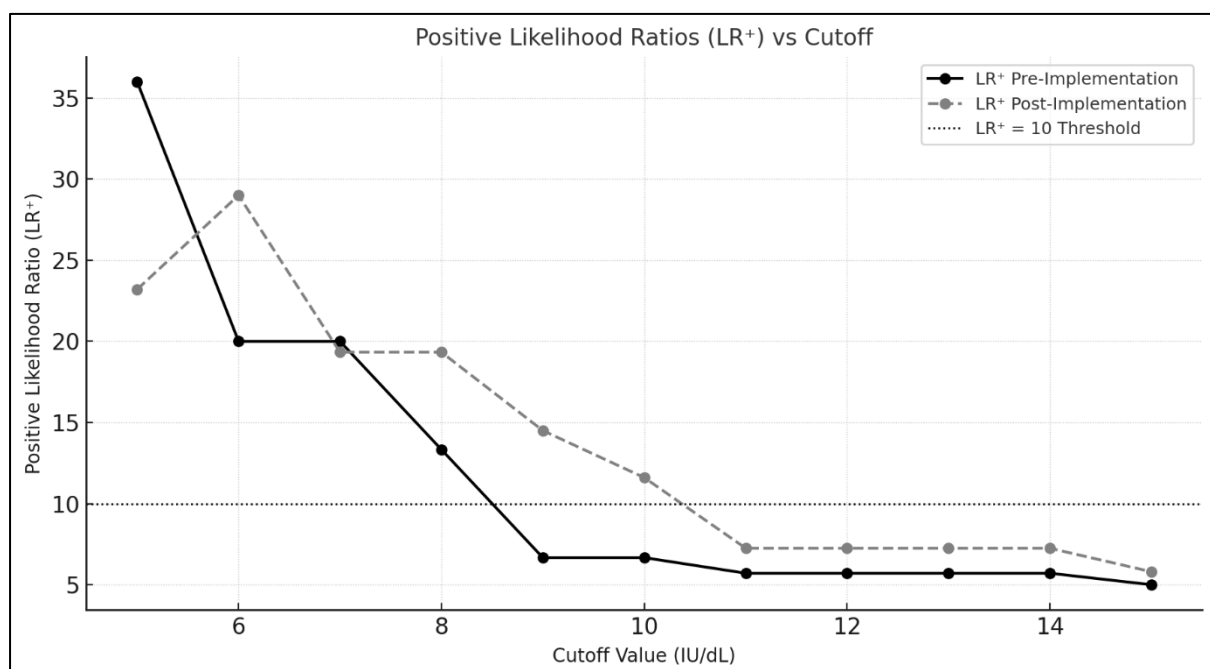


Figure 33: Cut-off determined positive likelihood ratio.

Positive Likelihood Ratio (LR^+) values for the pre-implementation group (solid black line) and post-implementation group (dashed grey line) across cutoff values ranging from 5 to 15 IU/dL. A horizontal dotted line at $LR^+ = 10$ is included to represent the commonly accepted threshold for strong diagnostic evidence. The pre-implementation data were derived from selected known positive and negative samples, while the post-implementation data were based on consecutive prospective samples submitted for routine testing.

As the cutoff increases from 8 to 15 IU/dL, LR^+ values progressively decline in both datasets (Figure 33). By cutoff 10, which is commonly used as the reference diagnostic threshold, LR^+ has fallen to 6.67 in the pre-implementation dataset and 11.60 in the post-implementation dataset. Beyond this point, diagnostic strength diminishes further, with LR^+ reaching 5.00 in the pre-implementation and 5.80 in the post-implementation dataset at a cutoff of 15.

Despite this downward trend, LR^- values remain consistently at 0.00 across all cutoffs above 5, reinforcing the high sensitivity of the test. However, this preservation of sensitivity comes at the cost of reduced specificity as cutoffs increase.

Lower cutoffs, particularly in the range of 6–7 IU/dL, offer the most favourable balance between high LR^+ and perfect sensitivity ($LR^- = 0.00$). This balance reflects strong diagnostic performance, as LR^+ values remain well above the commonly accepted threshold of 10 for strong evidence. Although the post-implementation values differ slightly from those in the more controlled pre-implementation setting, the overall performance trend remains consistent, reinforcing the value of using lower thresholds to optimise diagnostic confidence. This consistency is important because the diagnostic interpretation often assumed by clinicians—derived from assay verification data—is largely preserved following implementation into real-world use.

For both datasets, LR^+ values exceed the threshold of 10 only at lower cutoff values (Figure 33). In the pre-implementation dataset (black line), LR^+ remained above 10 until the cutoff exceeded 8 IU/dL. In the post-implementation dataset (grey line), this threshold persisted until the 11 IU/dL cutoff, after which LR^+ declined below 10, reflecting a diminished capacity to confidently rule in disease beyond these points.

The verification (pre-implementation) data consistently yielded higher LR^+ values at each cutoff compared to the post-implementation data. This disparity likely results from the controlled nature of the pre-implementation sample selection, where known positive cases typically yielded markedly low values—well below any of the cutoffs tested. Notably, patients presenting with initial episodes of

iTTP often have ADAMTS13 activity levels <1 IU/dL (or $<1\%$ on the AcuStar®), which can inflate estimates of diagnostic performance when used in verification studies. In contrast, the post-implementation dataset, derived from unselected, consecutive clinical samples, includes greater variability and more results in the equivocal range, providing a more realistic reflection of routine diagnostic performance.

5.5 ADAMTS13 Error Grid: Visualising Misclassification Risk

5.5.1 Methods

5.5.1.1 *Defining the error grid structure*

The error grid concept originates from Clarke's glucose error grid (129, 130), which classifies discrepancies based on clinical risk rather than statistical deviation. In this context, the ADAMTS13 error grid will be customised for method comparison studies involving ADAMTS13 assays, with clinically relevant decision thresholds set at:

- 10 IU/dL: indicating severe ADAMTS13 deficiency and a high risk of acute TTP according to current international guidance (1); (3).
- 20 IU/dL: representing moderate deficiency, typically used for monitoring patients at risk of relapse or requiring immunosuppressive therapy according to state of the art clinical outcome recommendations (93)
- 50 IU/dL: as a rough approximation of the lower boundary of the normal reference range, relevant for long-term follow-up, including scenarios such as pregnancy planning.

The reference method for this comparison will be the Technozym® ELISA, evaluated against two rapid methods: the HemoSIL AcuStar® and Technofluor® assays. To enhance interpretability and clinical relevance, the error grid will be structured into discrete zones reflecting varying degrees of clinical risk. In addition to standard zones representing acceptable or unacceptable agreement, guard bands—transition zones around clinical thresholds—will be introduced to account for MU. This will prevent misclassification due to inherent analytical variability near critical decision points.

Ultimately, this study aims to establish a clinically driven error grid model tailored to ADAMTS13 testing. The resulting framework will support method harmonisation, regulatory submissions, and

post-market surveillance by providing a standardised approach for evaluating the clinical validity of assay comparisons.

The error grid will be divided into four distinct clinical risk zones, adapted from Clarke's glucose error grid structure (Table 33, below):

Table 33: Definition of clinical zones for comparison when deriving the error grid

Zone	Definition
Zone A	Results fall within an acceptable clinical range, meaning the rapid method results are clinically interchangeable with the manual reference method.
Zone B	Results show minor deviations that are unlikely to cause harm but may still be clinically relevant.
Zone C	Results exhibit potentially clinically significant discrepancies, warranting further investigation or intervention.
Zone D	Results are unacceptable, with errors large enough to pose a risk of patient harm.

5.5.1.2 Guard Banding to represent MU impact

MU will be incorporated into the error grid to account for potential misclassification risks at the zone boundaries. This is essential to ensure that method comparisons reflect true analytical performance rather than limitations arising from random variability.

In this study, guard bands of fixed absolute width ($\pm 5\%$) will be applied across the entire ADAMTS13 measurement range to define transition zones between clinical risk categories. A standardised 5% MU was applied to define grey zones within the error grid where decision boundaries intersect. This MU value represents only the repeatability component of uncertainty, excluding factors such as calibration or between-run effects. Although the full quantified MU is approximately 15%, and allowable performance specifications extend to 20%, a lower value was selected to isolate the

influence of random variation under stable conditions during method comparison. This approach allows for a conservative but visually interpretable representation of potential misclassification risk due to repeatability alone. These grey zones indicated by guard bands serve to identify results that fall close to clinically critical thresholds and may be at risk of misclassification due to inherent assay imprecision.

The decision to apply a constant-width guard band, rather than a dynamic band reflecting changing MU across the measurement range, is based on some model assumptions introduced here. A constant-width approach facilitates practical implementation and visual clarity, particularly for assay verification, quality assurance, and regulatory evaluation. Dynamic MU modelling, while more reflective of heteroscedasticity, can introduce complexity without necessarily improving clinical decision accuracy when specific thresholds are used. The assays under investigation (e.g. Technozym® ELISA, AcuStar®, Technofluor®) are designed to maintain analytical performance across the measurement range. In verification studies, the observed MU does not vary substantially at the designated decision points. Therefore, a fixed guard band width of $\pm 5\%$ adequately captures the typical uncertainty expected in routine practice, without requiring dynamic adjustment.

To address misclassification risks where MU exceeds these guard bands, a conditional acceptance category will be introduced. This describes results that fall within a transition zone between risk categories and may be considered clinically acceptable, provided that MU remains within predefined tolerances. If MU exceeds this threshold, the reliability of method comparisons may be compromised, particularly in real-world clinical implementation. Additionally, borderline values between zones B, C and D all require individual inspection. All results where potentially clinically significant discrepancies exist all must be reviewed on their own merits in a clinical context.

5.5.2 Results

5.5.2.1 Rationale for requiring Error grids – local verification example

The local verification data comparing the Technozym® ELISA and the AcuStar® rapid assay demonstrates inconsistent performance across the measurement range, highlighting potential clinical implications (Figure 34). This variability suggests that a single overall performance metric may not adequately capture the most clinically significant differences, particularly where the risk of patient harm is greatest.

The data indicate that the AcuStar® assay tends to underestimate ADAMTS13 activity, particularly at higher levels. However, at these higher concentrations, small deviations have less clinical impact compared to discrepancies near the critical threshold of 10 IU/dL, where diagnostic and treatment decisions are most urgent.

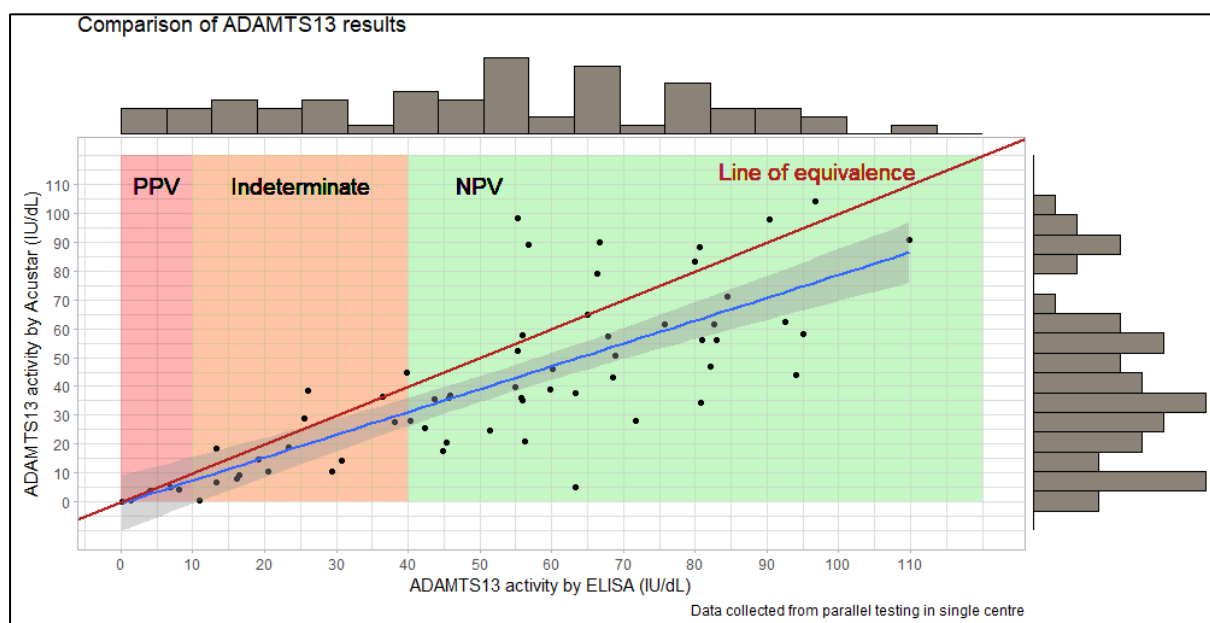


Figure 34: ADAMTS13 method comparison results.

Regression based comparison of two ADAMTS13 activity assay measurement systems ELISA – Technozym® ELISA, AcuStar –

HemoSIL AcuStar. PPV – Positive predictive value, NPV – Negative Predictive Value

Figure 34 above illustrates the diagnostic accuracy measures used to assess clinical utility at different ADAMTS13 levels. At very low levels (<10 IU/dL), a high PPV is essential, as a confirmed diagnosis of TTP requires urgent intervention. At normal levels (>50 IU/dL), the assay is primarily used to rule out TTP, where a negative result prevents unnecessary and potentially harmful treatments. In the intermediate range (10–50 IU/dL), diagnostic requirements are influenced by pre-test probability, other clinical results, and scoring systems, requiring careful interpretation of assay discrepancies.

This figure underscores that the consequences of assay discrepancies vary depending on ADAMTS13 concentration. Therefore, the use of an error grid—directly linking discrepancies to their clinical impact—provides a valuable framework for assessing method comparison results and ensuring clinically meaningful performance evaluation.

5.5.2.2 Defining the error grid based on clinical risk of harm and clinical guidelines

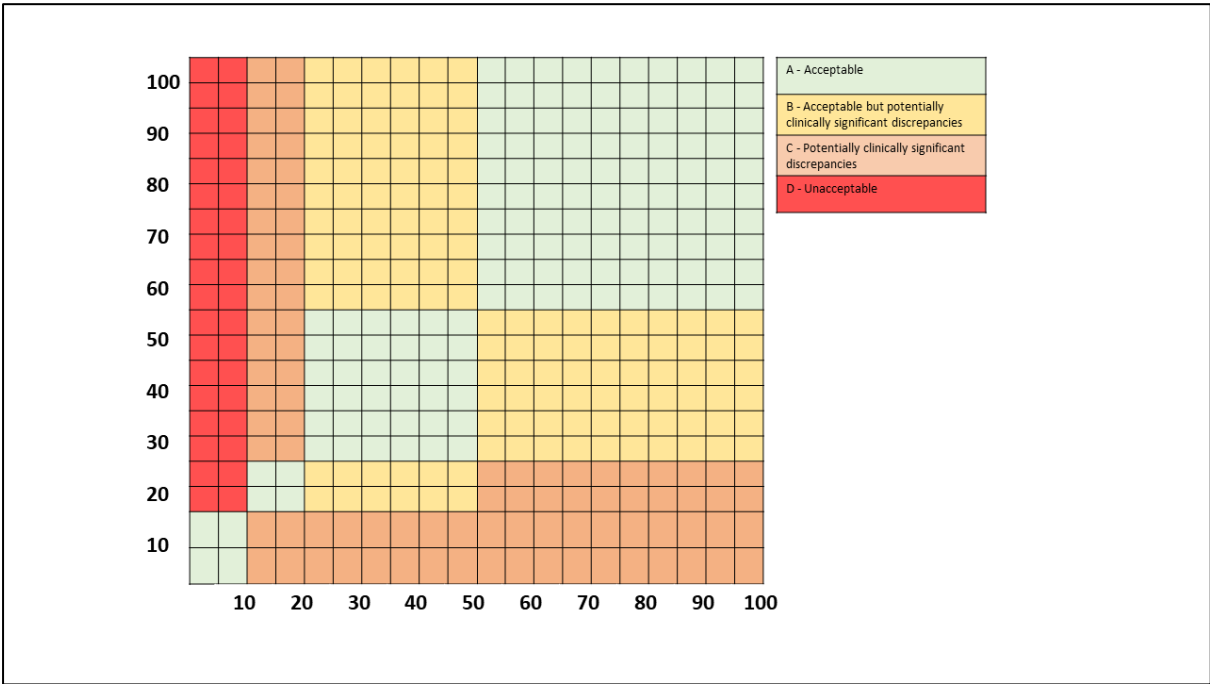


Figure 35: ADAMTS13 activity Error Grid.

Zones are defined based upon data from clinical guidelines indicating important areas for clinical decisions including diagnosis and monitoring

Figure 35, above illustrates the constructed zones of the ADAMTS13 error grid, developed based on clinical guidelines. With the Technozym® ELISA as the reference method plotted on the x-axis, the error grid highlights asymmetry in method discrepancies, particularly when comparing the AcuStar® rapid assay. While the ELISA is treated as the reference method in this analysis, it is important to acknowledge that it does not hold metrological reference status in the strictest sense.

At the critical ≤ 10 IU/dL threshold (Zone A), maintaining clinical accuracy is paramount, as this level represents the diagnostic cut-off for severe ADAMTS13 deficiency, where failure to identify true cases of TTP could have life-threatening consequences. In this region, the error grid demands high concordance between the ELISA and rapid methods, with minimal tolerance for discrepancies. Above 10 IU/dL, the grid allows for increased tolerance in method agreement, as the risk of immediate clinical harm decreases and the clinical significance of minor discrepancies is reduced.

The implications of misclassification within the error grid vary depending on ADAMTS13 activity levels. If a result obtained by ELISA exceeds 10 IU/dL but is misclassified as <10 IU/dL by the rapid method, the patient could be incorrectly diagnosed with severe ADAMTS13 deficiency, leading to unnecessary plasma exchange and caplacizumab therapy. This misclassification not only exposes the patient to the risks associated with blood product transfusion but also increases the likelihood of acquired von Willebrand syndrome (aVWS) due to inhibition of the VWF A1/GpIb α interaction. Conversely, if a true TTP patient with an ADAMTS13 activity <10 IU/dL is incorrectly classified as >10 IU/dL, the failure to recognise their severe deficiency could result in delayed or absent treatment, placing them at significant risk of clinical deterioration. Due to the critical nature of this error, the error grid adopts a strict classification approach in this region, rejecting any false-negative results as unacceptable.

In the intermediate range between 10 IU/dL and 50 IU/dL, discrepancies between the ELISA and rapid methods are more clinically acceptable but still require careful evaluation. While immediate patient harm is less likely in this region, misclassification can still carry significant long-term

implications, particularly for patients previously diagnosed with TTP who are considering pregnancy. Accurate ADAMTS13 measurement is essential in these cases, as misclassification could impact clinical decision-making regarding pregnancy safety. The error grid accounts for these factors by recognising that results in this range require individual review rather than outright rejection, ensuring that method comparison assessments reflect both analytical performance and real-world clinical considerations.

5.5.2.3 Defining the error grid based incorporating MU as guard bands

It is recognised that the traditional depiction of an error grid typically employs hard boundaries between zones, which do not account for the risk of MU falsely classifying results. To address this limitation, a guard banding approach was implemented, where the width of the guard band is determined by the repeatability portion of combined MU (5%).

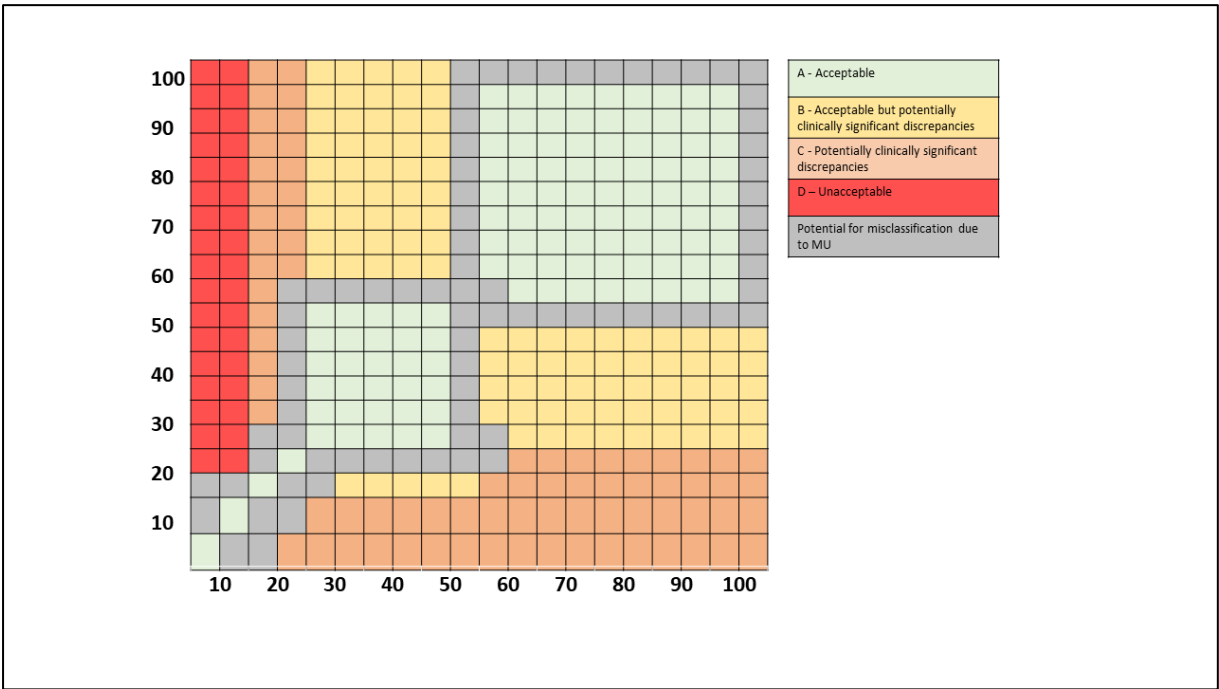


Figure 36: ADAMTS13 activity Error Grid with guard banding. Guard bands represent an absolute measurement uncertainty of 5% across the measurement range.

As shown in Figure 36, the regions with >95% confidence of agreement between methods are reduced in areas where MU is higher, particularly in a clinical context. Consequently, results that fall within these grey transition zones may be classified into either adjacent zone, necessitating careful interpretation. This is particularly critical when considering adjacent zones with significantly different clinical implications, such as Zone D (red), where ELISA results are <10 IU/dL but the rapid test reports >10 IU/dL. Misclassification in this area could lead to incorrect treatment decisions, including failure to diagnose or inappropriate intervention.

Samples falling within this uncertainty range should be carefully reviewed, with consideration given to the potential consequences of misclassification on patient management.

5.5.2.4 Error Grid Analysis used for verification of rapid ADAMTS13 activity assays

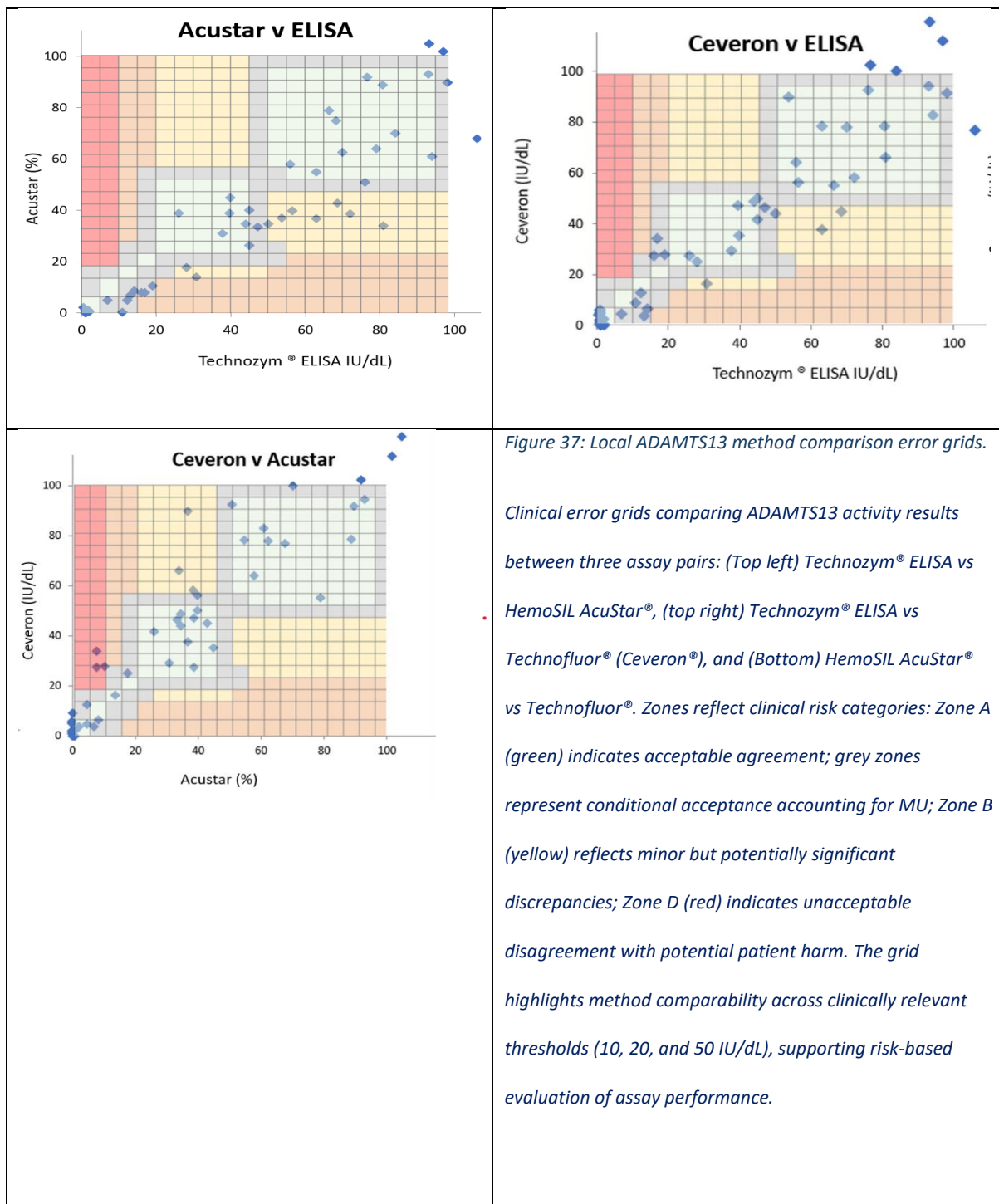
To evaluate the clinical utility of the ADAMTS13 error grid, three real-world method comparisons were assessed using the newly developed framework. Each error grid chart plots results from two methods: one serving as the reference, and the other as the test method. Zones were defined by clinical risk and transition guard bands accounting for MU. Results outside the defined grid boundary (100 IU/dL) were excluded from zone classification as they exceed the clinically relevant range for TTP diagnosis and monitoring.

5.5.2.5 Technozym® ELISA vs HemoSIL AcuStar®

The comparison between Technozym® ELISA (reference method) and the HemoSIL AcuStar® (incoming method) revealed strong clinical concordance. The majority of results fell within Zone A and the adjacent grey guard bands, representing either full agreement or conditional acceptance due to proximity to zone boundaries (Figure 37, top).

Eight results were classified in Zone B, denoting minor but potentially clinically relevant discrepancies. These split into two subtypes. One cluster demonstrated lower AcuStar results (~20%

lower) in samples where ELISA results ranged up to 50 IU/dL. In these cases, the AcuStar may misleadingly suggest borderline TTP-level activity, potentially affecting decisions about immunosuppressive treatment or monitoring. The other group of Zone B discrepancies was observed at higher measurement levels, with AcuStar lower than ELISA. While these lie outside the most clinically sensitive range (<10 IU/dL), they may still influence decisions about remission monitoring or pregnancy planning in patients with a history of iTTP. Notably, at the critical threshold of 10 IU/dL, the AcuStar[®] results generally showed good clinical agreement with the ELISA, though a trend toward lower numerical values was observed. These differences, however, fell within the conditional acceptance zone, supporting the use of AcuStar in urgent diagnostic contexts.



This pattern reflects prior findings that the AcuStar assay tends to underestimate ADAMTS13 activity, but the risk of clinically significant misclassification in this dataset appears low, particularly at the diagnostic threshold. The grid effectively highlights these trends, reinforcing the importance of zone-specific evaluation rather than global performance metrics.

5.5.2.6 *Technozym® ELISA vs Technofluor® (Ceveron® Platform)*

The comparison between the Technozym® ELISA and the Technofluor® assay, both manufactured by the same vendor and run on the Ceveron® analyser, demonstrated excellent agreement across the clinically relevant range. Only three results fell within Zone B (Figure 37, middle), indicating a high degree of consistency between these methods. Agreement was particularly robust around the critical 10 IU/dL threshold, where diagnostic accuracy is paramount. Some samples produced very high results on the Ceveron®, exceeding the 100 IU/dL upper boundary of the grid. While these are outside the evaluable range, they are not clinically concerning, as they represent supranormal results without diagnostic ambiguity.

The strong alignment is likely attributed to harmonisation in reagents and instrumentation, underscoring the value of platform consistency in method standardisation.

5.5.2.7 *HemoSIL AcuStar® vs Technofluor®*

This comparison between two newer rapid methods (AcuStar® and Technofluor®) yielded reasonable agreement overall, but also highlighted important discrepancies in the context of patient care. Three results were classified in Zone D (Figure 37, bottom), representing clinically unacceptable differences. In each case, the AcuStar® returned unexpectedly low values, while the Technofluor® reported significantly higher activity levels. These samples originated from low pre-test probability TTP cases, and it is plausible that interfering substances or other biological factors (e.g. lipids, antibodies) introduced a negative bias in the AcuStar® assay. Without access to the patients' full clinical context or confirmatory reference results, it is not possible to definitively determine which assay reflects the true physiological state in these cases.

The presence of Zone D misclassification—particularly at levels where clinical decisions hinge on accurate detection of severe ADAMTS13 deficiency—emphasises the need for cautious interpretation of discordant results between rapid platforms. The error grid provides a valuable

visual and analytical tool for flagging such critical discrepancies, which may otherwise be obscured in numerical comparisons.

5.5.2.8 Error Grid Analysis used for published data on verification of rapid ADAMTS13 activity assays

To further evaluate the robustness and generalisability of the ADAMTS13 error grid, published method comparison studies were reviewed and interpreted using the established zone classification framework. The error grid enabled stratification of assay discrepancies according to clinical risk, allowing a more nuanced interpretation beyond statistical agreement.

Favaloro *et al.*'s large-scale verification study revealed a notable proportion of results in Zone B (Figure 38, top left from (251), mirroring trends observed in our local comparison between the Technozym® ELISA and the HemoSIL AcuStar®. The study specifically identified misclassification risks at the critical thresholds of 10 and 40 IU/dL, consistent with the decision limits embedded in our error grid.

A key finding was the occurrence of Zone C discrepancies, absent from our local dataset (Figure 38 top-left, below). 10 patient samples demonstrated ELISA results in the 20–50 IU/dL range, while the AcuStar® returned values below 10 IU/dL. Such discrepancies are clinically significant, as they may trigger inappropriate interventions—such as plasma exchange or caplacizumab therapy—based on falsely low ADAMTS13 activity. These cases highlight a greater potential for clinical harm than observed locally and reinforce the importance of contextualising numerical agreement with clinical thresholds.

The study also reaffirmed a consistent trend of numerically lower values reported by AcuStar®, a feature seen in multiple evaluations, including that presented in this chapter.

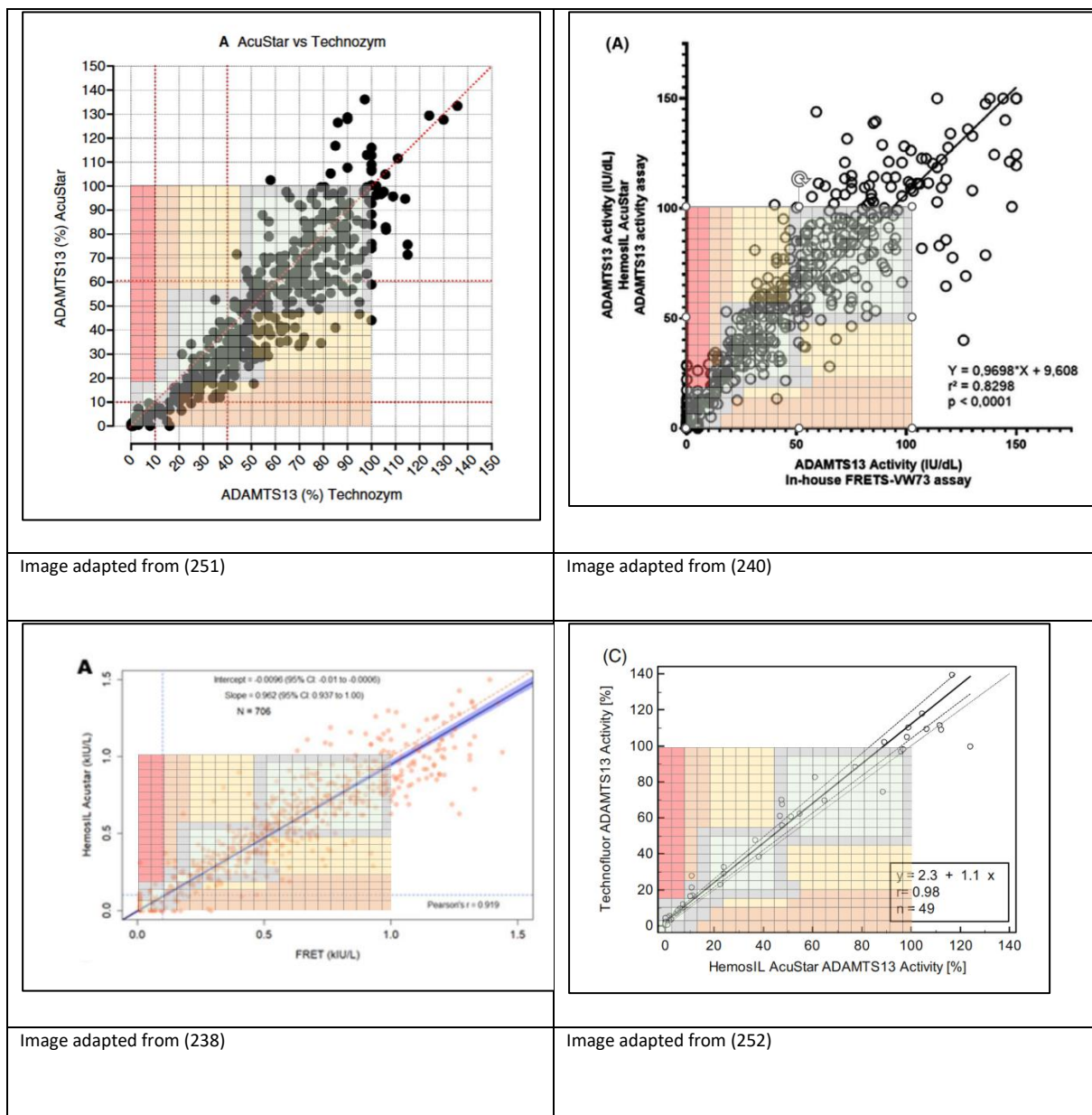


Figure 38: ADAMTS13 method comparison error grids applied to published studies.

Application of the ADAMTS13 error grid to four published studies where scatter graphs of comparison between ELISA and rapid methods were published

In contrast to the Favaloro findings (Figure 38, top left), the Beranger study (Figure 38, top right) identified several Zone D misclassifications—between four and six samples—where the AcuStar® reported falsely elevated ADAMTS13 activity, while the Technozym® ELISA showed results <10 IU/dL. These are false negative results with high clinical risk, as they could lead to the missed diagnosis of acute TTP and a failure to initiate timely treatment.

The dataset also included a number of Zone B and Zone C cases, with discrepancies generally within less urgent clinical ranges. However, the presence of Zone D errors underscores the potential unpredictability of certain assay platforms in critical diagnostic windows.

Dimopolous's study (Figure 38, bottom left), using >700 samples reinforced the broader findings of method comparison variability. A substantial number of samples fell into Zone B and Zone C, suggesting a moderate risk of clinically meaningful discrepancies.

Notably, several Zone D cases were identified, with the AcuStar® producing falsely high results relative to a FRET-based reference method. These errors raise similar concerns as seen in Beranger's work, suggesting the potential for systematic overestimation of ADAMTS13 activity in specific contexts or sample types.

An additional complication in this dataset was the presence of a large number of 0% activity results, which were visible in the error grid and hindered interpretability. These data points may reflect true severe deficiencies or analytical floor effects, and their inclusion further supports the use of structured frameworks like the error grid to contextualise outliers.

5.6 Chapter 5 Discussion

This chapter critically evaluated ADAMTS13 method comparison practices, introducing novel approaches that reframe assay verification around clinical interpretative risk and measurement uncertainty. For the first time, MU was explicitly integrated into method comparison analyses for ADAMTS13 activity assays, revealing its critical role in understanding inter-assay variability and its clinical consequences.

Traditional statistical methods—including correlation coefficients, regression analysis, and Bland-Altman plots—were shown to be inadequate for assessing clinical equivalence between assays. Correlation, although widely reported in published studies (P. J. Cornbleet & N. Gochman, 1979;

Linnet, 1998), was demonstrated here to mask clinically significant discrepancies. Stratified analysis revealed that correlation was strongest near the critical <10 IU/dL threshold, yet agreement deteriorated at higher levels where remission monitoring decisions are often made. Importantly, Monte Carlo simulations demonstrated that even modest MU significantly degraded correlation coefficients, particularly affecting Pearson's r due to its sensitivity to variance structure. These findings underscore the inadequacy of correlation coefficients as a primary verification metric, particularly when measurement uncertainty is not considered.

Regression analyses further demonstrated the limitations of traditional method comparison practices. While Passing-Bablok regression remains widely used, Deming regression provided more robust slope and intercept estimates in the presence of MU. The choice of regression method was shown to meaningfully influence parameter estimates, particularly when assessing proportional and constant bias. This analysis highlighted the need to incorporate MU modelling directly into regression frameworks and recommended the use of Deming regression with appropriately derived lambda values when MU differs between methods. Together, these findings emphasise that traditional analytical verification approaches, focused exclusively on analytical agreement, fail to capture clinically relevant assay variability, particularly near diagnostic thresholds.

A key methodological advance in this chapter was the introduction of likelihood ratio (LR) analysis as a tool for method verification. By evaluating how changing clinical decision limits affected diagnostic likelihood ratios (LR^+), this study demonstrated that adherence to historical cut-offs, established using legacy methods, may reduce diagnostic certainty in modern assays. Local verification data showed that by lowering the diagnostic threshold, LR^+ could be preserved above the clinically significant value of 10, maintaining consistent diagnostic performance despite inter-assay variability. This novel, proof-of-concept application of LR^+ as a criterion for cut-off optimisation offers a clinically meaningful alternative to fixed decision limits, enabling laboratories to harmonise result interpretation across different analytical platforms. This approach represents an important

conceptual shift in method verification: rather than adhering rigidly to historical thresholds, laboratories could instead seek to maintain diagnostic evidence strength—defined as preservation of LR^+ —as the target of assay verification. While further evaluation is required, this strategy offers a platform-agnostic method to reduce diagnostic uncertainty and supports the transition towards a Bayesian, probability-based interpretative framework discussed in subsequent chapters.

The development of an ADAMTS13-specific error grid incorporating MU guard bands represents the most practical and actionable outcome of this chapter. Adapting the CLSI EP27-A framework for glucose method comparisons, the error grid translated analytical discrepancies into clinical risk categories using established diagnostic thresholds (10, 20, and 50 IU/dL). The novel inclusion of guard bands, explicitly representing measurement uncertainty, provided a critical enhancement, delineating regions where uncertainty alone could result in patient misclassification. Local application of the grid revealed that discrepancies between the Technozym® ELISA and AcuStar® platforms were generally confined to low-risk zones, although systematic trends in bias were apparent. Retrospective application of the grid to published studies (238, 240, 251) demonstrated its ability to uncover clinically significant errors previously obscured by traditional analyses. Notably, discrepancies judged acceptable in published method comparison studies, based on statistical agreement alone, would have been flagged as clinically concerning using the error grid framework. This tool therefore offers laboratories a transparent and clinically meaningful method to assess comparability between ADAMTS13 assays, prioritising patient safety over purely analytical agreement.

Collectively, these findings highlight that measurement uncertainty is not merely a technical consideration but a critical determinant of clinical interpretability in ADAMTS13 testing. The failure to incorporate MU into existing verification frameworks represents a significant methodological limitation of current practice. This chapter demonstrates, for the first time, that MU must be considered at all stages of method comparison and verification to ensure clinically reliable reporting.

In conclusion, this chapter recommends the adoption of error grids incorporating measurement uncertainty as a standard component of ADAMTS13 method comparison and verification studies. This approach provides a clinically focused, risk-based framework for assessing assay comparability, moving beyond purely analytical agreement. While formal validation is required, the error grid represents the most practical tool for immediate adoption within laboratory quality assurance protocols. Likelihood ratio-based cut-off optimisation is presented as a novel, clinically aligned verification strategy, supporting harmonisation of result interpretation across platforms.

By integrating measurement uncertainty, clinical decision thresholds, and patient-centred risk assessment, this chapter advances method comparison methodology towards a more clinically meaningful and uncertainty-informed framework. These findings support a transition from analytical verification to interpretative verification, ensuring that ADAMTS13 assay results are evaluated not only for analytical comparability but for their potential impact on clinical decision-making and patient outcomes.

6 Chapter 6 – Measurement Uncertainty in ADAMTS13: A Practical Framework

6.1 Introduction

In haemostasis assays, particularly for ADAMTS13 activity, quantifying uncertainty presents unique challenges due to assay complexity, variable precision, and different methodologies. This chapter addresses these challenges by comparing two established frameworks for combined uncertainty estimation from the CLSI (160) and ISO (169). Both offer structured approaches to MU quantification but differ fundamentally in methodology, practical application, and treatment of variability sources.

CLSI EP29-A advocates detailed variance component analysis and ANOVA to dissect and quantify individual sources of imprecision (160). This method quantifies within-run and between-run, sources of uncertainty. ISO/TS 20914:2019 provides a pragmatic, top-down approach aligned with ISO 15189 accreditation requirements (ISO, 2022). It simplifies MU estimation by combining long-term IQC data and manufacturer-supplied calibrator uncertainty (u_{cal}) to generate a combined uncertainty estimate (u_c). This assumes that significant sources of variability are adequately captured through routine monitoring, without requiring decomposition of all contributing factors (103, 152, 154). Systematic bias is addressed theoretically within this framework: ISO/TS 20914 assumes bias has been minimised by manufacturers, and only mandates inclusion of bias correction uncertainty (u_{bias}) when an explicit correction is applied.

A critical distinction between these frameworks is in their data requirements. The top-down approach requires long-term IQC data, which is unavailable during early assay verification when implementing new methods. Also, rarely performed batched tests often do not have high volumes of IQC data, and in the case of ADAMTS13 activity by Technozym[®] ELISA measures QC performance by the mean result of replicates. To address this, this chapter presents variance component analysis

using ANOVA as a solution for both preliminary and established assays. Specifically, it demonstrates MU estimation using:

1. Controlled 5x5 replication studies for preliminary MU estimation (AcuStar® platform).
2. Routine duplicate IQC data for long-term MU assessment (Technozym® ELISA platform).

In both cases, the same statistical framework quantifies within-run and between-run variability to calculate imprecision contribution to uncertainty. This approach has not previously been framed in the context of preliminary MU estimation or applied systematically to duplicate IQC data in haemostasis assays.

A further focus of this chapter is the comparison of manual and automated ADAMTS13 assay platforms. The Technozym® ELISA and HemosIL AcuStar® methods differ fundamentally in assay design, calibration strategy, and operational stability, potentially affecting both imprecision and overall uncertainty profiles. Direct comparison of their respective MU profiles informs understanding of methodological limitations and supports best practice selection in clinical laboratories.

By applying both the detailed CLSI EP29-A variance component framework and the routine-focused ISO/TS 20914:2019 top-down approach, this chapter examines the applicability of each framework in for ADAMTS13 activity assays, considering both the requirements for routine implementation and method verification.

6.2 Aims

This chapter aims to establish a structured approach for estimating MU in ADAMTS13 activity assays, addressing both preliminary and routine testing contexts. It seeks to evaluate the suitability of

different uncertainty estimation frameworks, specifically comparing CLSI EP29-A and ISO/TS 20914:2019, for their applicability to haemostasis assays and their compliance with APS.

The chapter examines the impact of analytical platform differences, comparing manual ELISA and automated chemiluminescent methods, to determine their respective uncertainty profiles and practical limitations. Together, these analyses aim to inform best practice for MU estimation in haemostasis laboratories where different MU methods may be required depending on the stage of assay implementation.

6.3 ANOVA-Based Measurement Uncertainty Estimation for

ADAMTS13 Assays

6.3.1 Methods

The first methodological approach applies ANOVA-derived variance component analysis for preliminary and routine uncertainty estimation in ADAMTS13 activity assays. This section details the statistical framework used to quantify within-run and between-run variability.

6.3.1.1 Statistical methods used

Variance component analysis using one-way ANOVA partitions the total observed variance into within-run (repeatability) and between-run (intermediate precision) components. The key steps include calculation of within-run mean square (MS_{wth}) and between-run mean square (MS_{btw}) from ANOVA and derivation of within-run standard deviation (s_{wth}):

$$s_{wth} = \sqrt{MS_{wth}}$$

Equation 18: Within run standard deviation.

According to CLSI EP29-A. s_{wth} – standard deviation within run. MS_{wth} – Mean squared error within

The between run SD is somewhat more complex and is influenced by the number of data points used. The s_{btw} is chosen as the maximum value of either 0, or a calculated function according to the equation below:

$$s_{btw} = \sqrt{\text{MAX}\left(0, \frac{MS_{btw} - MS_{wth}}{n_0}\right)}$$

Equation 19: The between run standard deviation.

According to CLSI EP29-A. s_{btw} – between run standard deviation, MS_{btw} – mean squared error between run, MS_{wth} – mean squared error within run, n_0 – number of runs

Using the difference between the between and within Mean Squares ($MS_{btw} - MS_{wth}$) and dividing that by the number of samples (n_o), this is compared to a value of 0. Whichever is largest is passed to the square root function and determines the SD of the between run. Accordingly, if $MS_{btw} > MS_{wth}$, the value of s_{btw} must be 0. The use of $\text{MAX}(0, (MS_{btw} - MS_{wth})/n_o)$ ensures the between-group standard deviation (s_{btw}) is never negative, as variance cannot logically be less than zero. Negative estimates can occur due to sampling variability, especially when there is little or no true between-laboratory variation. Rather than interpret a negative value, the method conservatively assigns zero, implying no detectable between-lab variability. This truncation avoids spurious contributions to uncertainty and ensures estimates remain valid.

From the two equations above, the familiar root sum of squares methodology is used to determine the combined uncertainty (u_c), according to the below equation:

$$u_c = \sqrt{s_{wth}^2 + s_{btw}^2}$$

Equation 20: The combined uncertainty.

Calculated by the root sum of squares method, combining within and between run standard deviations determined above.

6.3.1.2 Calculations

All calculations were performed in Microsoft Excel[®] (Microsoft, USA), using the analysis tool pack and the single factor ANOVA analysis tool pack function. Output from ANOVA was checked for data integrity and resultant metrics including MS_{btw} and MS_{wth} were manually extracted and used in a bespoke spreadsheet workbook accompanying the analysis where the uncertainty was able to be calculated according to the statistical approaches above (An example is shown in Appendices B and C).

6.3.1.3 Application 1: Preliminary Uncertainty Estimation Using 5x5 Replication Study (AcuStar®)

In the absence of routine IQC data, a 5x5 reproducibility experiment was conducted using the AcuStar® ADAMTS13 activity assay. Over five consecutive days, five independent runs were performed daily, each including five replicates of VWF low and VWF Normal QC materials. The resulting 25 measurements per QC level were analysed using ANOVA.

Within-run and between-run components were calculated as described in Appendix C. Combined uncertainty estimates provided a preliminary estimate of uRW. This enabled assessment of whether the method fulfilled APS at verification, providing an interim uncertainty estimate until sufficient routine QC data became available.

6.3.1.4 Application 2: Long-Term Variance Component Analysis from Duplicate IQC Data

(Technozym® ELISA)

For the Technozym® ELISA method, routine IQC involves duplicate measurements per QC level in each analytical run. Traditional top-down MU estimation using the daily mean alone would disregard replicate variability, underestimating uncertainty.

To address this, ANOVA-derived variance component analysis was applied to the full dataset of duplicate QC measurements, using paired replicate data from 88 QC runs across eight months.

Within-run variance (reflecting replicate differences) and between-run variance (reflecting day-to-day variability) were estimated separately, then combined to calculate standard uncertainty.

6.3.2 Results

6.3.2.1 Preliminary MU Estimation: AcuStar® 5x5 Study

Analysis of VWF Low (mean 32.8%) and Normal (mean 74.8%) QC levels using the 5x5 replication study produced combined uncertainties of 5.1% and 5.8% respectively (Full calculations are available in Appendix C). Expanded uncertainties ($k=2$) approximated 10%, occupying approximately 50% of the maximum allowable uncertainty (20%, from Chapter 4) for ADAMTS13 activity reporting. This is expected as this MU method does not account for uncertainty of the calibrator and also doesn't include all variability that would be expected over long term IQC performance. When including the 10% value (for u_{cal}) from the manufacturer (Werfen) a combined uncertainty of 5.6% and 6.4 % at low and high QC levels was realised. This equates to expanded relative uncertainties of 11.2% and 12.8% respectively again within performance specification limits.

Interestingly, relative within-run imprecision remained stable across QC levels (Low – 2.86%, High – 2.84%), while between-run variability increased at higher concentrations (Low – 4.26%, High – 5.10%).

Overall, using this method the 5x5 method, using variance component analysis provides a convenient method to quantify MU. It also provides a valid albeit conservative interim estimate of MU, enhanced by supplementing with calibrator uncertainty, not used in the original CLSI method.

6.3.2.2 Long-Term MU Estimation: Technozym® Duplicate IQC Analysis

In the ELISA dataset, replicate variability contributed significantly to overall uncertainty. For high QC, within-run and between-run components were approximately equal (5.79 v 5.83 IU/dL, Appendix B) each contributing ~50% to total MU. For low QC, between-run variation dominated, contributing ~79% of total variance.

Standard uncertainties ranged from 10.6% to 13.4% CV for high QC across different reagent lots, and from 13.8% to 17.9% CV for low QC (Table 34, See Appendix B for full calculations). Ignoring within-

run variation (by analysing only daily means) systematically underestimated uncertainty, particularly at lower concentrations where replicate variability was proportionally greater. Therefore, accounting for replicate variability provides a more complete and accurate assessment of MU, critical for assays like ADAMTS13 where manual steps introduce pre-analytical variability.

Table 34: Standard uncertainty for three lot numbers from variance components analysis.

The table presents the standard uncertainty estimates derived from variance components analysis for three separate reagent lots (RU85B00, RU83B00, and RU88B00) of the Technozym® ADAMTS13 ELISA method. For both high- and low-level internal quality controls (IQC), uncertainty is presented in absolute terms (u_c), as a percentage relative to the mean ($u_c \%$), and as the standard uncertainty of the run means ($u_c(\bar{x})$), which reflects the uncertainty of the average value when replicates are available.

	Normal QC			Groups (n)	Low QC			Groups (n)
	u_c	$u_c \%$	$u_c(\bar{x})$		u_c	$u_c \%$	$u_c(\bar{x})$	
RU85B00	7.3	10.6	2.2	5	2.90	14.1	0.9	5
RU83B00	8.2	11.3	1.7	12	3.03	13.8	0.6	12
RU88B00	9.7	13.4	1.3	24	2.48	17.9	0.4	24

6.3.2.3 Comparative Interpretation

Both approaches demonstrated that ANOVA-derived variance component analysis can generate clinically meaningful uncertainty estimates for ADAMTS13 methods. The preliminary 5x5 method offers a practical interim solution, enabling risk-based decisions during method verification. The routine QC-based analysis using duplicate data represents the preferred long-term solution, fully compliant with ISO/TS 20914 and CLSI EP29-A.

Failure to account for within-run variability, particularly in manual methods, risks underestimating MU and misrepresenting analytical performance. This is particularly significant in ADAMTS13 testing, where accurate uncertainty estimation supports clinical decision-making at critical thresholds.

6.4 Benchmarking Combined Uncertainty Using Standard Top-Down

Method

The second approach follows the ISO/TS 20914:2019 top-down framework, combining routine internal quality control data with manufacturer-supplied calibrator uncertainties to calculate combined and expanded measurement uncertainty across both analytical platforms.

6.4.1 Methods

6.4.1.1 Internal Quality Control Data — Technozym® ELISA

Two levels of IQC material (Technozym® ADAMTS13 High and Low QC) were monitored. The High QC material targeted the normal range (80–100 IU/dL), and the Low QC material targeted the abnormal range (20–40 IU/dL). Data were collected over a seven-month period (1 January 2022 to 31 July 2022), comprising 44 QC measurements for each level.

External quality assessment (EQA) participation during the preceding 12 months (UKNEQAS for Blood Coagulation) included three surveys using commutable materials with consensus-assigned target values themselves with an associated uncertainty, assigned according to the derivation of the consensus value, individually for each survey. No medically significant bias was detected either across individual surveys or in the overall mean, and therefore no bias correction (*ubias*) was applied.

6.4.1.2 Internal Quality Control Data — HemosIL® AcuStar

IQC data for ADAMTS13 activity were collected across five reagent lots for both QC levels: Q-VWFL (Low QC) and Q-VWFN (Normal QC). Sample sizes varied across lots, for Q-VWFL, data points ranged from 2 (lot 3738) to 56 (lot 4879); for Q-VWFN, from 2 (lot 3738) to 57 (lot 4879), with intermediate sample sizes for lots 4189, 4284, and 4538. A minimum sample number was not pre-specified for this study due to the retrospective nature of the data collection and the primary aim of evaluating real-world analytical performance across multiple reagent lots. The sample sizes available for each lot

reflected routine laboratory testing volumes rather than controlled experimental design. Imposing a minimum threshold would have excluded available data that, while limited in size, still contributed valuable information regarding between-lot variability. This is particularly important in specialist haemostasis testing, where large datasets may not be routinely achievable due to low test frequency and the clinical rarity of some indications.

The variation in sample numbers across lots was addressed by calculating pooled standard deviations, ensuring that lots with limited data did not disproportionately influence the overall estimate of measurement variability (See Appendix E). Bias evaluation using EQA material indicated no medically significant bias, and therefore no bias correction was applied.

6.4.1.3 MU Calculation

MU was assessed using the ISO/TS 20914:2019 top-down approach, combining IQC data with calibrator uncertainty (u_{cal}) estimates provided by manufacturers. Imprecision was evaluated across multiple IQC levels and reagent lot numbers for both the Technozym® ELISA and HemosIL® AcuStar platforms.

Within-laboratory imprecision (u_{Rw}) was estimated by calculating the standard deviation of IQC measurements over the study period. Imprecision estimates were stratified by QC level to reflect performance variability across the analytical range. Manufacturer-reported calibrator uncertainties (u_{cal}) were then incorporated into the combined uncertainty (u_c) calculations. An attempt to incorporate uncertainty from higher references (u_{ref}) was limited to the uncertainty of the fill volume of the WHO 1st international standard (12/252) and was universally applied as 0.4%.

For the Technozym® ELISA, individual calibrator uncertainties were provided for each calibrator concentration. The calibrator value nearest to the IQC level was selected, and its uncertainty combined with u_{Rw} and the uncertainty of the reference material to calculate combined uncertainty.

For the HemosIL® AcuStar, a single relative calibrator uncertainty of 10%, as advised by the manufacturer, was applied uniformly across all QC levels. While necessary given available information, this approach may not fully capture potential variations in calibration performance across the assay's calibration hierarchy.

Uncertainty calculations followed the statistical models outlined in ISO/TS 20914:2019. Final combined and expanded uncertainties were summarised in tabular form according to best practice guidance, consistent with the International Council for Standardization in Haematology (ICSH) MU Working Group recommendations for haemostasis assays (under development, examples are shown in Appendices D and F).

6.4.2 Results

6.4.2.1 Combined uncertainty of Technozym® ELISA

Across the seven-month data collection period, the Technozym® ADAMTS13 High QC material had a local mean ADAMTS13 activity of 72.6 IU/dL, closely aligned with the locally assigned target value of 72.0 IU/dL, based on previous performance and manufacturer's instructions for use. The observed local standard deviation was 4.64 IU/dL, resulting in a standard uncertainty under routine conditions (u_{RW}) of 4.64 IU/dL, at the higher level. u_{RW} was 1.71 at the lower QC level. After combining the contributions from calibrator uncertainty (u_{cal}) and reference material uncertainty (u_{ref}), the total combined standard uncertainty was calculated as 4.82 IU/dL. Applying a coverage factor of 2 ($k=2$) to account for 95% confidence, the expanded uncertainty (U) for the High QC material was a relative expanded uncertainty ($\%U_{rel}$) of 13.2% and 13.4% at the low level.

The maximum allowable MU was predefined at 20% for both the normal and low ranges following local agreement with the clinical lead, aligning with clinical outcome based APS previously presented in this thesis (Chapter 4). In both cases, the observed relative expanded uncertainties were below this threshold, confirming that the MU remained within clinically acceptable limits throughout the

study period. No bias correction was applied, as external quality assessment participation across three surveys during the preceding year demonstrated no medically significant bias at either level.

6.4.2.2 Combined uncertainty of HemosIL AcuStar®

IQC performance for ADAMTS13 activity was assessed across five consecutive reagent lots for two QC levels: Q-VWFL (Low) and Q-VWFN (High). For Q-VWFL, the CV ranged from 4.75% to 17.82%, with mean activity values between 0.28 and 0.38. For Q-VWFN, mean values ranged from 0.77 to 0.90, with CVs between 7.58% and 14.10%. Overall, a CV of approximately 10% was observed for the QC Low material, and around 9% for the QC High material.

Variability in SD was observed across lots, ranging from 0.02 for lot 3738 to 0.06 for lot 4284. The very low SD seen in lot 3738 was based on only two data points, and therefore had limited reliability. Without appropriately weighting by sample number, such small datasets could exert disproportionate influence on overall estimates of variability. In contrast, lot 4284 showed visibly increased variability at both QC levels, reflected in the highest CVs recorded (17.82% for Q-VWFL and 14.10% for Q-VWFN), but these findings were based on sufficiently large sample numbers to be considered robust. Despite the increased variability associated with lot 4284, all results remained within both local and manufacturer-recommended performance specifications. A trend toward improved precision was observed with later lot numbers, particularly lot 4879, which demonstrated the lowest CV for Q-VWFN (7.58%).

To account for the variability between lots and avoid over- or under-representation of individual lot contributions, the pooled SD was calculated in accordance with ISO/TS 20914:2019 guidance. This method appropriately weighted each lot's SD by its sample size, ensuring that lots with low numbers of observations did not exert disproportionate influence on the overall variability estimate.

For Q-VWFL, individual lot SDs ranged from 0.02 to 0.06. Lot 4284 exhibited the highest individual SD (0.06), while lots 3738 and 4879 both showed lower SDs (0.02). After pooling across all lots, the final pooled SD for the Low QC level was 0.03.

For Q-VWFN, SDs across lots ranged from 0.04 to 0.12, with lot 4284 again demonstrating the highest variability (SD 0.12). The pooled SD across all lots for the Normal QC level was calculated as 0.08.

The use of the pooled SD approach, rather than treating each lot independently, provided a more robust and representative estimate of long-term imprecision. This was particularly important given the variation in sample sizes across lots, with some lots (e.g., 3738) having very limited data ($n=2$), which could otherwise distort the overall estimate. By applying this method, a stable and clinically appropriate estimate of u_{RW} was achieved for both QC levels.

After combining the uncertainties associated with imprecision (u_{RW}), the calibrator (u_{cal}), and the reference standard (u_{ref}), the total combined standard uncertainty was calculated as 12.8 % for the High QC and 10.4 % for the Low QC, here representing an absolute rather than relative uncertainty. Applying a coverage factor of 2 ($k=2$) resulted in expanded uncertainties (U) of 25.6 % and 20.8 %, respectively.

When expressed relative to the mean values, this corresponded to an uncertainty of $\pm 16.0\%$ for the VWF High QC and $\pm 29.8\%$ for the VWF Low QC, at approximately 95% confidence. The maximum allowable MU was predefined locally at 20% for both levels, as previously. The uncertainty for the High QC was within the allowable limit. Although the Low QC initially exceeded the 20% threshold, it was accepted after discussion, recognising that the information provided with the calibrator uncertainty (u_{cal}) was not in sufficient detail and likely overestimated the u_{cal} contribution, and that the u_{RW} component alone accounted for a U_c of 10%, an acceptable u_{RW} for an APS of 20%. Clearly, review of the u_{cal} estimate from the manufacturer is required to improve its inclusion in the top down method.

6.5 Chapter 6 Discussion

ANOVA-derived variance component analysis proved an effective and adaptable method for estimating MU in ADAMTS13 activity assays, applicable to both preliminary and routine data. The controlled 5x5 replication study using the HemosIL AcuStar® platform demonstrated that preliminary MU estimates could be generated rapidly during assay implementation, providing combined uncertainties of approximately 5–6% CV at both low and normal control levels. These preliminary estimates yielded expanded uncertainties near 10%, occupying approximately 50% of the predefined maximum allowable uncertainty (20%) for ADAMTS13 activity reporting. This confirms the suitability of the 5x5 method as a conservative, risk-based approach for verifying compliance with APS during early method adoption.

For the manual Technozym® ELISA, long-term routine IQC data subjected to ANOVA revealed significant contributions to uncertainty from both within-run and between-run variability, with between-run variation dominating at lower QC levels. Combined uncertainty estimates ranged from 10.6% to 17.9% CV across different control levels and reagent lots, approaching the maximum allowable limits defined locally. Crucially, ignoring within-run variability—by treating daily means as single observations—consistently underestimated MU, particularly at lower concentrations where replicate variability was proportionally greater. This finding underscores the need to include replicate-level variability in uncertainty calculations for one-at-a-time assays like the Technozym® ELISA, where manual steps increase susceptibility to pre-analytical variation.

The application of ISO/TS 20914:2019 further reinforced these findings, demonstrating that the top-down approach using pooled routine IQC data, combined with manufacturer-provided calibrator uncertainties, enabled consistent estimation of combined and expanded MU for both platforms. However, neither the Werfen calibrator uncertainty nor the Technozym® ELISA calibrator uncertainty explicitly incorporated uncertainty propagated through the full metrological traceability chain. Both calibrators, although traceable to the WHO 1st International Standard for ADAMTS13 (12/252), relied

on commercial manufacturing assignments of calibrator values. Communication with Werfen to confirm the 10% calibration uncertainty stated lacked clarification as to whether it was either absolute or relative, and no breakdown of components was available. Similarly, Technozym® reported calibrator uncertainties ranging from 1.28% at high concentrations to approximately 3% near clinical decision limits, without accounting for upstream uncertainty sources.

The uncertainty associated with the assigned value of the WHO International Standard 12/252 itself was not integrated into either manufacturer's calibrator uncertainty statements. Although Hubbard *et. al.* reported a group geometric CV of 12.4% for between-laboratory variability during standard assignment (65), this reflects inter-laboratory variability, not uncertainty of the certified value. The only quantitative uncertainty associated directly with the WHO standard was a 0.4% CV based on repeatability of vial mass measurements, representing physical consistency rather than functional concentration uncertainty.

This incomplete propagation of uncertainty from the primary reference material through to end-user calibrators represents a significant limitation to strict top-down MU estimation and full metrological traceability application (146, 149, 151). Consequently, while MU estimates remained within clinically acceptable limits for both platforms, the laboratory formally acknowledged this traceability gap in its quality management system and initiated manufacturer engagement to request greater transparency in calibrator uncertainty reporting.

In summary, this chapter demonstrates that ANOVA-derived variance component analysis provides a robust methodological solution for MU estimation across both preliminary and routine laboratory contexts. The ISO/TS 20914:2019 framework supports routine application of top-down MU estimation, though its effectiveness is constrained by incomplete calibrator uncertainty transparency. Comparative analysis confirms the advantage of automated platforms in achieving lower imprecision estimates, reinforcing the importance of platform selection and comprehensive uncertainty accounting to support reliable ADAMTS13 activity reporting in clinical practice.

7 Chapter 7 - Decomposing Internal Quality Control

Data for MU

7.1 Introduction

The estimation of within-laboratory measurement uncertainty, particularly the standard uncertainty component u_{RW} , typically relies on the analysis of routine IQC data. The standard approach involves calculating the standard deviation SD from IQC results as a proxy for random variation within the measurement process (169). This top-down methodology, recommended for its practicality over more complex bottom-up calculations defined in the GUM (JCGM:100), is widely adopted in medical laboratories (149, 152, 253). However, a fundamental statistical assumption underpins this process: that IQC data are independent and identically distributed (i.i.d.), meaning each result reflects a random outcome, unaffected by preceding or subsequent measurements (161).

In clinical laboratory practice, this assumption is inherently problematic. IQC procedures are designed to detect non-random variation—such as drift, calibration shifts, or systematic measurement bias. Consequently, routine IQC datasets inherently include both random variability and systematic influences. ISO/TS 20914:2019 acknowledges that long-term IQC data may absorb minor systematic variations, such as lot changes, into the apparent random variability (Clause 5.3, (169)). This practical interpretation assumes that such influences become indistinguishable from random variation over time (154, 188). However, it leaves unexamined the extent to which systematic components distort uncertainty estimates.

The reliance on the SD of IQC results to estimate u_{RW} , without validating the underlying data structure, risks incorporating short-term systematic bias into the random uncertainty estimate. This is particularly concerning when such bias manifests as subtle drift or periodic variation, artificially inflating imprecision estimates and potentially misrepresenting analytical performance.

To address this methodological concern, this chapter explores the internal structure of IQC datasets using a combined analytical framework incorporating autocorrelation detection and time series decomposition. Together, these techniques interrogate whether IQC data violate the i.i.d. assumption and quantify the extent of systematic influence within routine QC monitoring.

Autocorrelation analysis examines whether sequential IQC measurements are statistically dependent, identifying whether earlier results partially predict later values (254). The detection of autocorrelation suggests that systematic factors—such as reagent degradation, calibration drift, or environmental effects—contribute to apparent variability in the control process. Recognising such dependencies is essential, as autocorrelation can mask performance shifts, undermine Westgard rule sensitivity, and reduce method reliability.

Complementing this, time series decomposition separates IQC datasets into trend, seasonal, and residual (random) components (254, 255). By isolating the truly random component, decomposition enables direct evaluation of the variability that should define u_{RW} . This approach provides a practical means to investigate whether conventional SD-based methods inadvertently misclassify systematic variation as random error, potentially leading to inaccurate MU estimation. Both additive and multiplicative decomposition models are considered to assess whether fluctuations scale with mean activity levels, with model choice reflecting assay-specific behaviour (256) Shumway (254) (257).

APTT and ADAMTS13 activity assays serve as exemplar workflows within this analysis. APTT represents a continuous, high-volume, automated system with prolonged reagent and control use, where autocorrelation and systematic trends might be expected. In contrast, ADAMTS13 assays, conducted using manual or “one-at-a-time” methods such as ELISA, typically involve fresh reagents and independently processed controls, theoretically reducing systematic dependencies. These contrasting workflows allow exploration of how analytical context influences IQC data structure.

This chapter evaluates the assumption that IQC data represent independent, identically distributed measurements, and assesses whether standard MU estimation practices based on SD calculation

remain valid when systematic structure is present. By highlighting these limitations, the chapter provides a methodological foundation for future refinement of uncertainty estimation processes in medical laboratories.

7.2 Aims

This chapter aims to determine whether IQC data meet the statistical assumptions required for valid MU estimation using standard deviation as a surrogate for within-laboratory random uncertainty (u_{RW}). The findings aim to inform the suitability of current top-down MU estimation practices and assess whether systematic variability is being misclassified as random error in uncertainty calculations.

7.3 Randomness of IQC data used for uRW determination

7.3.1 Methods

This analysis focused on evaluating whether IQC data, when viewed sequentially, exhibit patterns that suggest autocorrelation, drift, or other non-random behaviours. Three assays were selected to represent different analytical platforms, control materials, and time intervals – the APTT (ACLTOP 750CTS), ADAMTS13 by Technozym[®] ELISA and HemosIL AcuStar[®].

For the APTT assay, IQC data were extracted over a one-month period (July 2021) from the ACLTOP 750 CTS analyser. This included High Abnormal QC levels as an exemplar. 191 IQC results were analysed using the runs test to statistically assess the presence of non-random sequences, and histograms were generated to visually assess the normality of the data distributions. Run plots were produced for each control level to examine the order and progression of results over time.

Similarly, IQC data from the Technozym[®] ELISA method for ADAMTS13 activity were analysed. These data spanned a longer timeframe, from January 16th, 2018 to November 20th, 2018 to ensure sufficient data – 30 samples was considered appropriate for the study being undertaken. A single high and low quality control lot (RUY71BO, n= 47 and RUZ71BO, n=47, respectively, Table 35) was used during this period. The results were arranged in run order rather than by date and subjected to the same runs test and visual histogram analysis to assess randomness and distribution shape.

Finally, the same approach was applied to IQC data from the AcuStar[®] platform, which performs rapid ADAMTS13 activity testing. Here, IQC data were collected between October 17th, 2022 and December 28th, 2022 using normal and low vWF-based control material (Lot Number 4783, n=36, Table 35). Again, the data were ordered by sequence of acquisition and analysed using the runs test alongside histogram review and run plot visualisation.

Table 35: Assays and control material used as the focus of this study.

APTT – activated Partial thromboplastin Time, ELISA – Enzyme Linked Immunosorbent Assay, CTS – Closed Tube Sampling

Assay	Platform	Control Materials	Time Interval
APTT	ACLTOP 750 CTS	High Abnormal QC	One-month period (July 2021) n=191
ADAMTS13 activity (ELISA)	In-house ELISA method	RUY71BO (high), RUZ71BO (low)	Jan 16 th , 2018 - Nov 20 th 2018, n=47
ADAMTS13 activity (AcuStar)	AcuStar platform	vWF-based control material (Lot Number 4783)	Oct 17 th , 2022 - Dec 28 th 2022, n=36

This multi-assay comparison provides insight into the appropriateness of applying standard deviation to IQC data for estimating random uncertainty and evaluates whether the assumption of independence holds across different platforms, methodologies, and QC timeframes.

7.3.1.1 The runs test

The runs test is a non-parametric statistical method used to assess the randomness of a sequence of observations, typically binary or sign-based (+/-). It evaluates whether the order of elements in a data sequence is too predictable to be by chance, which is crucial when testing for independence or randomness in time series, residuals, or binary outcomes. A “run” is defined as a consecutive sequence of the same category (e.g. all + or all -), and the test compares the observed number of runs to the expected number under randomness. Too few or too many runs indicate a departure from randomness.

The runs test was implemented using the *randtests* package in R, and plots were generated using the *ggplot2* package, also in R.

7.3.2 Results

7.3.2.1 APTT (High volume automated routine assay)

Plotting the individual IQC data points in the order they were generated (Figure 39) reveals considerable variability between measurements, with no obvious trend or correlation discernible by visual inspection. To formally assess the presence of correlation across successive measurements, the runs test for randomness was applied.

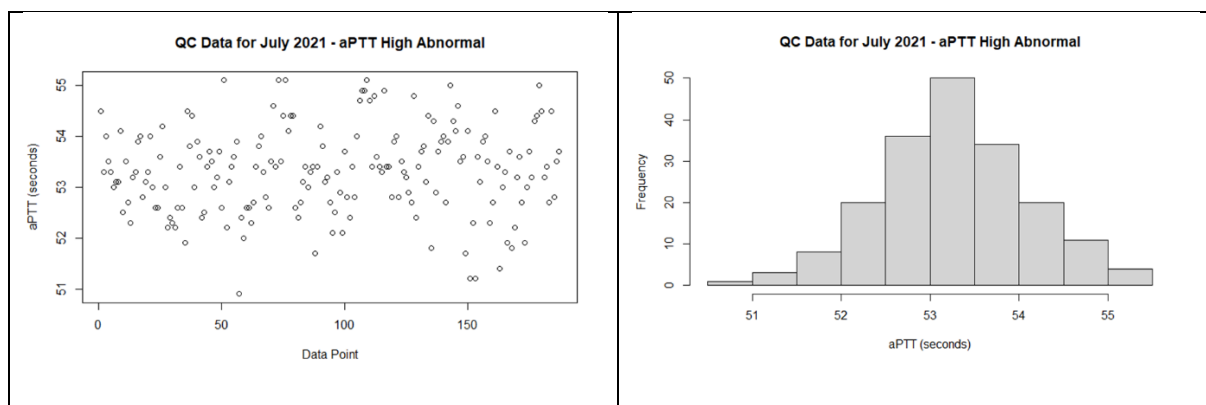


Figure 39: APTT IQC performance.

(Left) - activated Partial Thromboplastin Time (aPTT) Internal Quality Control (IQC) data for July 2021 ($n = 191$). The IQC level investigated was the HemosIL High Abnormal Quality Control, Lot number . (Right) - Distribution of results from the IQC dataset used for this analysis

The test yielded a runs statistic of 77, with a two-sided p-value of 0.0243. Using a significance level of 0.05 (critical value = ± 1.96), this result provides evidence of non-randomness in the data. Specifically, it suggests that consecutive data points are statistically correlated, rather than independent.

This finding has important implications and merited further investigation in subsequent sections. The presence of correlation between successive IQC results indicates the potential influence of an underlying systematic trend—possibly a developing systematic error. Although current guidelines allow such systematic effects to be absorbed within the estimate of long-term imprecision (154),

interpreting uRW as purely random variation becomes questionable. Potentially, caution should be exercised when using standard deviation of IQC data as a direct surrogate for random uncertainty without further evaluation of temporal dependencies.

7.3.2.2 One at a time ADAMTS13 activity assay method (Technozym® ELISA)

The runs test for randomness was applied to quality control (QC) data from both the AcuStar® rapid immunoassay and the manual ELISA for ADAMTS13 activity to determine whether the assumption of independent and identically distributed (i.i.d.) IQC data holds.

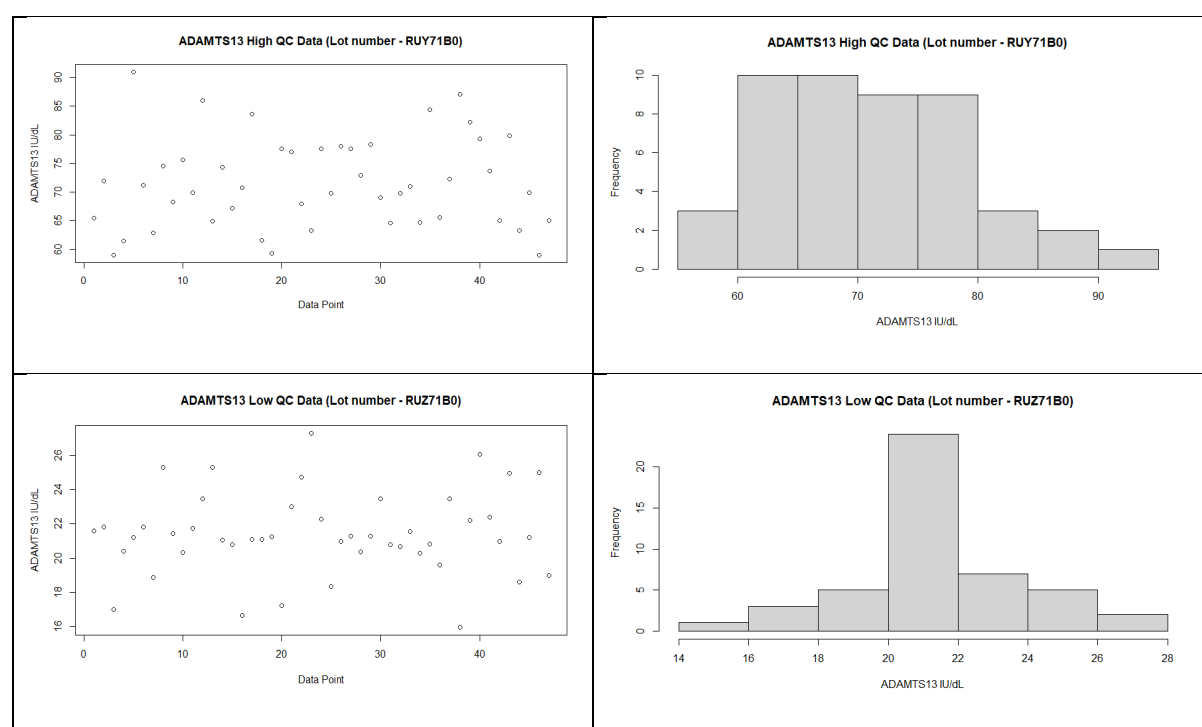


Figure 40: ADAMTS13 IQC run plots (Technozym®).

(A and C) – Run plots of a single ADAMTS13 High Quality Control (QC), Lot Number RUY71B0 (Top-left, n=47) or Low Quality Control (QC), Lot Number RUZ71B0 (Bottom-left, n=47) between January 16th 2018 and November 20th 2018. Points are ordered according to the order they were run (X-axis not scaled for date/time). Histograms of ADAMTS13 High Quality Control (QC), Lot Number RUY71B0 (Top right) or Low Quality Control (QC), Lot Number RUZ71B0 (Bottom right) between January 16th 2018 and November 20th 2018.

The ELISA-based ADAMTS13 assay did not show statistically significant departures from randomness (Figure 40). The low QC returned $z = 1.33$, $p = 0.183$, and the high QC yielded $z = 0.44$, $p = 0.656$, indicating that the sequence of IQC values is consistent with random variation and independent sampling.

For the AcuStar® platform, both the low and high QC data sets demonstrated statistically significant evidence of non-random behaviour (Figure 41). The low QC yielded a runs statistic of $z = 4.45$, $p < 0.001$, while the high QC returned $z = 5.12$, $p < 0.001$. These results indicate strong evidence of correlation between successive measurements, suggesting that consecutive QC results are not independent and may be influenced by shared factors in the measurement process, such as analyser performance.

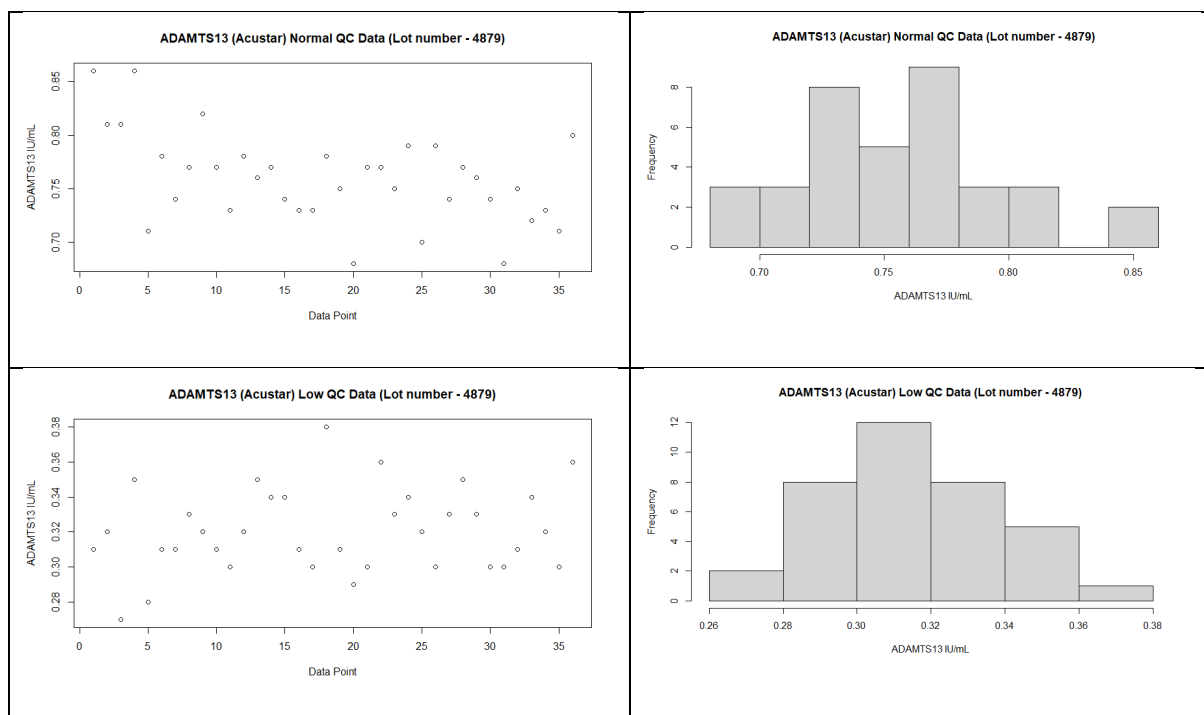


Figure 41: ADAMTS13 IQC run plots (AcuStar®).

Run plots of a single ADAMTS13 (Acustar) Normal vWF Quality Control (QC), Lot Number 4783 (Top-left, $n=36$) or Low vWF Quality Control (QC), Lot Number 4783 (Bottom-left, $n=36$) between October 17th 2022 and December 28th 2022. Points are ordered according to the order they were run (X-axis not scaled for date/time). Histograms of ADAMTS13 (Acustar) Normal vWF Quality Control (QC), Lot Number 4783 (Top right) or Low vWF Quality Control (QC), Lot Number 4783 (Bottom-right) between October 17th 2022 and December 28th 2022.

These contrasting results align with the expected behaviour of the two platforms. The ELISA is a manual, single-well assay where each measurement is performed independently, making it less susceptible to autocorrelation introduced by the analytical system. Conversely, the AcuStar® is a fully automated platform, where repeated measurements are subject to shared system characteristics such as reagent stability, calibration drift, or mechanical factors. As a result, temporal correlation between consecutive IQC measurements is more likely in the AcuStar® system, contributing to the observed non-random behaviour.

7.4 Autocorrelation of IQC data used for u_{RW} determination

7.4.1 Methods

A proof-of-concept study was conducted to evaluate the presence and structure of autocorrelation in IQC data across three analytical platforms: a high-volume coagulation assay (APTT), a manual ELISA (Technozym[®]) for ADAMTS13, and a rapid random-access chemiluminescent immunoassay (AcuStar[®]) for ADAMTS13.

7.4.1.1 Sample data

7.4.1.1.1 Full dataset

For each assay, retrospective IQC data were extracted from the laboratory information management system (LIMS) or from laboratory data collection spreadsheets.

The full dataset used to investigate performance across multiple lot numbers is described in Table 36.

The Technozym ELISA data represented the full dataset available in the laboratory, the AcuStar data represented a full 12 months data collection on the automated platform.

Table 36: IQC data characteristics used for the study of autocorrelation and partial correlation

Assay	QC	Start Date	End Date	Number of QC points used for time series algorithms
APTT	High Abnormal	01/02/2021	31/08/2021	213
Technozym [®] ELISA	High QC	01/06/2012	01/01/2021	285
	Low QC	01/06/2012	01/01/2021	286
HemosIL AcuStar [®]	VWF High	04/01/2022	31/12/2022	116
	VWF Low	04/01/2022	31/12/2022	106

7.4.1.1.2 Single lot dataset

The dataset used in section 7.3 was used for this analysis, consisting of 47 ELISA datapoints (both high and low QC) and 36 AcuStar QC datapoints (Again for both high and low QC). These QC represent routine QC used to control assays, and calculate u_{RW} , and also are at measurand concentrations intended to challenge performance of the assay at clinically meaningful levels.

7.4.1.2 Time series analysis

Time series analysis was performed using R (version 4.1.0), with the *acf()* and *pacf()* functions used to assess the autocorrelation and partial autocorrelation structures within each IQC dataset. Data were visually inspected for normality and stationarity, and where necessary, first-order differencing was applied to stabilise the series. Autocorrelation at each lag was assessed against 95% confidence intervals to determine statistical significance. ACF and PACF plots were generated for each assay to characterise the temporal structure of variation and compare across the three platforms.

7.4.1.3 Autocorrelation and partial correlation in R

The autocorrelation function (ACF) in R measures the linear relationship between observations separated by k time units (lags). It provides a visual and statistical method for detecting temporal dependencies within a time series. On the ACF plot, blue lines typically represent 95% confidence intervals, and autocorrelations that exceed these bounds suggest statistically significant relationships at those lags. In contrast, the PACF quantifies the correlation between a time series observation and a specific lag, while controlling for the correlations of all shorter lags. This makes PACF particularly useful for identifying the appropriate order of autoregressive models, even though no formal modelling is undertaken in the present study.

In R, both *acf()* and *pacf()* are part of the base stats package, which is loaded by default and requires no additional installation. The *acf()* function computes and plots autocorrelations for univariate or multivariate time series, helping to reveal repeating patterns or potential seasonality. The *pacf()*

function complements this by isolating the direct influence of individual lags. Together, these tools are commonly used in model selection, such as determining lag terms in ARIMA models. Example usage in R includes *acf(ts_data)* and *pacf(ts_data)*, where *ts_data* indicates the tibble data format derived from IQC data.

When applied to IQC data, ACF and PACF plots can help identify non-random behaviour over time. This insight can support laboratories in refining quality control protocols and in understanding underlying sources of variation.

7.4.1.4 *Augmented Dickey fuller test for stationarity*

To assess the stationarity of IQC time series data, the ADF test was applied. The test evaluates the null hypothesis that a unit root is present in the data, indicating non-stationarity, against the alternative hypothesis of stationarity. The ADF test was performed using the *adf.test()* function from the *tseries* package in R (version 4.1.0). A significance level of 0.05 was used to determine whether the null hypothesis could be rejected, with p-values less than 0.05 indicating evidence of stationarity in the series.

7.4.2 Results

All datasets were determined to be stationary using the ADF test. In total 9467 HemosIL High Abnormal APTT (Werfen, Bedford, USA) QC results, representing 6 months collection, were available. For the data to be implemented in the time series algorithms equally spaced data was required, the mean result for each days data was calculated and used to detect daily autocorrelation resulting in 213 individual data points. for autocorrelation and partial correlation, results of analysis are shown in Figure 42, below.

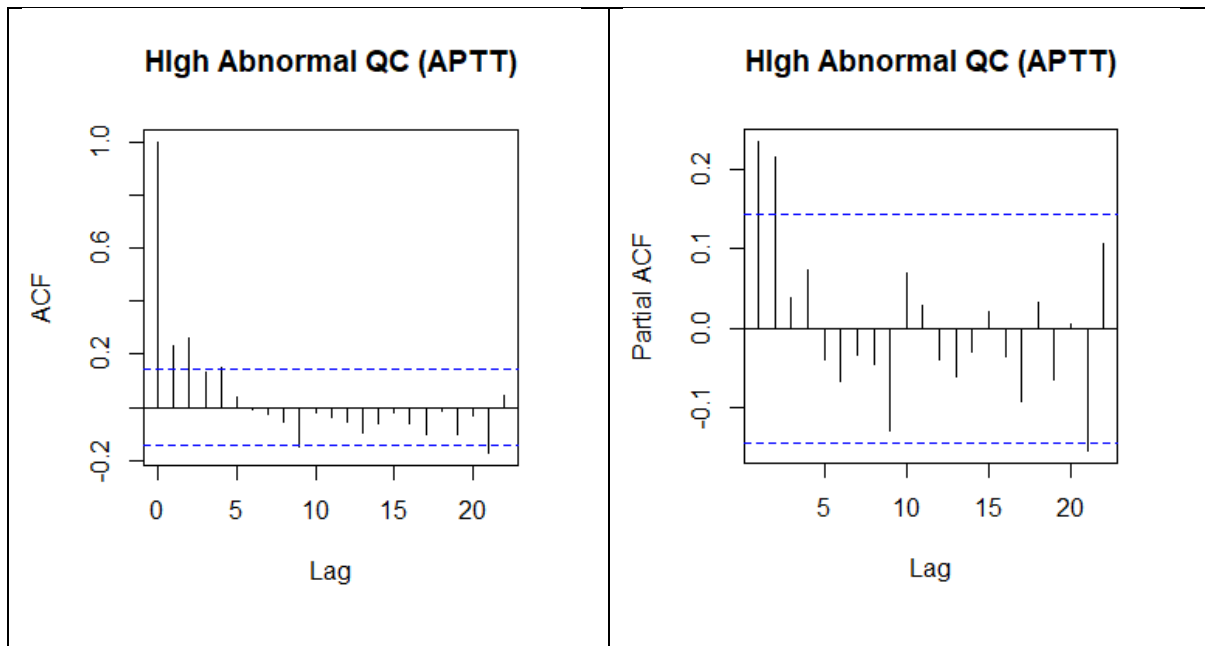


Figure 42: Autocorrelation in APTT IQC.

(Left) Autocorrelation function (ACF) plot for APTT High Abnormal Internal Quality Control data with 95% confidence intervals (dashed blue lines). Bars extending beyond the limits indicate statistically significant autocorrelation. (Right) Partial autocorrelation function (PACF) plot for the same data, highlighting lag 1 as the only significant partial correlation after accounting for earlier lags.

Figure 42 shows the ACF of IQC data from the APTT assay. Statistically significant autocorrelation was observed at lags 1 and 2, with lags 3 and 4 also approaching the significance threshold. This indicates that each result is, to some extent, influenced by preceding measurements up to approximately 3–4 runs. The PACF shown in Figure 42 (right), isolates the direct correlation between observations after accounting for prior lags and confirms that only lag 1 remains significant, suggesting a first-order autoregressive structure.

7.4.2.1 One at a time ADAMTS13 activity assay method (Technozym[®] ELISA)

7.4.2.1.1 Low QC all data

Analysis of the ELISA method revealed notable autocorrelation in both the low and normal QC data when multiple reagent lots were included. For the low control, the ACF (Figure 43, left) demonstrated statistically significant autocorrelation up to lag 20. Similarly, the PACF (Figure 43, right) indicated partial autocorrelation up to approximately 4–5 lags, even after accounting for prior lag effects. These findings were unexpected given the one-at-a-time nature of ELISA assays.

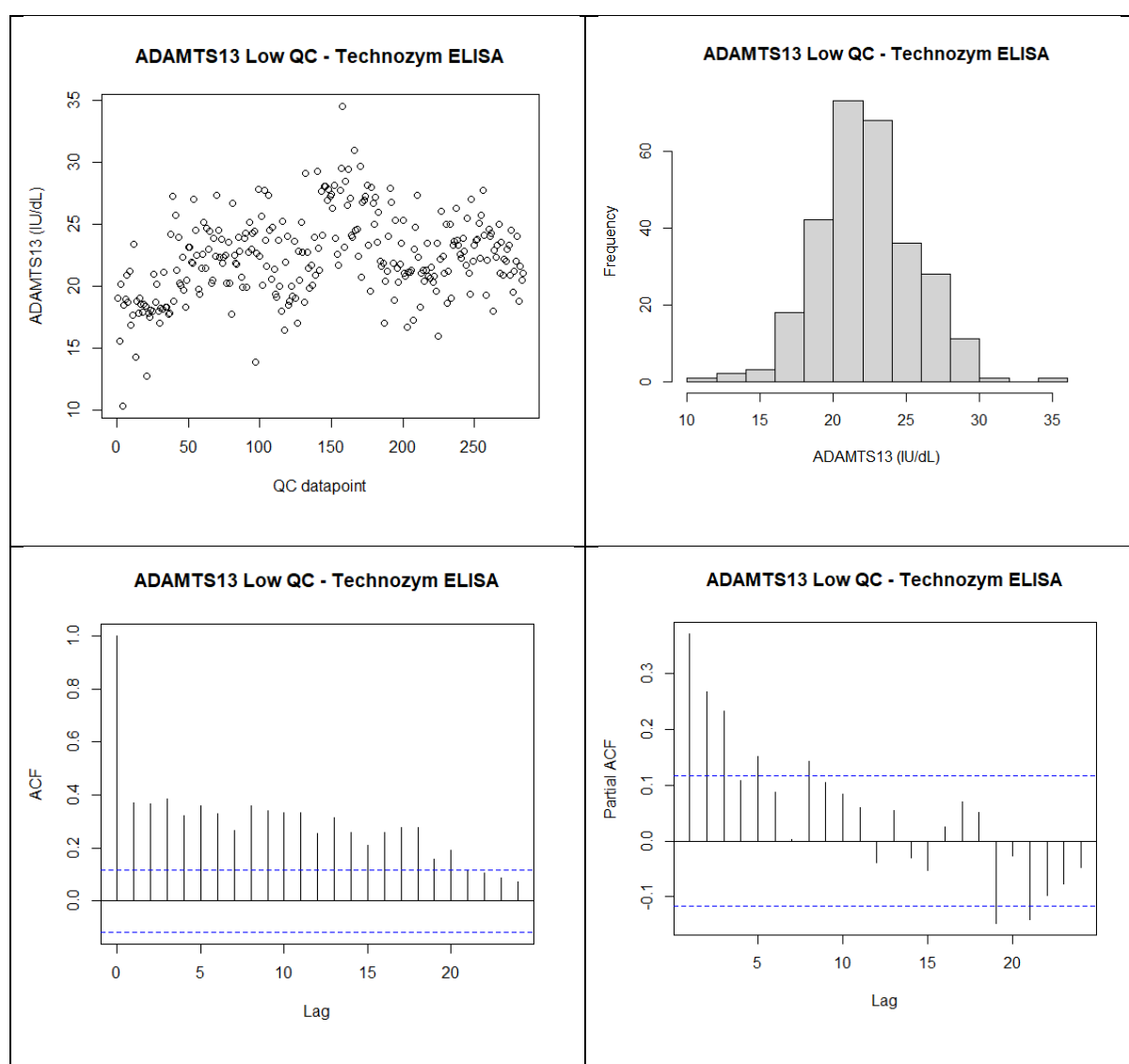


Figure 43: Autocorrelation in ADAMTS13 Low IQC (Technozym[®]).

(Left) ACF plot for ELISA low QC data across multiple reagent lots. (Right) PACF plot for the same dataset. Significant autocorrelation is observed up to lag 20, while partial autocorrelation remains up to lag 5.

Comparable results were observed for the normal control (Figure 44). The ACF remained statistically significant up to lag 17, while the PACF showed significance up to lag 3. However, when the analysis was restricted to a single lot number (Figure 45), both autocorrelation and partial autocorrelation disappeared for both control levels. This suggests that the observed dependencies were primarily due to mean shifts between lots rather than underlying process-related trends.

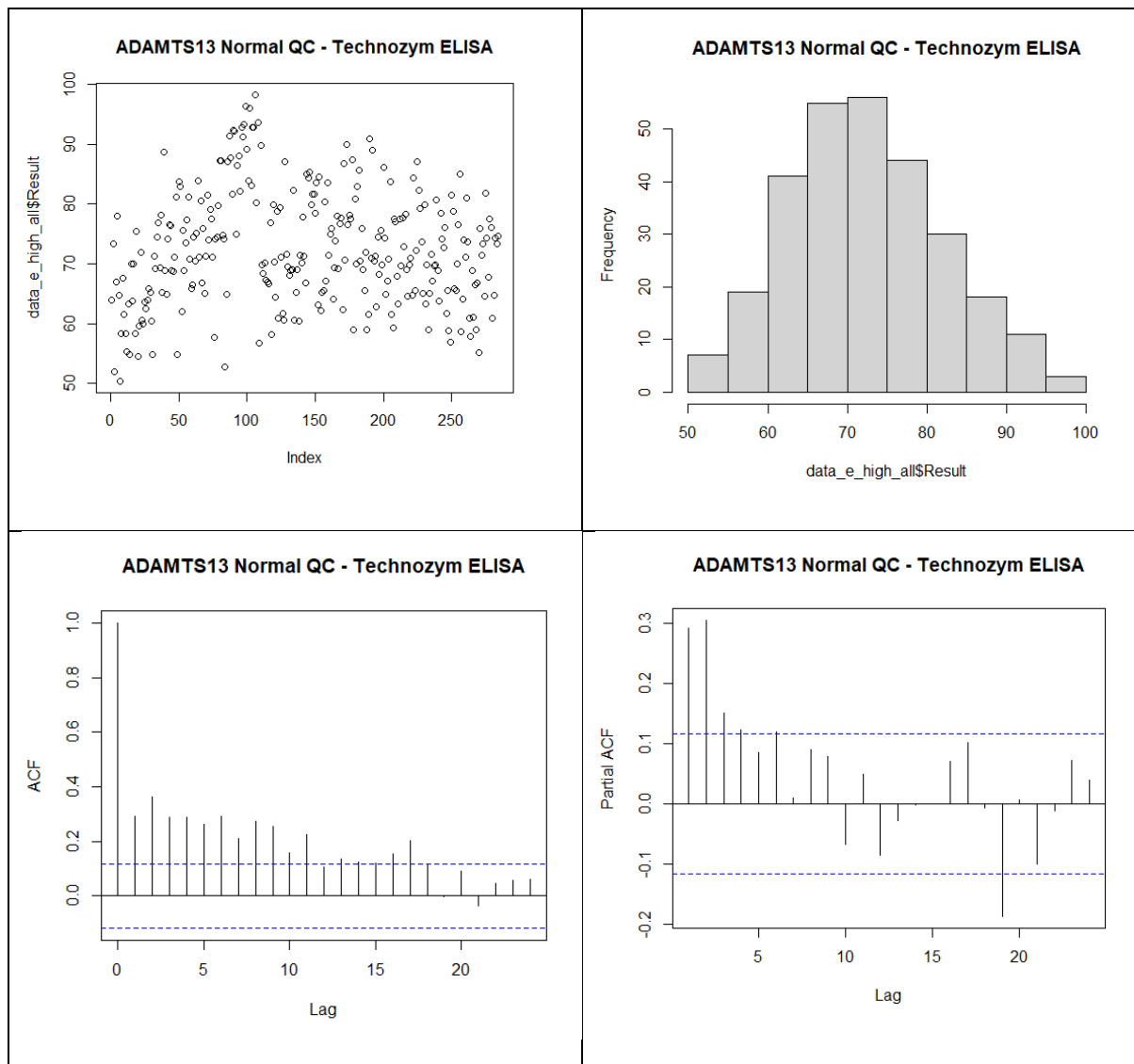


Figure 44: Autocorrelation in ADAMTS13 Normal IQC (Technozym ®).

Autocorrelation Function (ACF, left) and Partial Autocorrelation Function (PACF, right) plots for ELISA low and normal QC data over multiple lots.

7.4.2.1.2 Single lot number

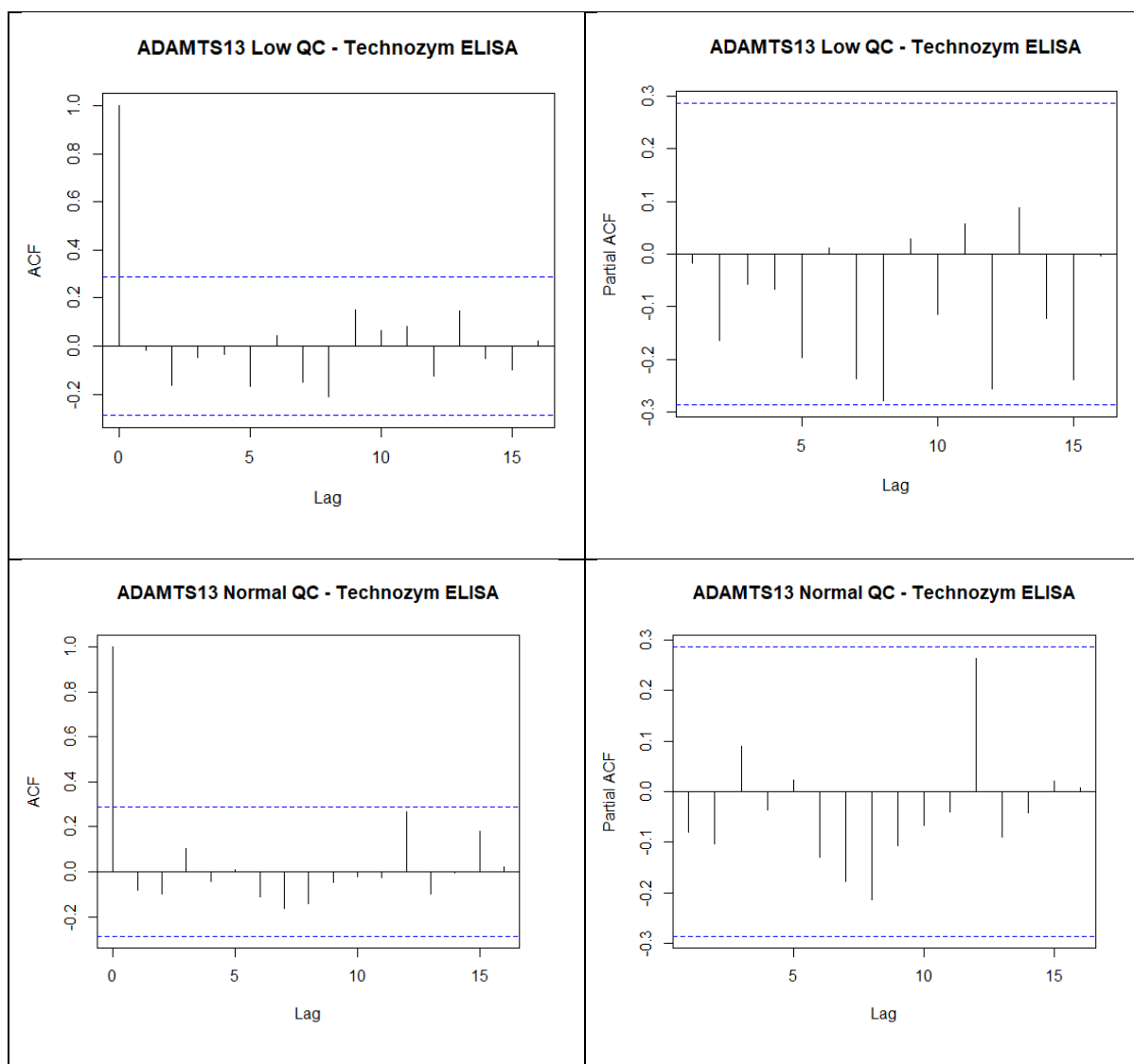


Figure 45: Autocorrelation in ADAMTS13 IQC (Technozym[®]) single lot number.

Autocorrelation Function (ACF, left) and Partial Autocorrelation Function (PACF, right) plots for ELISA low and normal QC data for a single lot number

7.4.2.2 Rapid ADAMTS13 activity assay (Werfen, HemosIL ADAMTS13 activity Acustar[®])

The AcuStar[®] platform showed similar behaviour to the ELISA method. When multiple lot numbers were included in the analysis, both low and normal QC data (Figure 46) exhibited strong autocorrelation and partial autocorrelation patterns. As with the ELISA, these effects were absent when the analysis was limited to a single lot number (Figure 45), supporting the interpretation that lot-to-lot variability introduces temporal structure into the IQC data.

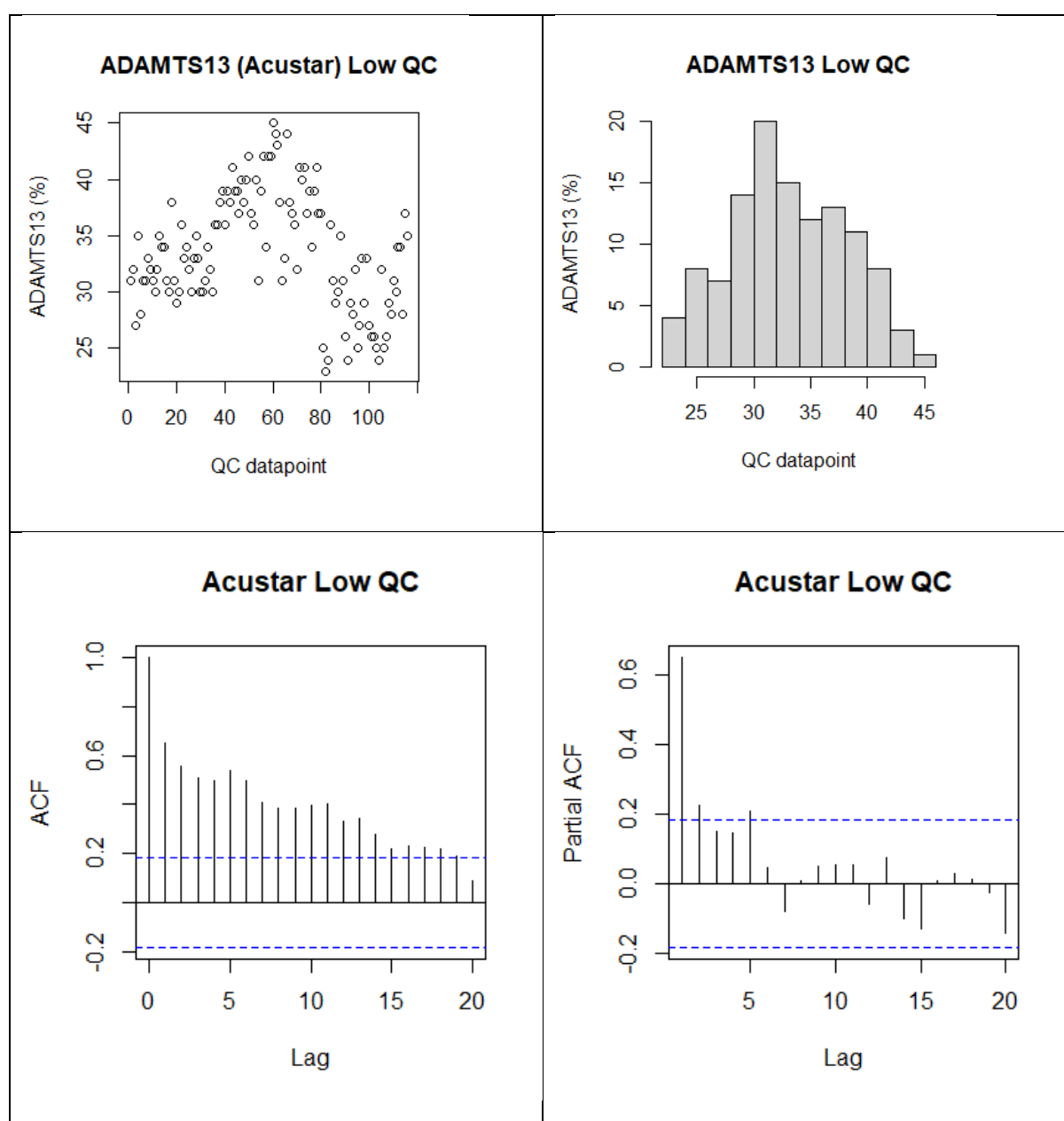


Figure 46: Autocorrelation in ADAMTS13 Low IQC (AcuStar[®]).

Autocorrelation Function (ACF, left) and Partial Autocorrelation Function (PACF, right) plots for AcuStar[®] low QC data over multiple lots.

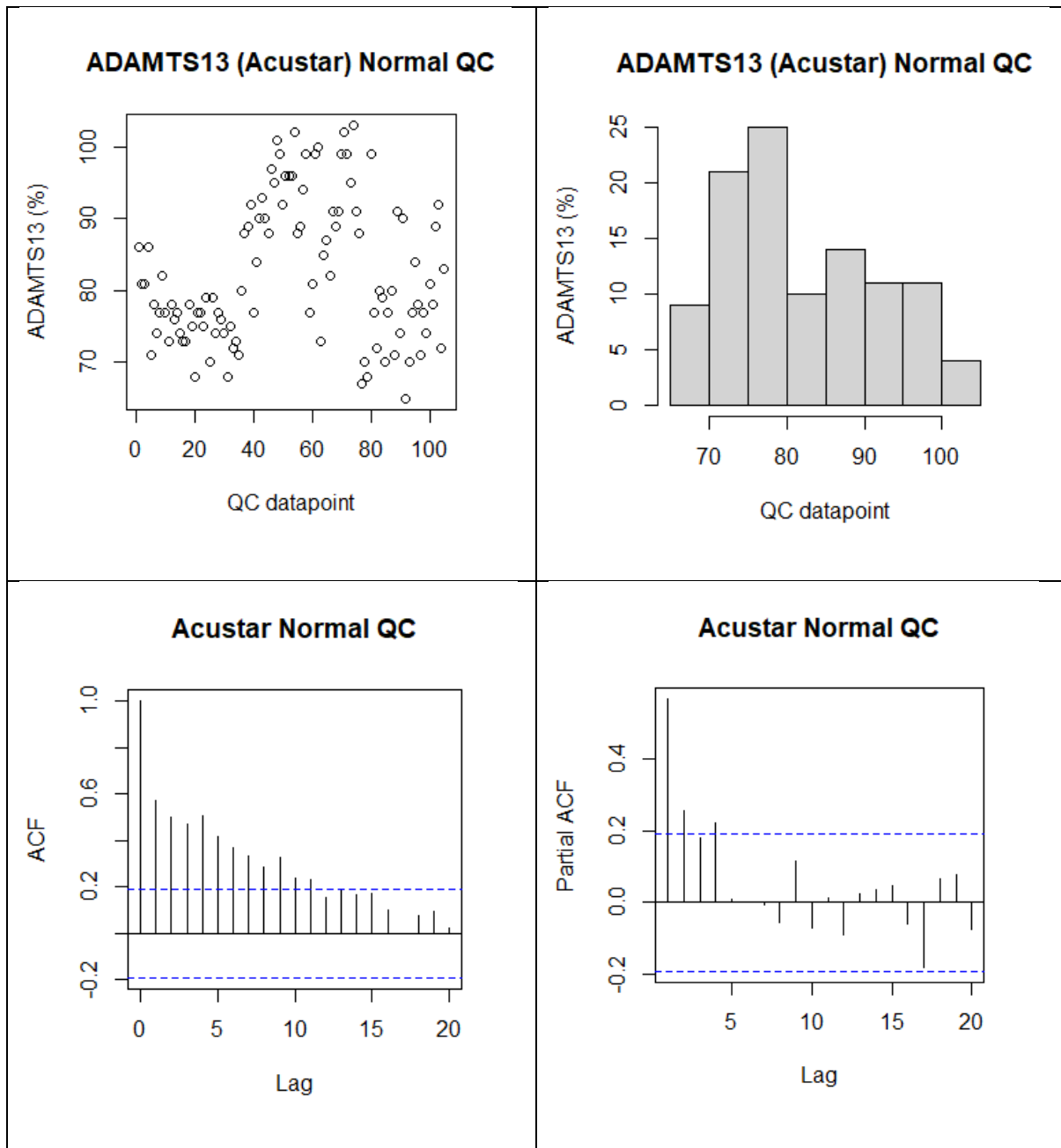


Figure 47: Autocorrelation in ADAMTS13 Normal IQC (AcuStar®).

ACF and PACF plots for AcuStar normal QC data across multiple lots. Significant autocorrelation patterns are visible, similar to those seen in ELISA data.

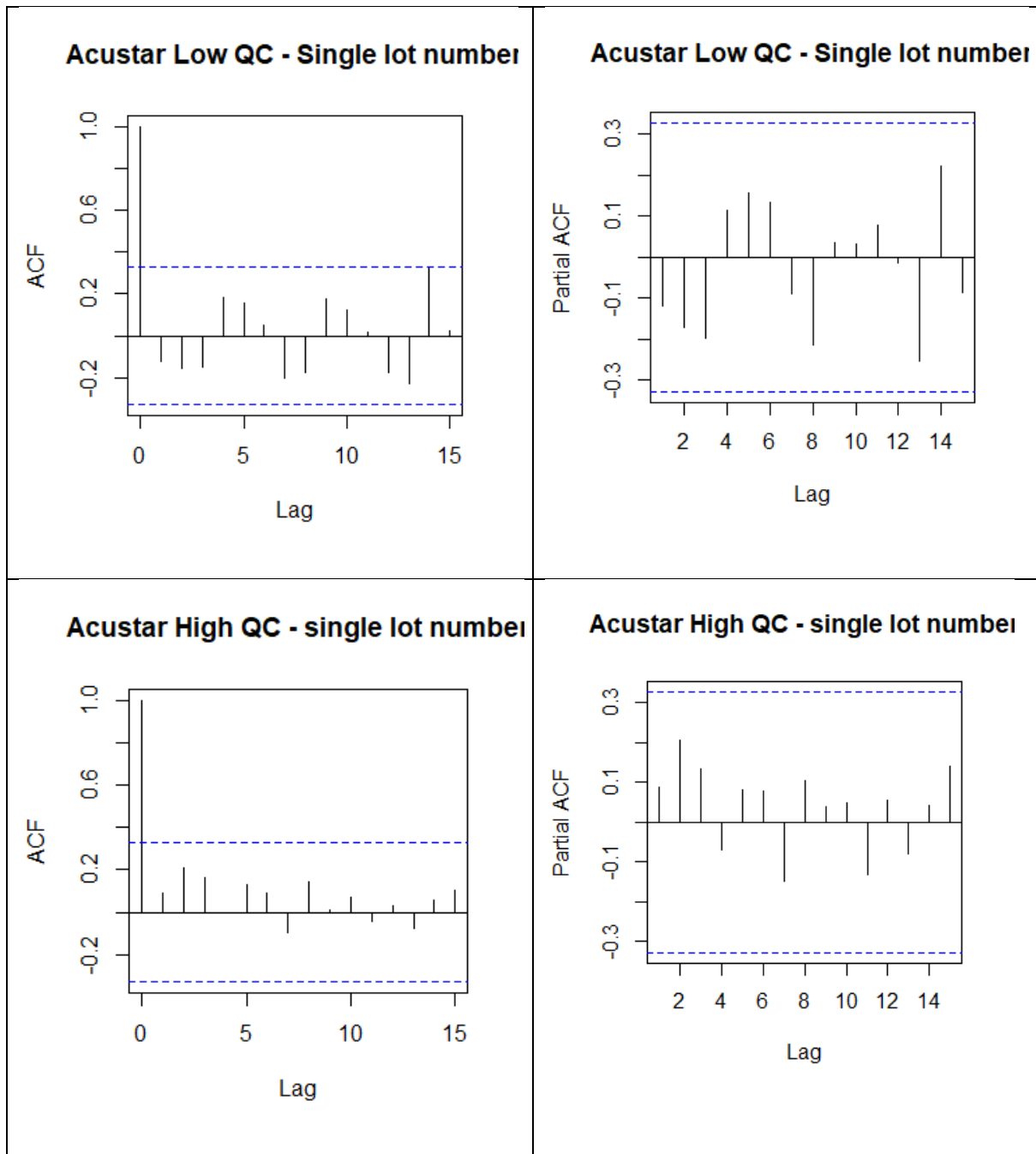


Figure 48: Autocorrelation in ADAMTS13 IQC (AcuStar[®]) single lot number.

ACF and PACF plots for AcuStar QC data restricted to a single lot. Absence of significant autocorrelation confirms that correlation in earlier plots was due to lot shifts.

7.5 Decomposition of IQC data used for u_{RW} determination

7.5.1 Methods

7.5.1.1 Study Design and Objective

This study employed time series decomposition to evaluate the contributions of trend and random variation within IQC data across multiple assay platforms. The primary objective was to refine the estimation of within-laboratory random uncertainty (u_{RW}) by isolating the residual component of variation after modelling potential temporal structures. Particular attention was given to the contribution of short-term trend development to overall imprecision, an aspect not easily discernible through traditional visualisation tools such as Levy-Jennings charts.

7.5.1.2 Assay Platforms and Quality Control Materials

The analysis encompassed three assay types representing a spectrum of testing approaches in clinical laboratories:

- A high-throughput, fully automated coagulation test, APTT, performed on the ACL-TOP 750CTS platform using HemosIL High Abnormal QC material.
- A manual one-at-a-time ELISA for ADAMTS13 activity (Technozym®), tested using both Low and Normal QC levels.
- An automated chemiluminescence assay for ADAMTS13 activity (HemosIL AcuStar®), similarly using Low and Normal QC levels.

The datasets used for each method were identical to those described in Section 7.4, including data extraction strategy, sample selection, and duration of monitoring. QC data from these platforms were routinely collected during ongoing laboratory operation and were suitable for assessing within-laboratory precision over extended timeframes.

7.5.1.3 Data Preparation and Preprocessing

To ensure the suitability of the data for time series analysis, a series of preprocessing steps were undertaken. For the aPTT assay, which involved multiple daily runs, QC data were aggregated to daily means to provide a consistent temporal structure. The ELISA and AcuStar[®] methods, performed once per run (typically once per day), were analysed as per-run series without additional aggregation.

All datasets were reviewed for completeness and consistency. Time stamps were aligned across records, and duplicate entries or gaps due to instrument downtime or reagent absence were identified and excluded. Outlier detection was performed using the Tukey method, applying the conventional $1.5 \times \text{IQR}$ rule. Outliers were removed only if attributable to known reagent errors, instrument faults, or failed internal QC performance reviews documented at the time of testing. No transformations were applied to the data, and all analyses were conducted on the original scale of measurement.

7.5.1.4 Time Series Decomposition Procedure

Seasonal and Trend decomposition using Loess (STL) was employed to separate each IQC series into its constituent components: trend, seasonal (if present), and residual. The STL procedure was implemented using an additive model, appropriate for data exhibiting roughly constant variance across the measurement range. STL was selected due to its ability to flexibly handle non-linear trend structures and accommodate non-equidistant sampling, particularly relevant for methods like AcuStar that are not run every day.

All analyses were conducted in R (version 4.3.1), using the *stats*, *forecast*, and *tsibble* packages. STL decomposition was implemented via the *stl()* function in the *stats* package. A weekly frequency cycle (frequency = 7) was assumed during model construction, though seasonal patterns were expected to be minimal and were not analysed further.

7.5.1.5 Parameters Extracted

From the STL decomposition output, both trend and residual components were extracted for further evaluation. The residual (random) component was summarised by its mean and SD. To determine the extent to which random variation contributed to overall imprecision, the SD of the residual component was expressed as a percentage of the total SD of the raw, undecomposed data series. Trend components were retained for analysis, as their relative contribution to total variability was a specific focus of this study.

No seasonal structures of statistical or clinical significance were observed. Normality testing of the residuals was performed using the Shapiro-Wilk test to confirm the assumption of Gaussian distribution, which underpins the interpretation of standard deviation as a proxy for random uncertainty.

7.5.1.6 Comparative Analysis

Comparative analyses were conducted to evaluate the proportion of random versus trend-derived variability across different platforms and QC levels. In particular, results from the Low and Normal QC levels for the Technozym ELISA and AcuStar platforms were compared to assess whether increased variability at low concentrations could be attributed to transient biases or persistent trends. The results were summarised as tabulated percentages of the total SD attributable to the residual (random) component.

Annotated trend plots were also produced to visually illustrate local inflection points, periods of apparent drift, and stability zones. These visual outputs were used to support interpretation of statistical findings and to identify assay- or lot-specific shifts not readily captured by traditional QC methods.

7.5.2 Results

7.5.2.1 Proof of principle using APTT (High volume automated routine assay)

To assess the feasibility of applying time series decomposition to IQC data, a high-frequency assay was selected as a proof of concept.

The APTT data set was chosen specifically to include a known performance shift caused by a QC lot change in late April 2021, allowing evaluation of whether decomposition could detect such a shift (Figure 49).

7.5.2.2 QC Performance February to August 2021

Visual inspection of the raw data revealed:

- A downward trend in mean aPTT times during the first two months (A)
- Stabilisation at a lower mean level (B)
- A sudden step change in the running mean following the QC lot change (C)
- Subsequent stabilisation at a new mean of approximately 53 seconds
- Occasional outliers (D and E)

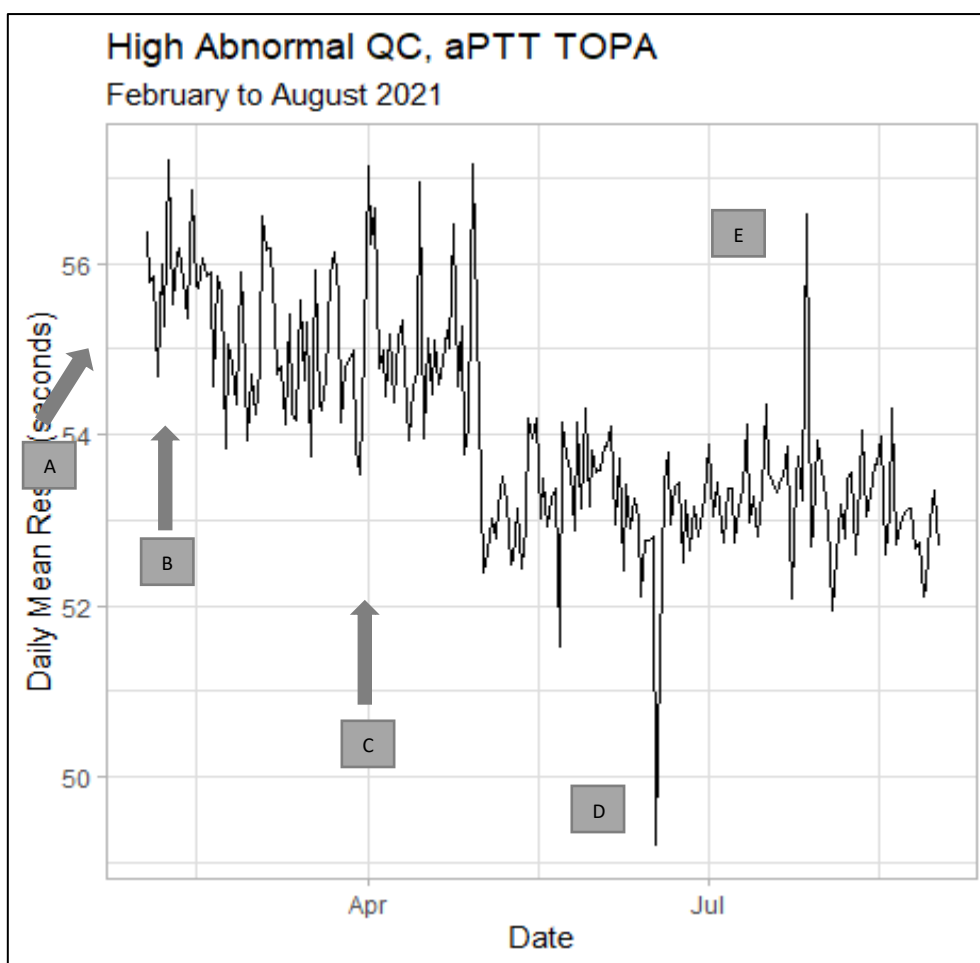


Figure 49: APTT High Abnormal Internal Quality Control data.

Six months consecutive Internal Quality Control data between March to September 2021 for the activated Partial Thromboplastin Time (APTT). Controls were performed on the ACL-TOP 750CTS analytical platform using HemosIL SynthaSIL reagent and HemosIL High Abnormal QC. As multiple quality control values were available for each date, data was represented as the daily mean to facilitate modelling. The arrow indicates a lot change of aPTT reagent at the end of April 2021 (indicated by the arrow).

Two outlier results were observed post-lot change (D and E), each corresponding to documented reconstitution errors. These were removed prior to modelling, based on detection via the Tukey method ($1.5 \times \text{IQR}$ criterion) and confirmed root cause analysis.

The apparent change in variability across lot numbers warranted further analysis. While there was a higher degree of dispersion before the lot change, the overall variance was judged sufficiently consistent (i.e., heteroscedasticity was not observed) to justify a unified model.

7.5.2.3 Decomposition of QC data from February to August 2021 (APTT assay)

7.5.2.3.1 Trend

The underlying noise of the data makes it challenging to detect short term trends other than the lot change (Figure 49). To identify the short-term biases and trends, decomposition extracts that trend component. Several interesting characteristics are visible in the trend pane of the decomposed data. There are clear short-term biases being introduced, two of which are highlighted as short term downward shifts (represented by filled squares in Figure 50), the significant shift due to the lot change was identified as such (Filled triangle, Figure 50). Other areas of stable operation are also clear with variation around a stable mean (unfilled rectangle in Figure 50).

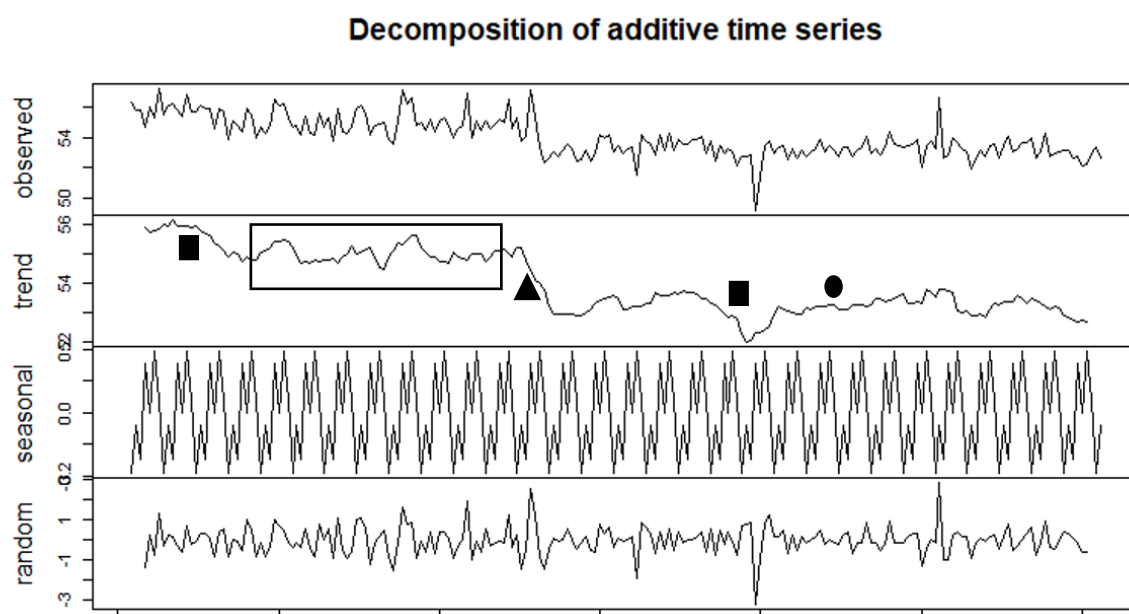


Figure 50: Decomposed APTT High Abnormal Internal Quality Control data.

Decomposed time series for HemosIL High Abnormal QC for the aPTT performed on the ACL-TOP750CTS system. ▲ = Shift due to lot change. ■ = short term downward trend, ● = Gradual increasing trend, □ = stable operation at a consistent mean.

7.5.2.3.2 Random variability

To quantify the proportion of variability attributable to the random (residual) component, the standard deviation of the residual component (random) was: 0.708 seconds and the standard deviation of the original (non-decomposed) aPTT data was: 1.273 seconds. The mean of the decomposed random values was negligible at -0.008 seconds.

Accordingly, the total variability in the original dataset could be attributed to true random variation after removal of trend and short-term biases.

$$\text{Random proportion} = (0.708/1.273) \times 100 = 55.6\%$$

This finding highlights that nearly half of the apparent variability in raw QC data may be due to non-random sources, such as drift, bias, or reagent-related shifts—components that are typically conflated with imprecision in traditional statistical approaches to u_{RW} estimation.

One at a time ADAMTS13 activity assay method (Technozym[®] ELISA)

Data were extracted for two QC levels (Low and Normal), each derived from two reagent lots again to detect the shift of mean associated with the lot change. Due to the manual nature of the assay and lower testing frequency, the data set contained fewer points over an extended period, but still allowed for time series modelling after preprocessing and daily aggregation where needed.

Figure 51, below, represents the decomposed time series outputs for each QC level.

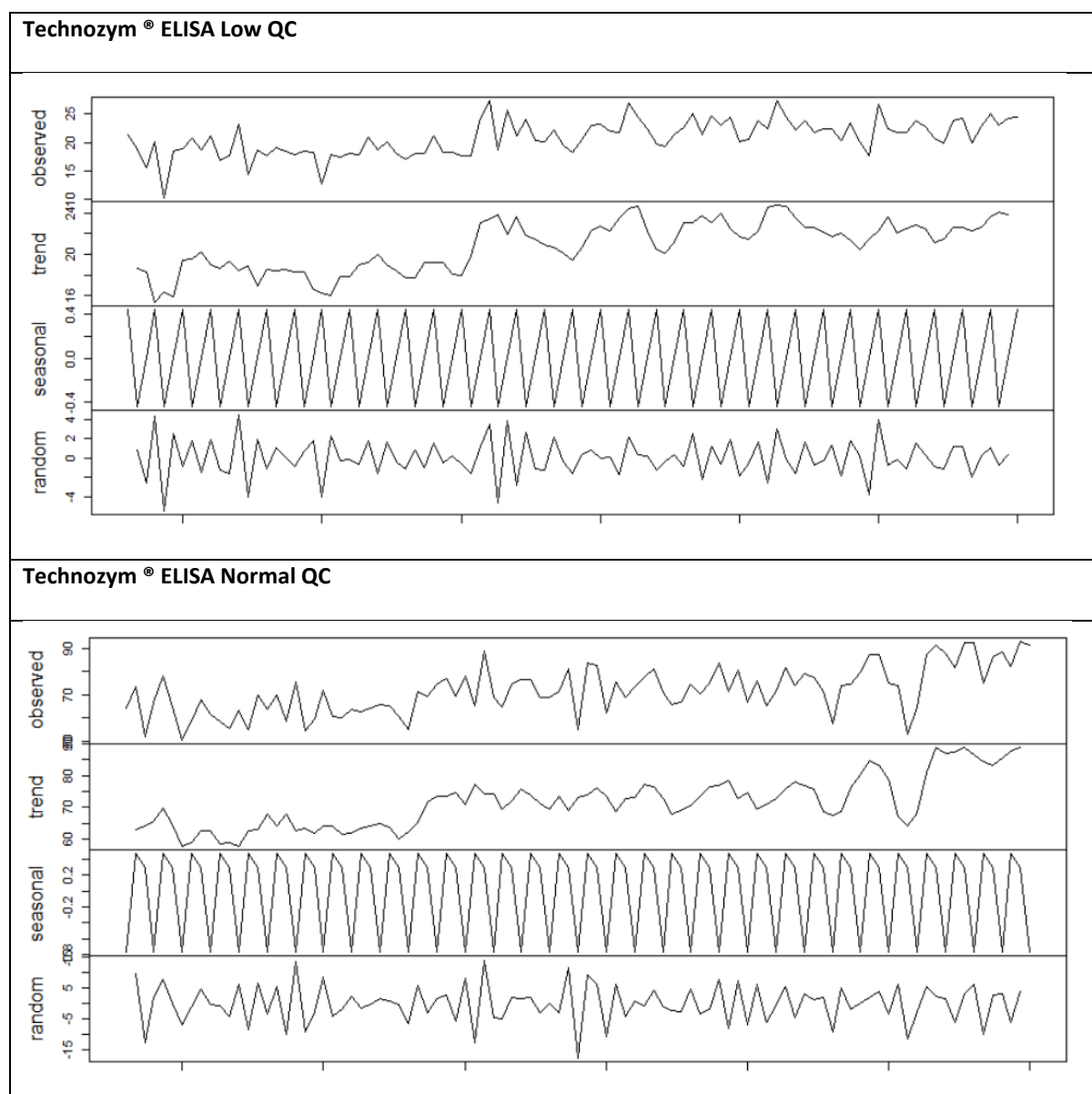


Figure 51: Decomposed Technozym®ELISA Low (Top) and Normal (Bottom) Quality control (QC).

The trend component for both QC levels revealed small, transient fluctuations initially followed by a shift in mean upon the lot change. There also appeared to be significant swings in the trend component at latter points in the dataset, particularly in the Normal QC. The minor shifts likely represent day-to-day variation in manual handling, incubation, or plate washing procedures, which are not easily standardised in manual ELISA protocols.

Table 37: Random variation contribution for Technozym[®] ELISA Quality Control (QC).

Calculations are based on overall standard deviation (SD Raw) and decomposed random components (SD random)

QC Level	SD (Random)	SD (Raw Data)	% Random Contribution
Low	0.122	0.212	57.7%
Normal	0.139	0.219	63.6%

These findings suggest that between 57% and 64% of the observed variation can be attributed to true random error in the ELISA method (Table 37). The remaining 36–43% is likely due to short-term biases or manual procedural inconsistencies not quantified by traditional imprecision metrics.

7.5.2.4 Rapid ADAMTS13 activity assay method (Werfen, HemosIL ADAMTS13 activity AcuStar[®])

The decomposition revealed significant variability across both QC levels both within the raw data and when decomposing the trend. Interestingly the shifts observed in the data are not due to a lot change as was previously the case. This was reflective of a genuine drift and/or shift in underlying performance exemplified by the return to the baseline trend level after time.

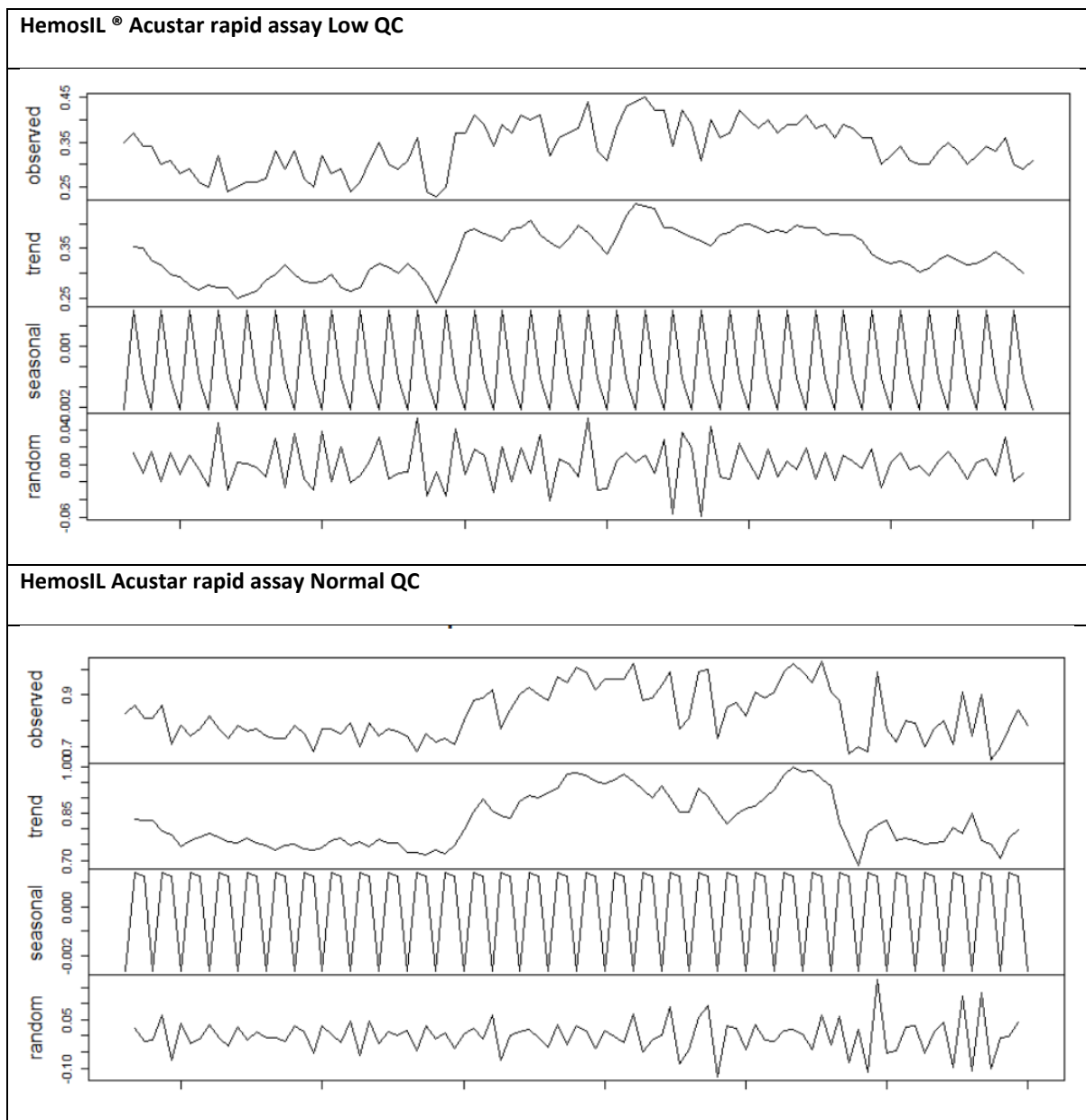


Figure 52: Decomposed Internal Quality Control - HemoSIL AcuStar ADAMTS13 activity assay ®

Decomposed Time Series analysis of ADAMTS13 activity Quality Control data performed using the HemoSIL AcuStar ADAMTS13 activity assay ®. (A) – Low vWF quality control data, (B) – Normal vWF Quality Control data

Performance is clearly significantly degraded at later time points compared to early in the study (Figure 52), this is itself reflect in the random component of the decomposition where increased random variability (and therefore by extension u_{RW}) is seen. Quantification of the random component yielded that shown in Table 38.

Table 38: Random variability in internal quality control (IQC) measurements.

SD – Standard deviation, QC – Quality Control

QC Level	SD (Random)	SD (Raw Data)	% Random Contribution
Low	0.103	0.199	51.9%
Normal	0.112	0.201	55.8%

Approximately 52–56% of the overall variation in QC values was attributable to true random variability (Table 38), with the remainder likely explained by short-lived shifts, calibration adjustments, or minor automated system adaptations that would typically go undetected in traditional QC evaluation. This was consistent with that seen for the APTT and ELISA methods. Interestingly stratifying on late data only, where the trend component had returned to baseline and random component had increased, the contribution of decomposed random variability to the overall variability increased to 74%.

7.5.2.5 Comparison across methods

The proportion of overall variability attributable to random variation differed modestly across assays and QC levels.

Table 39: True random variability in APTT and ADAMTS13 by two methods.

IQC - internal quality control

Assay	IQC Level	Random contribution to overall variability
aPTT	High Abnormal	55.57%
Technozym [®] ELISA	Low	57.68%
	Normal	63.57%
HemosIL [®] ADAMTS13 Activity	Low	51.94%
	Normal	55.81%

For the aPTT assay, 55.57% of total variability was due to random effects (Table 39). The Technozym[®] ELISA exhibited a higher random contribution, particularly at the Normal QC level (63.57%) compared to the Low QC level (57.68%). In contrast, the HemosIL[®] ADAMTS13 Activity assay showed slightly lower random contributions, with 51.94% for the Low QC and 55.81% for the Normal QC level.

7.6 Chapter 7 Discussion

This chapter demonstrated, for the first time in any medical laboratory context, the combined application of autocorrelation analysis and time series decomposition as a methodological framework for interrogating the structure of internal quality control data used in measurement uncertainty estimation. The findings challenge the prevailing assumption that IQC data are independent and identically distributed (i.i.d.) and suggest that standard deviation (SD)-based approaches may systematically overestimate the within-laboratory random uncertainty component (u_{RW}) by incorporating systematic variation. This represents a significant methodological contribution to uncertainty estimation and laboratory quality assurance.

Initial runs tests and subsequent autocorrelation analyses revealed significant temporal dependencies in the AcuStar IQC data, indicating that consecutive measurements were not statistically independent. This temporal structure is likely attributable to systematic factors inherent to automated analysers, such as instrument drift and on-board reagent stability. In contrast, the ELISA-based ADAMTS13 method showed no evidence of sequential dependence, consistent with its one-at-a-time, manual processing workflow. These findings demonstrate that analytical format influences the temporal structure of IQC datasets, with continuous, automated systems inherently more prone to autocorrelation effects. As a result, the i.i.d. assumption underlying the use of SD as a surrogate for random variability is not consistently valid, particularly in automated platforms—a concern previously acknowledged in ISO/TS 20914:2019 (161).

Further analysis confirmed that lot-to-lot variation was a principal driver of observed autocorrelation in both the ELISA and AcuStar® data. Aggregation of IQC data across reagent lots introduced systematic shifts that manifested as autocorrelation, while analysis restricted to individual lots eliminated these dependencies. These findings support existing ISO/TS 20914:2019 guidance recommending lot-specific calculation of u_{RW} , ensuring that systematic shifts between lots do not distort long-term imprecision estimates.

The application of time series decomposition provided additional insight into the internal structure of IQC datasets, separating variability into random, trend, and seasonal components across the APTT, ELISA, and AcuStar® platforms. Notably, decomposition consistently demonstrated that only approximately 50–65% of total variability represented random variation, with the remaining variance arising from systematic sources. This finding does not contradict current top-down uncertainty estimation practice; rather, it quantifies a phenomenon implicitly accepted within ISO/TS 20914:2019—that systematic influences become absorbed into long-term variability estimates, contributing to the apparent random variation captured within u_{RW} (154, 188).

This study brings this practical compromise into sharp focus, raising a conceptual question regarding the nature of uncertainty estimation in medical laboratories. From a strictly metrological perspective, random and systematic sources of variation should be treated distinctly to ensure that u_{RW} represents only random error. However, the top-down approach pragmatically aggregates all long-term sources of variability, including minor systematic effects, into a single uncertainty estimate for practical and operational reasons. The findings presented here directly quantify this trade-off, demonstrating that a substantial proportion of what is currently classified as ‘random’ uncertainty is, in fact, systematic in origin.

Whether this methodological approximation remains acceptable, or whether future frameworks should seek to distinguish these components explicitly, remains an open question. While current practice provides a practical solution that avoids the complexity of full bottom-up uncertainty decomposition, laboratories and guideline developers may need to consider whether this convenience justifies the potential for overestimated uncertainty. This issue is particularly relevant in assays operating near clinical decision limits, where the consequences of inflated uncertainty, while not formally evaluated in this study, could affect confidence in result interpretation.

In summary, the combined application of autocorrelation analysis and time series decomposition, as demonstrated in this chapter, represents a novel and exploratory framework for critically evaluating

IQC data used in MU estimation. This study confirms that IQC datasets routinely violate the statistical assumptions underpinning current top-down MU estimation methods, particularly in automated, continuous analyser platforms. It also demonstrates that conventional SD-based methods likely overestimate uRW by absorbing systematic variation into random uncertainty estimates.

The framework presented here should be regarded as a proof-of-concept approach. While no formal recommendations for routine implementation are made, the consistent findings across multiple assay formats—including high-volume coagulation testing, manual ELISA, and automated immunoassays—highlight the need for further investigation. Time series decomposition offers a structured methodology to quantify and separate systematic and random variability, addressing limitations of traditional QC monitoring approaches, such as Levy-Jennings charts, which cannot reliably distinguish transient biases from inherent system noise.

From a quality management perspective, isolating the true random component of measurement variability better aligns with the intent of ISO 15189:2022 and ISO/TS 20914:2019, supporting the development of MU estimates that reflect fitness for clinical use. Furthermore, integrating decomposed IQC metrics into automated MU calculation tools, as part of a future research agenda, could enable laboratories to refine uncertainty estimation processes and improve method surveillance over time.

While promising, this approach requires sufficient longitudinal IQC data and careful interpretation of seasonal components, which may reflect operational frequency rather than true cyclic variation. These limitations underscore the exploratory nature of the present work.

In conclusion, this study demonstrates that while current top-down approaches to MU estimation offer practical advantages, they are underpinned by a methodological compromise that allows systematic variation to be treated as random for convenience. This work quantifies the extent of that compromise for the first time, providing a novel analytical framework that challenges conventional assumptions and offers a foundation for future refinement of laboratory uncertainty estimation.

8 Chapter 8 – Calibration as a Source of Measurement

Uncertainty in ADAMTS13 Assays

8.1 Introduction

Following the identification of hidden contributors to measurement variability within internal quality control datasets in Chapter 7, this chapter examines whether calibration procedures function as an overlooked source of measurement uncertainty in ADAMTS13 activity assays. Current top-down MU estimation frameworks, including ISO/TS 20914:2019, treat calibration uncertainty solely as the uncertainty associated with assigned calibrator values (u_{cal}), if the calibration process itself does not contribute additional variability. However, in one-at-a-time immunoassays such as the Technozym® ADAMTS13 ELISA, calibration is performed independently for each assay run, raising the possibility that the calibration model itself acts as a dynamic, unrecognised source of measurement variability.

This chapter addresses this hypothesis through a four-stage approach. The first section explores how different calibration models affect interpolation uncertainty and curve behaviour within individual assay runs. By comparing commonly applied curve-fitting approaches—including four-parameter logistic (4PL), five-parameter logistic (5PL), cubic spline, point-to-point interpolation, and linear regression—the analysis illustrates how model selection affects result estimation, particularly in critical clinical regions. This mechanistic evaluation establishes the foundation for understanding how calibration models differ in their structural behaviour.

Building on this, the second section assesses whether calibration model choice contributes to real-world calibration variability across multiple assay runs. By analysing repeated calibrations across different models, this section investigates whether calibration model selection introduces systematic error and run-specific variability, representing a hidden structural contributor to between-run variability. The analysis aims to determine whether such model-dependent fluctuations are absorbed

into within-laboratory reproducibility (u_{RW}) and therefore misclassified as random variability within current uncertainty frameworks.

The third section progresses from comparative evaluation to quantification, applying Monte Carlo simulation methods using Statlia Matrix[®] software to assess calibration-derived interpolation uncertainty across the assay's analytical range. Using the 4PL model as an exemplar, cumulative interpolation variability is modelled across multiple runs, enabling calibration uncertainty to be quantified as a discrete contributor to overall measurement uncertainty. This approach supports the assessment of calibration performance against analytical performance specifications and explores whether calibration modelling effects should be incorporated into laboratory MU estimation frameworks.

Finally, the fourth section addresses calibration in automated systems, focusing on the HemosIL[®] ADAMTS13 assay. In such platforms, master calibration curves are designed to minimise calibration-related variability, yet local bias correction factors are routinely applied to adjust calibration output to laboratory-specific performance. This section evaluates whether local bias correction represents a hidden source of systematic uncertainty, contributing to measurement variability even in automated systems typically considered stable.

Together, these analyses reposition calibration from a background process to a potential structural contributor to measurement uncertainty in ADAMTS13 activity assays. By exploring calibration model selection, run-specific calibration variability, cumulative interpolation uncertainty, and bias correction in automated platforms, this chapter examines whether calibration procedures—manual and automated—should be reconsidered as active contributors to result variability and uncertainty estimation.

8.2 Aims

This chapter aims to determine whether calibration processes function as an unrecognised source of measurement variability in ADAMTS13 activity assays. It examines the hypothesis that calibration model selection contributes both systematic error and between-run variability in one-at-a-time assays, representing a hidden structural source of uncertainty. The chapter seeks to challenge current assumptions regarding calibration and assess whether calibration procedures should be considered as active contributors to measurement uncertainty in both manual and automated assay workflows.

8.3 The Impact of Calibration Model Selection on Interpolation

Uncertainty

To examine whether calibration model choice affects result generation, this section evaluates interpolation behaviour across different curve-fitting models within individual calibration runs. By comparing linear and polynomial regression models, the analysis explores whether inherent differences in interpolation introduce structural variability in result estimation, forming the mechanistic basis for subsequent analyses.

8.3.1 Methods

To assess the impact of calibration model selection on MU, a single assay run of the Technozym ADAMTS13 ELISA was used as an exemplar. This assay was chosen to evaluate how different regression models influence confidence intervals around interpolated concentrations. The analysis was conducted using R statistical software, incorporating linear regression modelling and data visualisation with the *ggplot2* package to compare confidence interval widths at different concentrations.

8.3.1.1 Statistical Methods

OLS regression was applied to fit a linear calibration curve, estimating the relationship between absorbance values (response) and known standard concentrations. The regression model was expressed as:

$$y = \beta_0 + \beta_1 x + \epsilon$$

where y represents the measured absorbance, x the standard concentration, β_0 the intercept, β_1 the slope, and ϵ the residual error. The confidence bounds around the regression line were calculated using the standard error of the parameter estimates, generating 95% confidence intervals to assess the reliability of interpolated values at different concentrations.

To visualise uncertainty propagation, the *ggplot2* package in R was used to generate calibration curves with confidence bands. This enabled comparison of confidence interval width at low, mid, and high concentrations, highlighting areas where uncertainty is most pronounced. Residual diagnostics were performed to assess model validity, checking for violations of linear regression assumptions such as homoscedasticity and normality of residuals.

8.3.1.2 Interpolation Error in OLS when Estimating x from y

To estimate the uncertainty in x (concentration) given a measured y (absorbance), the linear regression equation is inverted and the uncertainty propagated. The standard error of the estimated x is given by:

$$SE(x) = SEP / |b|$$

where:

- x is the estimated concentration from the measured absorbance y
- b is the slope of the OLS regression line ($y = \beta_0 + \beta_1 x + \epsilon$)
- SEP is the standard error of prediction, which is calculated as:

$$SEP = s \sqrt{1 + \frac{1}{n} + \frac{(y^* - \bar{y})^2}{\sum (y_i - \bar{y})^2}}$$

Equation 21: Interpolation Error in OLS when Estimating x from y.

where:

- s = residual standard deviation (root mean square error, RMSE),
- n = number of calibration points,
- y^* = measured absorbance,
- $\bar{y}$ = mean absorbance of calibration points,
- y_i = individual calibration absorbance values.

Dividing by the absolute value of the slope ($|b|$) converts the uncertainty from the y-scale to the x-scale, allowing for a direct estimation of uncertainty in the predicted concentration from a given absorbance measurement. This transformation highlights that uncertainty is not uniform across the calibration range; it is most pronounced at the extremes, particularly at lower ADAMTS13 concentrations, where even small fluctuations in absorbance can lead to significant variations in the interpolated concentration. Consequently, this approach relies on the assumption that a linear regression model is appropriate for the data. However, if the underlying relationship is non-linear, alternative models such as polynomial or logistic regression may be required to provide a more accurate estimation of concentration uncertainty.

8.3.1.3 Higher-Order Regression Models

Beyond linear regression, quadratic and cubic polynomial models were applied to the same data to evaluate whether higher-order fits reduced confidence interval widths and improved interpolation accuracy. These models were expressed as:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

where additional terms ($\beta_2 x^2$ and $\beta_3 x^3$) allowed for additional curvature in the calibration function.

The width of confidence intervals for linear and polynomial models was compared at clinically relevant concentration ranges to assess whether non-linear fits provided more precise interpolations.

8.3.1.4 Calibration variability across multiple assay runs

30 consecutive calibration curves from the Technozym® ADAMTS13 ELISA were analysed, modelling absorbance values against calibrator concentrations to assess overall variability. Additionally, residuals were examined across the measurement range, identifying patterns of error propagation that may influence MU estimation. By comparing inter-assay variability to single-run uncertainty, this approach aims to determine whether aggregating calibration curves provides a more stable and representative measure of MU for ADAMTS13 activity assays.

8.3.2 Results

8.3.2.1 Linear regression calibration curve fit

A single assay run for the Technozym[®] ELISA was used as an exemplar to evaluate the imprecision of interpolated results from a standard curve fitted using different linear models.

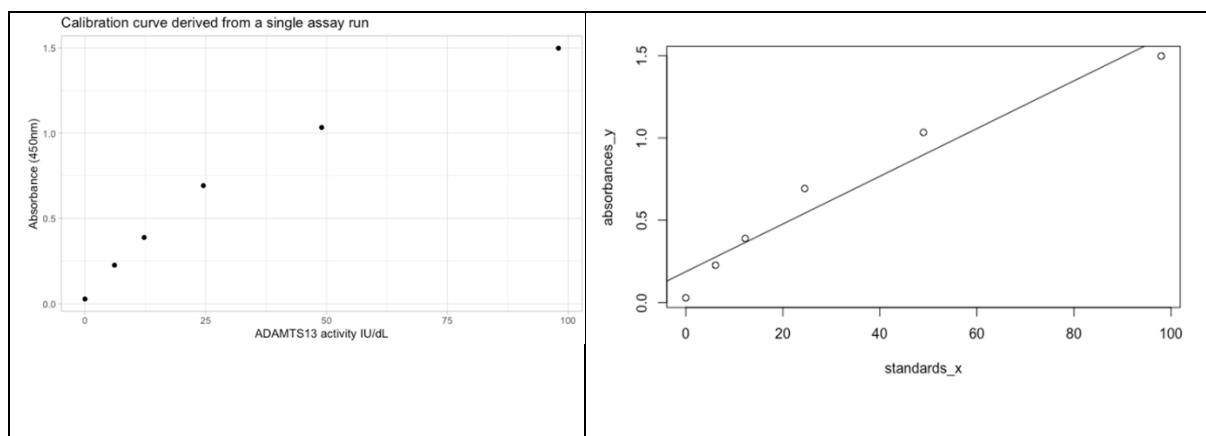


Figure 53: Linear regression calibration fit - Technozym[®] ADAMTS13

(Left) - Example calibration curve for a single assay run using the Technozym[®] ADAMTS13 activity ELISA (Technoclone, Austria). (Right) - Linear regression fit by ordinary least squares regression using Maximum Likelihood Estimation for parameter estimation.

The calibration points visually deviate from a linear relationship, indicating that a linear regression model is inappropriate for this dataset (Figure 53, right). Applying OLS regression results in systematic biases across the measurement range: low concentrations are overestimated, high concentrations are also overestimated, and mid-range values (15–50 IU/dL) are consistently underestimated. This pattern suggests that the linear model fails to adequately capture the underlying relationship between absorbance and concentration.

The consequence of this misfit is the introduction of excessive interpolation uncertainty, particularly at clinically significant decision points. When a model does not accurately represent the data,

confidence intervals around interpolated results become unreliable, leading to potential misinterpretation of patient results. This issue becomes evident when examining regression residuals (Figure 54), which exhibit systematic deviations across the calibration range, further reinforcing the inadequacy of a linear model for ADAMTS13 assay calibration.

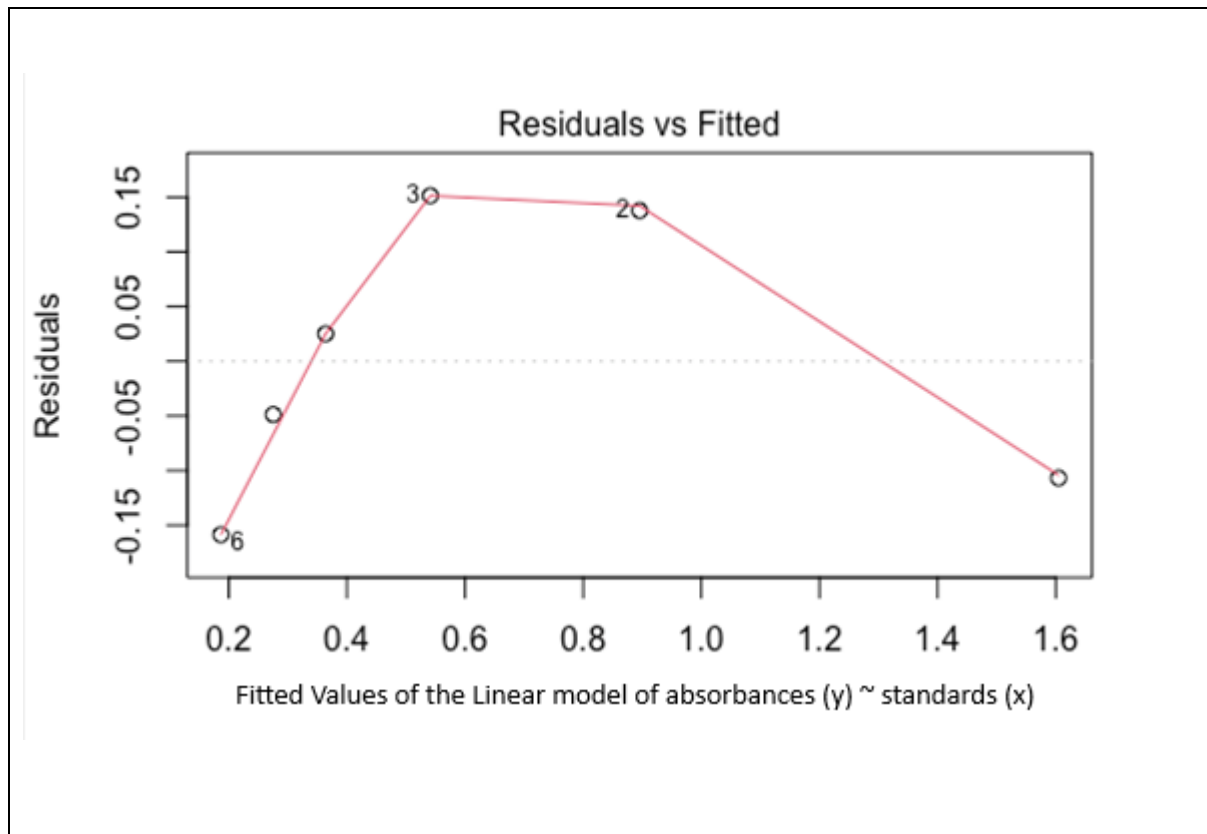


Figure 54: Linear calibration residual plot - Technozym ® ADAMTS13

Residual diagnostic plot of fitted v measured values using ordinary least squares linear regression. A clear non-linear pattern in the residual values indicates a linear regression fit by this method is not statistically valid due to violation of the assumptions of the model.

8.3.2.2 Interpolation of a result ignoring MU in the calibrator and measured value

The calibration function, as an optimised representation of the underlying “true” relationship is used to interpolate unknown results. Common approaches in commercial software, and when performing the analysis manually, is to ignore uncertainty in both the calibrators and measured values. In doing so the approach (and consequent interpolated result) shown in Figure 55 is achieved.

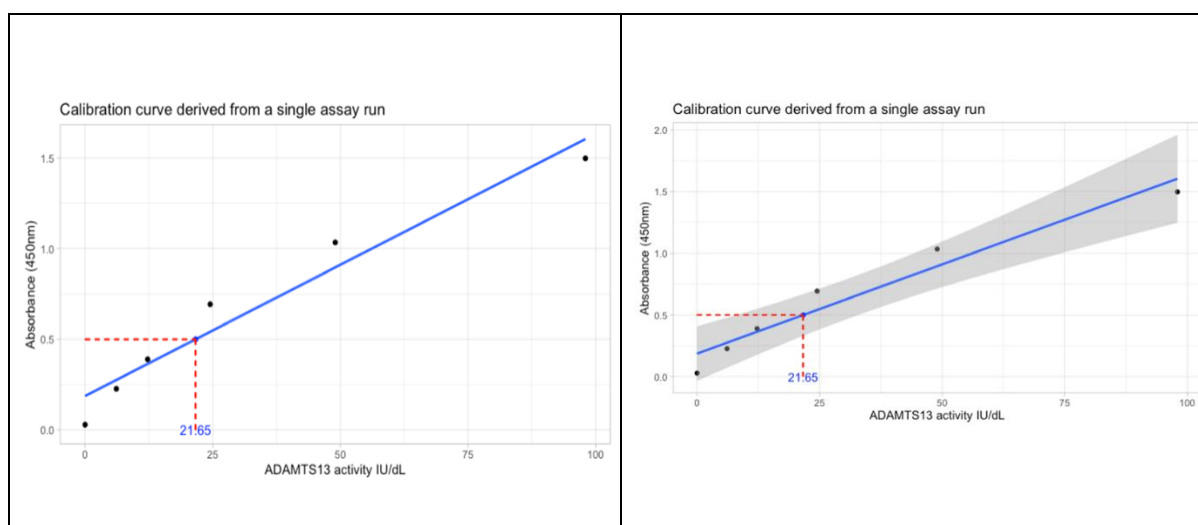


Figure 55: Interpolation of linear calibration - Technozym® ADAMTS13

(Left) - Interpolation of an example result from linear least squares regression fit. Such an interpolation does not account for uncertainty in the measured value (Absorbance - y-axis) nor uncertainty in the parameters fitting the standard curve.

(Right) - representation of the interpolation with uncertainty associated with parameter estimates included. The frequentist confidence intervals demonstrate the classical "bow-tie" shape across the measurement range

As shown in the figure above, the characteristic bow-tie shape of the confidence intervals in an OLS illustrates the increasing uncertainty at the extremes of the measurement range (right pane). While this representation is more informative than a simple point estimate that does not account for parameter uncertainty, it highlights a critical issue: uncertainty is greatest where diagnostic accuracy is most needed. In ADAMTS13 activity measurements the greatest uncertainty occurs at low concentrations, where precise interpretation is essential. The observed pattern confirms that a linear

regression model is inappropriate for these assays, as it fails to capture the true relationship between absorbance and concentration, violating key statistical assumptions (Figure 55).

Applying the standard error of prediction equation, the interpolated concentration for a given absorbance value—rather than a single-point estimate of 21 IU/dL—is shown to fall within a wide range of 10–32 IU/dL. This is further widened if uncertainty in y is considered to be 7–37 IU/dL using 10% as an example standard uncertainty. This substantial variation underscores the lack of reliability of linear regression for interpolation in the presence of MU. When results vary so significantly at critical clinical thresholds, they lose diagnostic utility, making linear regression unsuitable for ADAMTS13 activity estimation at clinically relevant decision points.

To improve calibration accuracy and reduce interpolation error, higher-order regression models could be considered. By incorporating non-linearity into the calibration function, polynomial or logistic models may provide a more robust fit, better representing the underlying assay characteristics while minimising the uncertainty seen in linear regression models.

8.3.2.3 Calibration variability across multiple assay runs

While the previous section focused on estimating MU from a single calibration curve, this section explores an alternative approach—deriving assay-level MU by aggregating calibration curves from multiple runs. In clinical laboratories, calibration is performed repeatedly over time, and the variability across different assay runs can provide valuable insights into the robustness and reproducibility of the calibration process.

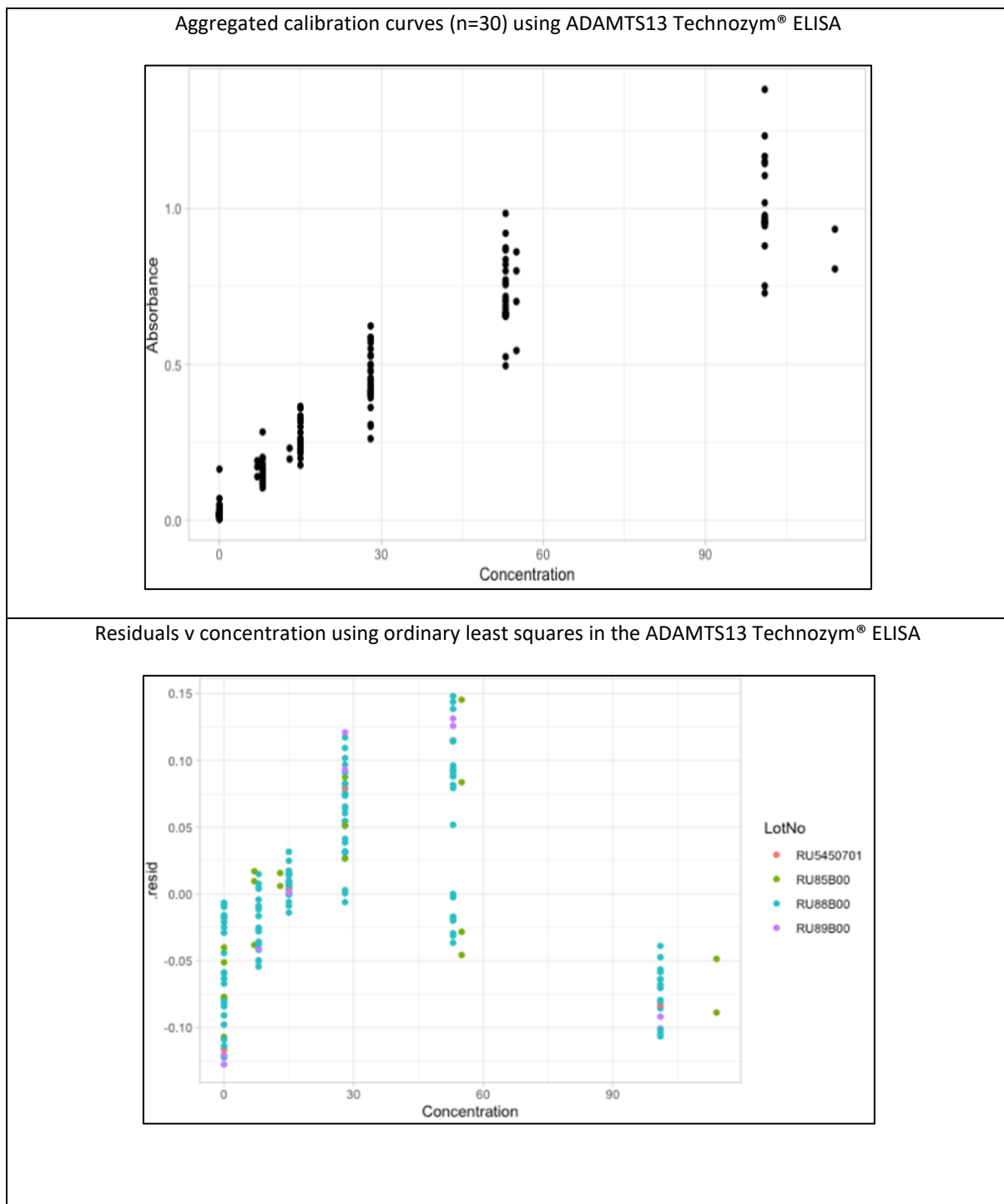


Figure 56: Aggregated absorbance responses - Technozym® ADAMTS13

(Top) - Aggregated calibration curves (n=30) demonstrating the range of values and different shapes of curves across different runs (Bottom) – Residuals v concentration using ordinary least squares in the ADAMTS13 Technozym® ELISA. A trend is clear with underestimation from the model that steadily resolves and continues to overestimate at concentrations > 30 IU/dL. This returns to underestimation at the area of the highest calibrator.

The aggregated responses from 30 consecutive calibration curves reveal a high degree of variability in calibration performance (Figure 56, left) in terms of absorbances measured. This variability is particularly pronounced at the upper end of the measurement range, where responses show greater spread across different assay runs. In contrast, responses at lower concentrations are more consistent, particularly around the clinical decision limit of 10 IU/dL, where maintaining precision is critical for accurate diagnosis. The data suggest a concentration-dependent trend in variability, with greater consistency at lower concentrations and increased variation at higher levels.

Examining the residuals from linear regression further highlights a similar concentration-dependent effect, but with different interpretation at clinical decision limits (Figure 56, right). At the lower end of the calibration range, residuals exhibit substantial variability, indicating a lack of reproducibility in calibration performance at low concentrations, in the presence of apparently reproducible absorbances (left). This suggests that uncertainty introduced to results in this region is not only significant but also difficult to predict, given the large degree of variability across assay runs, and it is introduced by the model used, not the measurement procedure itself. Notably, the pattern observed in a single calibration curve is replicated across all 30 calibrations, with the mid-range concentrations deviating most from the fitted regression line.

Interestingly, at the highest calibration level, residuals are comparatively lower than those seen in the mid-range concentrations. Similarly, while residuals at the lowest concentrations are also lower, they exhibit greater inconsistency than higher calibration points. This suggests that, despite a poor overall fit—both in single and aggregated calibrations—the extremes of the measurement range exhibit the lowest residuals, except for the lowest calibrator which as a blank is expected to have a degree of variability. These findings reinforce the limitations of a linear calibration approach, particularly in capturing variability across repeated assay runs.

8.3.2.4 Second order polynomial calibration curve fit

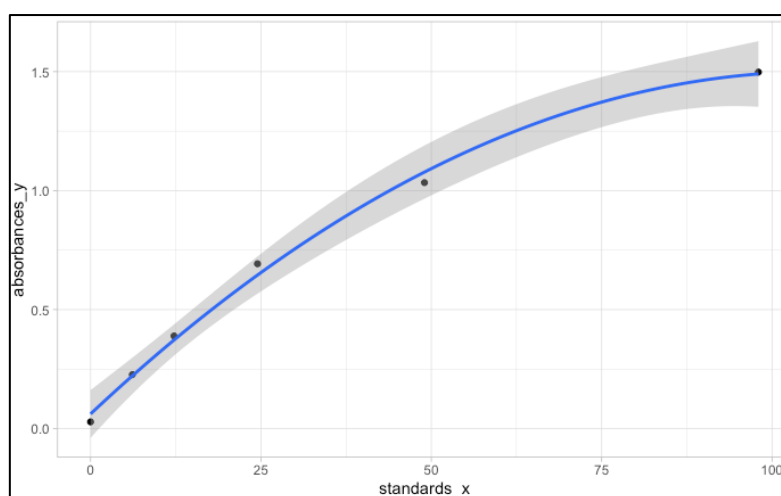


Figure 57: Quadratic polynomial calibration - Technozym® ADAMTS13

Second order polynomial fit to calibration data. This represents the curve as a non-linear function by adding a quadratic term to the formula of the curve.

The second-order polynomial function provides a noticeably better fit to the calibration data (Figure 57), as the measured points align more closely with the regression curve compared to the linear model. This improved fit is reflected in the narrower confidence intervals, which remain closer to the curve across the measurement range, thereby reducing interpolation uncertainty. The introduction of curvature into the model significantly improves accuracy, particularly at the extremes of the measurement range, where linear regression exhibited substantial deviations. While the widest confidence intervals still occur at the upper and lower extremes, their impact at low concentrations (<25 IU/dL) is far less pronounced than in the linear model, suggesting a more reliable estimation at clinically relevant decision points.

This reduction in uncertainty is particularly evident when comparing interpolated concentration ranges. For a given absorbance value, the second-order polynomial model predicts a range of 14–20 IU/dL, a considerable improvement over the 10–32 IU/dL range observed in the linear regression model. When applying an additional 10% MU to the absorbance measurement, the estimated range further adjusts to 13–22 IU/dL, still maintaining a much narrower uncertainty interval than the linear

approach. These findings indicate that using a second-order polynomial function significantly enhances calibration reliability, particularly at low concentrations where precise quantification is essential for diagnostic decision-making.

While the quadratic model improves fit quality, residual errors persist at higher concentrations, suggesting that additional curvature might further refine calibration accuracy. To test this, a cubic polynomial regression was applied to assess whether an additional degree of freedom would enhance interpolation precision.

8.3.2.5 Third order polynomial calibration curve fit

The cubic polynomial model further improves the calibration fit, as evidenced by the reduction in confidence interval width around the fitted line. However, this improvement comes with an unintended consequence: the introduction of a new pattern in the upper range of the calibration curve (Red square, Figure 58). This effect arises from the mathematical properties of cubic polynomials, where the additional curvature alters the shape of the fit beyond the measurement range, leading to an asymptotic shift that structurally influences the upper end of the curve. While this does not impact clinical decision-making at the critical threshold of 10 IU/dL, it introduces increased uncertainty at higher concentrations, which is undesirable for overall assay reliability.

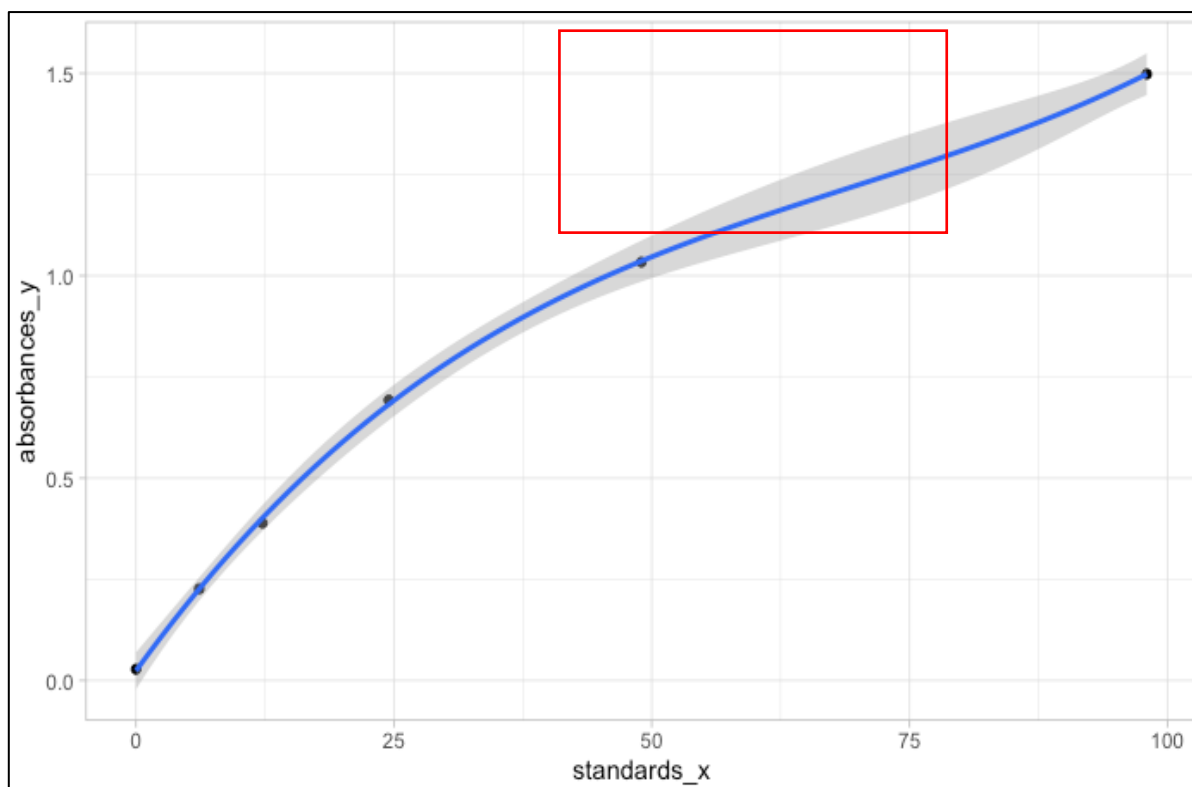


Figure 58: Cubic polynomial calibration - Technozym ® ADAMTS13

Increasing the non-linear fit to a third order polynomial (cubic) representation of the calibration curve

Despite this, the cubic polynomial model reduces uncertainty at the lower end of the calibration range, where clinical decisions are most relevant. The confidence intervals around the clinical decision limit remain narrow, leading to a more stable interpolation with lower uncertainty. For example, an absorbance of 0.5 yields an interpolated concentration range of 21–22 IU/dL, a significant reduction in uncertainty compared to the wider prediction intervals observed with lower-order models. Furthermore, incorporating a 10% MU in the absorbance value does not meaningfully alter this range, confirming the stability of the cubic model fit in the clinically relevant region.

8.4 Calibration Model Choice as a Structural Contributor to Run-to-Run Assay Variability

Building on this assessment of single-curve behaviour, this section evaluates whether calibration model choice contributes to variability across multiple assay runs. By comparing calibration performance across repeated runs using different models, the analysis investigates whether model selection acts as a structural contributor to between-run variability, potentially influencing measurement uncertainty through mechanisms currently unaccounted for.

8.4.1 Methods

8.4.1.1 Calibration Curve Modelling and Back-Calculation

Ten independent ELISA runs were analysed, each incorporating a single calibration event using Technozym® ADAMTS13 activity assays. A single lot number of kits was used to reduced the impact of lot-to-lot variability. For each calibration, raw optical density data from two replicates of each calibrator and control material were processed using MARS data analysis software (BMG Labtech, UK). Manufacturer calibrator materials at assigned values of 98 IU/dL (S1), 47 IU/dL (S2), 27 IU/dL (S3), 13 IU/dL (S4), 7 IU/dL (S5), and 0 IU/dL (S6). In addition, two levels of internal quality control material were included: Control C1 at 69 IU/dL and Control C2 at 21 IU/dL. The zero calibrator (S6) was included during curve fitting but was excluded from back-calculation analysis due to the non-comparable and uninterpretable nature of results near the lower asymptote for certain models.

For each calibration run, six curve-fitting models were applied: point-to-point interpolation, linear regression, third-order polynomial, cubic spline, four-parameter logistic (4PL), and five-parameter logistic (5PL). Each model was applied separately within MARS software, generating a distinct calibration curve per model for every run. Back-calculated concentrations for each calibrator and control material were extracted directly from MARS software for each model. Replicate measurements contributed to the curve fit in all models.

8.4.1.2 Bias and Variability Analysis

Bias for each material was calculated using both absolute and relative methods. Absolute bias was defined as the difference between the back-calculated concentration and the assigned target concentration for each calibrator or control material. Relative bias was calculated as the absolute bias divided by the target concentration, expressed as a percentage. Both bias calculations were performed using Microsoft Excel (Microsoft Corporation, USA).

For each calibration model, bias values were collated across the ten independent calibration runs. Average relative bias (% bias) was calculated for each model and material to assess systematic calibration error. To evaluate calibration-induced variability, the standard deviation of absolute bias (expressed in IU/dL) was calculated for each material and model across the ten runs. This approach provided a measure of run-to-run variability in calibration performance in true measurement space, aligning with the requirements for evaluating contributors to measurement uncertainty.

8.4.1.3 Data Analysis and Interpretation

The analysis was descriptive in nature, focusing on the comparison of bias and variability between calibration models. No statistical testing was performed. Results were stratified by model type, with bias and variability reported separately for each model across all calibrator and control levels. Complete datasets were analysed without exclusions.

8.4.2 Results

8.4.2.1 Average Bias

Average calibration bias across standards and controls demonstrated significant differences between calibration models. Both four-parameter logistic (4PL) and five-parameter logistic (5PL) models consistently produced the smallest average biases across all materials (Figure 59). The overall mean bias across all calibrators and controls was -0.32% for the 4PL model and -0.41% for the 5PL model, indicating minimal systematic deviation from target concentrations. Cubic spline and third-order

polynomial fits produced more variable performance. Cubic spline models showed an overall negative bias (-0.69%), largely attributable to pronounced underestimation at higher concentrations (notably Standard S1, with a mean bias of -12.05%). Similarly, the third-order polynomial model displayed a slight overall negative bias (-0.26%), with substantial underestimation evident at Standard S1 and Control C1.

Point-to-point interpolation demonstrated relatively stable performance, with a slight overall positive bias ($+1.39\%$). However, performance across individual standards varied, with inconsistent behaviour noted particularly for Standards S1 and S5.

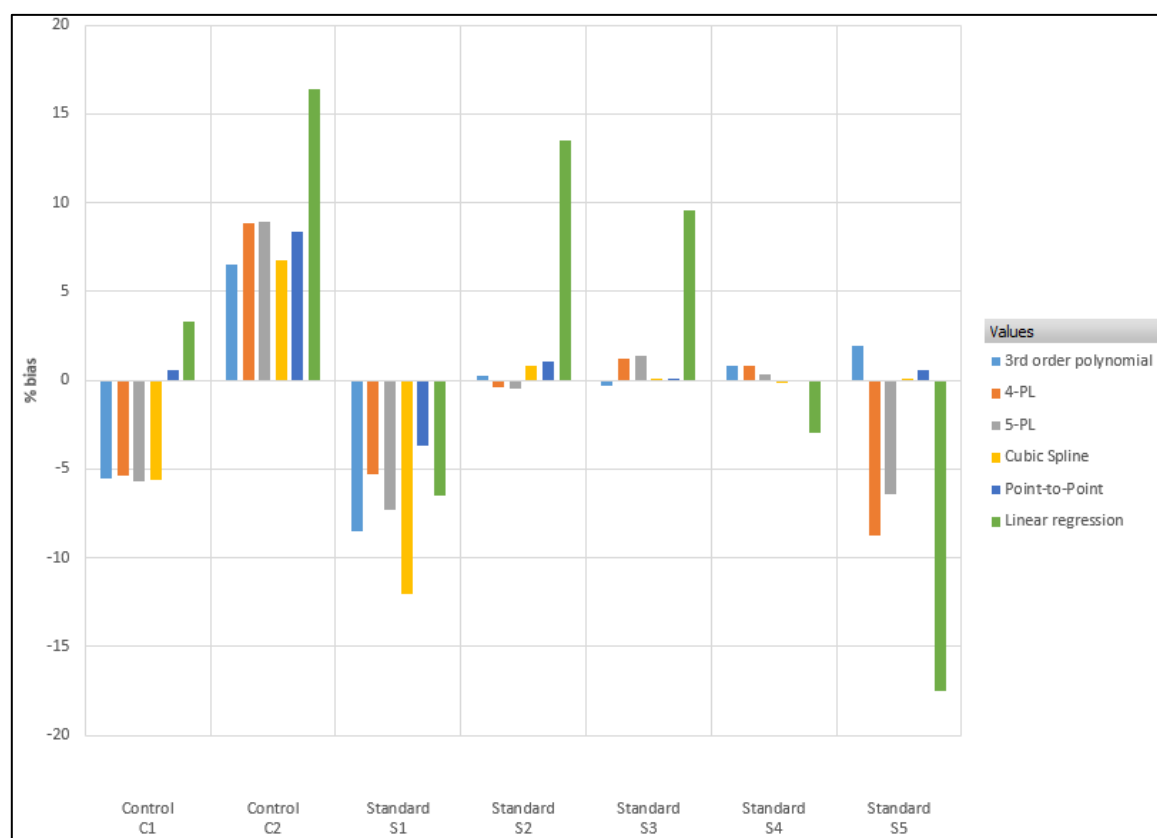


Figure 59: Relative Bias across 6 calibration modelling methods.

10 consecutive runs of ADAMTS13 activity assay by Technozym® ELISA. 4-PL – 4 parameter logistic, 5-PL – 5 parameter logistic.

Linear regression performed consistently worse than all other models. The method showed significant positive bias across controls and most standards, with an overall mean bias of +4.63 %. Overestimation was most marked in the low control (Control C2), where the average bias reached +16.41 %, indicating systematic overestimation of low-concentration results. A similar trend was observed across other calibrator levels, with severe underestimation at Standard S5 (–17.53 %), reflecting poor curve fit near the lower asymptote.

In summary, 4PL and 5PL models produced the most accurate calibration results, minimising systematic error across standards and controls. Alternative modelling approaches, particularly linear regression, demonstrated substantial model-dependent biases, reinforcing the importance of appropriate curve-fitting in reducing calibration-induced variability in one-at-a-time assays.

8.4.2.2 Variability (Standard Deviation of Absolute Bias)

The evaluation of calibration model stability, expressed as the standard deviation of absolute bias across multiple runs, demonstrated clear differences between models (Figure 60). Both four-parameter logistic (4PL) and five-parameter logistic (5PL) models consistently exhibited the lowest variability across all standards and controls, with overall standard deviations of 4.97 IU/dL and 5.46 IU/dL respectively. This indicates that 4PL and 5PL models deliver the most consistent calibration performance, with minimal run-to-run fluctuation in calibration error. Cubic spline modelling produced the highest overall variability at 7.11 IU/dL, reflecting unstable interpolation, particularly at the boundaries of the calibration curve. Linear regression exhibited moderately high variability (5.84 IU/dL), although this likely reflects consistently poor fit across the entire range, leading to repeatable miscalibration rather than genuine stability. Point-to-point interpolation performed comparably to 4PL and 5PL overall, with a standard deviation of 4.82 IU/dL.

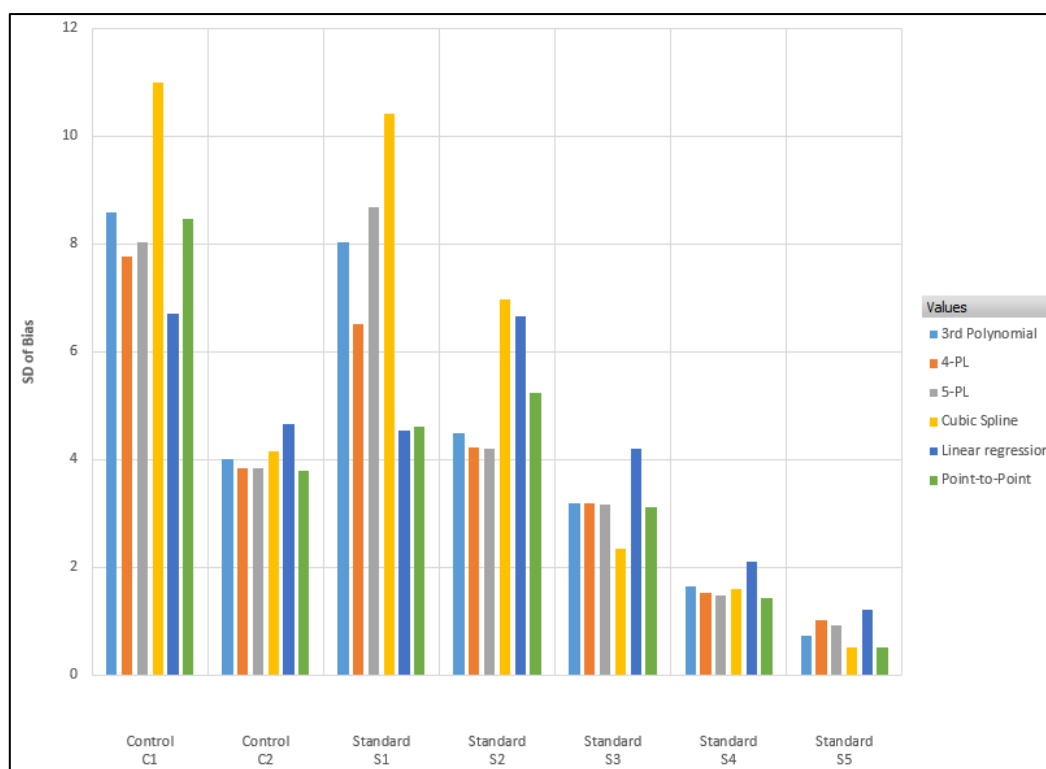


Figure 60: Variability of the absolute bias across the measurement range.

Each model tested across 10 runs produced variability (measured as the SD of the bias (in IU/dL) in each back calculated value) to differing extents. 4-PL – 4 parameter logistic, 5-PL – 5 parameter logistic

When assessed at the control levels, variability was notably higher for Control C1, where cubic spline modelling produced the greatest fluctuation (10.99 IU/dL). In contrast, at the low-concentration Control C2, all models exhibited reduced variability, with 4PL and 5PL models again providing the most consistent calibration, each maintaining variability below 4 IU/dL. Across the calibrator levels, cubic spline models demonstrated marked instability at both Standard S1 and Standard S2, while 4PL and 5PL maintained consistently lower variability across all standards. Linear regression produced deceptively low variability in certain standards, likely attributable to systematic model misfit rather than true calibration consistency.

Overall, the analysis demonstrates that calibration model selection directly affects calibration stability across runs, with 4PL and 5PL models providing both the lowest bias and the most consistent calibration performance. These findings confirm that calibration model choice functions as a

structural source of variability in one-at-a-time assays, contributing to result-level imprecision through mechanisms not currently accounted for in top-down uncertainty estimation frameworks.

8.5 Quantifying Calibration-Derived Measurement Uncertainty

Following the demonstration of model-dependent variability across calibrations, this section focuses on quantifying the contribution of calibration-derived uncertainty using Monte Carlo simulation.

Using the 4PL model as the standard approach, pooled calibration data are modelled to evaluate interpolation uncertainty across the assay range, supporting assessment against analytical performance specifications.

8.5.1 Methods

8.5.1.1 Data Collection and Aggregation

Twelve consecutive ADAMTS13 ELISA assay runs were included, conducted between 9th September and 29th November 2019, all using the same lot number of kits and calibrators. From the 15 runs originally performed during this period, three (Runs 3, 9, and 12) were excluded due to quality control failures.

Data from the remaining 12 runs were aggregated using the automated features within Statlia Matrix 5.0. The software generated response-versus-concentration plots with 95% confidence intervals at each calibrator level. It also calculated intra- and inter-assay coefficients of variation (%CV) and provided variance estimates for each calibration point, facilitating a robust assessment of uncertainty across the calibration range.

Given the non-linear nature of ELISA calibration, four-parameter logistic (4PL) regression models were applied to fit the calibration curves. The zero calibrator (0 IU/dL) was excluded from curve fitting due to its disproportionate influence on the regression model, a standard approach in ELISA calibration modelling (74). The five-parameter logistic (5PL) model was not employed, as the standard curves did not display the asymmetry required to justify its use.

8.5.1.2 *Monte Carlo-Based Estimation of MU*

Step 1: Aggregating Calibration Curves Across Multiple Runs

Each assay run produced a standard curve based on a 4PL model. Natural variability between runs—due to instrumental, reagent, and operator factors—resulted in minor fluctuations in the curve parameters (asymptotes, slope, and inflection point). Rather than evaluating each curve in isolation, Statlia Matrix aggregates these calibration models and applies Monte Carlo simulation to capture the expected variability in assay performance over time. This approach incorporates both intra- and inter-assay variation into the final uncertainty estimation.

Step 2: Simulating Curve Variability with Monte Carlo Methods

Monte Carlo simulation was used to model the range of possible calibration curve outcomes based on observed data. The procedure involved:

- Simulating calibration datasets that account for observed variability and the manufacturer-reported uncertainty (u_{cal}) for each calibrator.
- Fitting each simulated dataset with a 4PL model.
- Estimating distributions of key curve parameters across simulations.
- Interpolating the response of unknown specimens against these simulated curves to assess the range of plausible concentration estimates.

Ten thousand iterations were performed, providing a model-free, non-parametric method of estimating MU, particularly well-suited to the non-linearity of 4 parameter logistic regression.

Step 3: Interpolating Unknown Specimen Concentrations

Concentration estimates for unknown samples were derived by interpolating their absorbance values against the Monte Carlo-derived aggregate calibration model. The uncertainty of each estimate was influenced by several factors:

- Response variance, reflecting variability in signal at each concentration.
- Local slope of the curve, determining the sensitivity of interpolation.
- Spacing of standard calibrators, affecting how tightly the model constrains interpolation.
- Number of replicates, with two replicates used per unknown specimen in this study.

These factors collectively determined the width of the confidence interval for each interpolated value, which was ultimately used to calculate the percentage concentration error.

Step 4: Calculating Percentage Concentration Error

The percentage concentration error (%Conc Error) was calculated using the following formula:

$$\% \text{Conc Error} = \frac{(\text{Upper Bound} - \text{Lower Bound}) * 100}{(2 \times \text{Estimated Concentration})}$$

Equation 22: Determination of % concentration error.

Values derived from limits of interpolated values resulting from simulated, aggregated calibration curves

Alternatively, this can be expressed as an absolute uncertainty by multiplying the % concentration error by the interpolated value

8.5.1.3 Weighted Residuals and Calibration Curve Bias Analysis

To assess the adequacy of the 4PL calibration model, weighted residuals were calculated at each calibrator level. Squared residual plots were used to visualise discrepancies between observed and predicted values. Bias estimation was also performed to quantify systematic deviation from expected values. Residual Sum of Squares Error (RSSE) analysis was used to detect and localise systematic bias across the curve, enabling the identification of calibration points where model fit was suboptimal.

8.5.1.4 Error Profiling and Determination of Quantifiable Range

A further objective was to establish the assay's overall error profile and define its quantifiable range. Statlia Matrix's precision error profiling tool was used to compute the limits of quantitation (LOQ) and limits of detection (LOD) based on error thresholds pre-specified in this thesis. Error was modelled across the full calibration range to identify the lowest reliably quantifiable analyte concentration. From this, the lower and upper limits of quantitation (LLOQ and ULOQ) were defined. These values were aligned with regulatory guidance to establish the assay's reportable range, ensuring that patient results fall within clinically acceptable precision limits.

8.5.2 Results

8.5.2.1 Raw absorbance results of calibration curves from 12 consecutive assays

Higher concentrations exhibited greater variability between assays, with absorbance values at 94 IU/dL ranging from approximately 0.713 to 1.259, while responses at 51 IU/dL fluctuated between 0.525 and 0.894 (Table 40). The greatest variation at 94 IU/dL was likely caused by non-linearity due to substrate saturation at the upper end of the calibration curve, itself being the reason that a non-linear calibration curve fit was required as opposed to a linear one, as shown previously in section 8.3. In contrast, lower concentrations, such as 7 IU/dL and 14 IU/dL, demonstrated more consistent measurements, although some variation remains. At 7 IU/dL, absorbance values range from 0.116 to 0.164, whereas at 14 IU/dL, values fall between 0.185 and 0.286.

Table 40: Raw data (absorbance at 450nm) for calibrator values in IU/dL.

15 assay runs are presented, with each calibrator concentration run in replicate as indicated

Assay	Replicate	Absorbance at 450nm				
		94 IU/dL	51 IU/dL	29 IU/dL	14 IU/dL	7 IU/dL
1	1	1.118	0.712	0.449	0.259	0.157
1	2	1.099	0.835	0.448	0.269	0.155
2	1	1.052	0.673	0.404	0.234	0.139
2	2	1.057	0.675	0.394	0.239	0.14
4	1	1.156	0.764	0.45	0.259	0.144
4	2	1.109	0.763	0.474	0.276	0.151
5	1	0.978	0.658	0.381	0.185	0.116
5	2	1.1	0.676	0.411	0.203	0.127
6	1	1.186	0.869	0.583	0.384	0.235
6	2	1.259	0.894	0.57	0.434	0.225
7	1	0.932	0.687	0.453	0.287	0.154
7	2	0.909	0.676	0.393	0.247	0.162
8	1	1.105	0.782	0.467	0.268	0.152
8	2	1.08	0.742	0.433	0.257	0.149
10	1	1.194	0.789	0.475	0.261	0.161
10	2	1.114	0.795	0.482	0.286	0.164
11	1	0.765	0.577	0.393	0.252	0.129
11	2	0.754	0.585	0.398	0.262	0.142
13	1	0.849	0.577	0.381	0.213	0.125
13	2	0.935	0.573	0.379	0.228	0.142
14	1	0.89	0.613	0.366	0.217	0.143
14	2	0.981	0.642	0.402	0.219	0.158
15	1	0.713	0.534	0.317	0.197	0.118
15	2	0.774	0.525	0.335	0.199	0.122

Replicate measurements within each assay run generally showed good agreement, though some deviations existed most likely due to pipetting inconsistencies although all were within acceptable criteria between replicates (CV <20%). This represents a local acceptance criteria following those recommended by the manufacturer in the instructions for use). Certain assay runs, particularly assays 11, 13, and 15, display systematically lower overall absorbance values compared to others, highlighting inter assay differences.

Table 41 provides descriptive statistics for absorbance values obtained from 12 independent ADAMTS13 ELISA assay runs, across six calibrator levels: Bmin (blank), and standards at 94, 51, 29, 14, and 7 units. The table includes minimum, maximum, median, and mean values for each calibrator, along with measures of dispersion (standard deviation, CV), inferential statistics (two-tailed standard error, significance), and 95% confidence intervals (low and high confidence limits).

Table 41: Absorbance values across six calibrator levels over 12 independent ELISA runs.

The table summarises the minimum, maximum, median, mean, standard deviation (SD), coefficient of variation (CV%), total standard error (T-SE), and 95% confidence intervals (Low CL, High CL) for each calibrator level, including the blank (Bmin). The significance column indicates that all calibrator levels showed statistically significant variation across runs ($p < 0.01$).

Result	Bmin [1]	94	51	29	14	7
Assays	12	12	12	12	12	12
Min	0.016	0.744	0.530	0.326	0.194	0.120
Max	0.062	1.223	0.882	0.577	0.409	0.230
Median	0.03175	1.047	0.678	0.411	0.260	0.149
Mean	0.038083333	1.005	0.692	0.427	0.256	0.150
Std Dev	0.015115966	0.154	0.105	0.063	0.056	0.029
CV%	39.69181411	15.338	15.099	14.865	21.866	18.980
T-SE	0.015733196	0.160	0.109	0.066	0.058	0.030
Significance	0.01	0.01	0.01	0.01	0.01	0.01
Low CL	-0.01078093	0.506	0.354	0.222	0.075	0.058
High CL	0.086947596	1.503	1.030	0.632	0.436	0.243

The mean absorbance values follow the expected calibration trend, with the highest mean (1.005) at calibrator 94 and the lowest (0.038) at the blank (Bmin), confirming the direct relationship between analyte concentration and signal intensity. The median values closely align with the means, suggesting symmetrical distributions.

Signal variability increases as absorbance decreases. The CV% is lowest at mid-to-high signal levels (approximately 15% for calibrators 94, 51, and 29), increases at lower concentrations (14: 21.87%, 7: 18.98%), and is highest at the blank (Bmin: 39.69%), reflecting substantial instability in background signal.

The SD and total standard error (T-SE) follow a similar pattern, with larger values at higher concentrations but proportionally smaller relative to the mean, resulting in lower CVs. This supports the conclusion that absolute variability is highest at the top end, while relative variability is greatest at the bottom of the calibration range.

All calibrator levels show a statistically significant difference across assays (Significance = 0.01), and the confidence intervals for each mean confirm this variation. The wide confidence intervals at both ends of the calibration range (e.g., Bmin: -0.011 to 0.087; 94: 0.506 to 1.503) reflect greater uncertainty in the signal estimates at those levels.

8.5.2.2 Reproducibility of curve fitting using the 4-parameter logistic model

Aggregation of the twelve selected analytical runs allowed characterisation of the modelling process across assays as shown below (Figure 61).

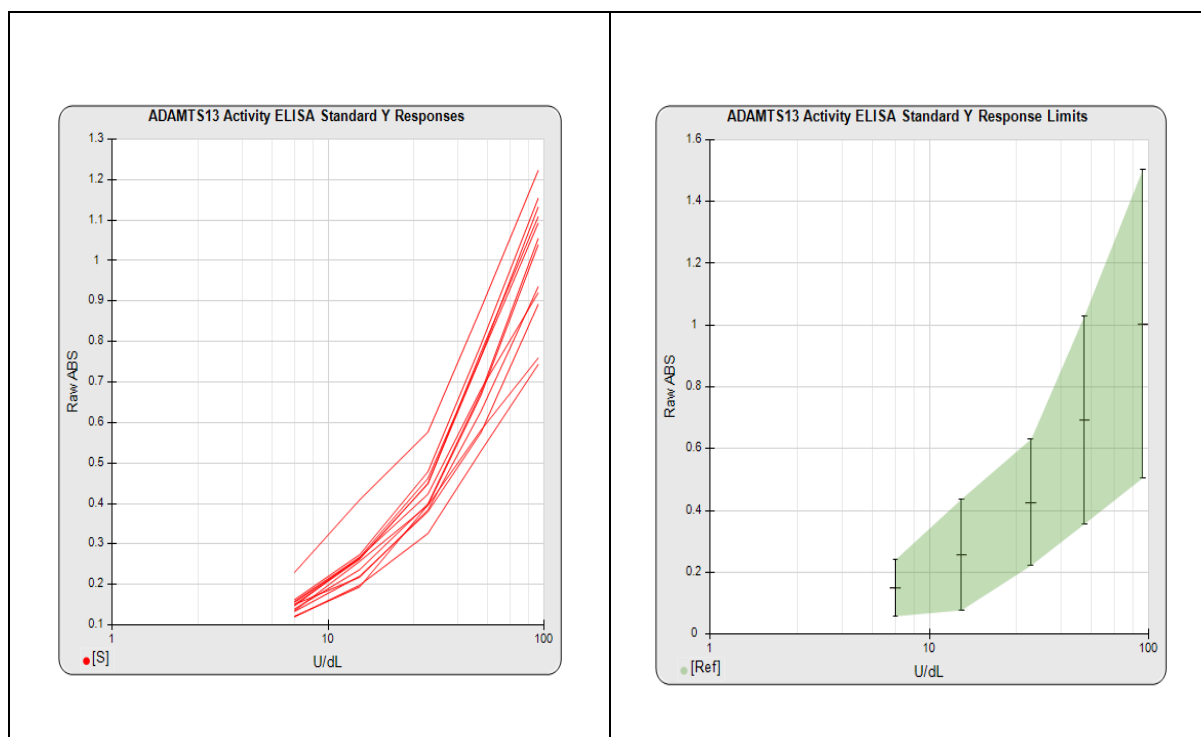


Figure 61: Aggregated Technozym® ELISA runs (n=12).

(Left) The Y responses of each standard dilution of an assay are linked together for each of the pooled reference assays (n=12). (Right) - The mean and confidence limits of the Y responses of each standard dilution are plotted and the region within and between the limits is shaded in green. The confidence limits are determined from the assigned significance level (95%). Graphs generated using Statlia Matrix Software (version 5.0).

Curve shapes were consistent (Figure 61). One assay run produced noticeably higher absolute absorbances however the overall shape of the curve was consistent with the other eleven assays. 95% confidence limits of the expected absorbances are shown in the green shaded area. Absorbances at lower concentrations of calibrator had higher repeatability in this dataset and the width of the confidence interval increased proportionally with calibrator concentration.

8.5.2.3 Trending of absorbance values for calibration plasmas over time

Figure 62 illustrates the mean raw absorbance (ABS) values for each calibrator level across the 15 ADAMTS13 activity ELISA runs. Each line represents a different calibrator concentration, with higher concentrations producing higher absorbance signals. The data demonstrates the expected direct relationship between analyte concentration and absorbance, consistent with a non-competitive (sandwich) ELISA format.

The highest calibrator levels (red series) show mean absorbance values consistently above 1.0 across most assay runs, while mid-range calibrators cluster between 0.2 and 0.9. The lowest calibrator levels, including the blank ([Bn], shown in yellow), remain near the baseline with minimal variation (<0.1 ABS), indicating stable background signal and low non-specific binding.

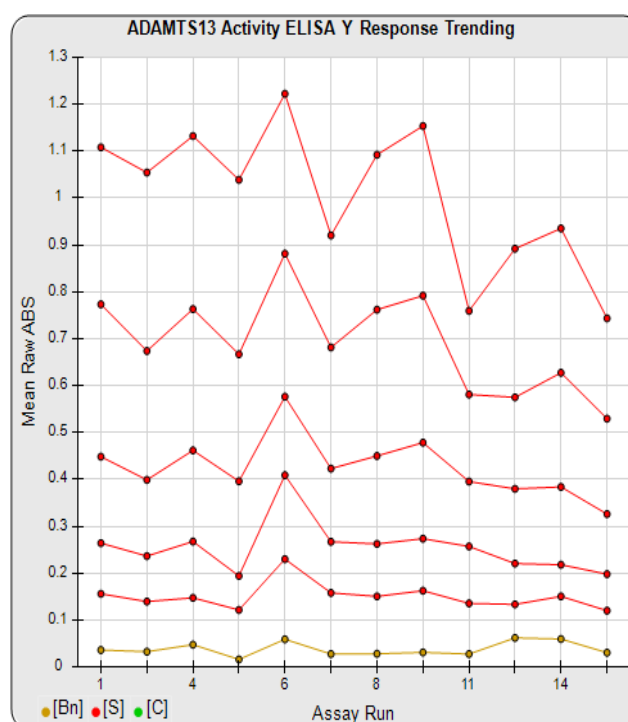


Figure 62: Absorbances across aggregated Technozym[®] ELISA runs

The Y response of each baseline, standard, and control dilution are plotted for each reference assay. The data are taken from the dilution responses in the Dilution Y Responses table. Figure generated using Stalia matrix software version 5.0

Although minor inter-assay variation is present, particularly in higher calibrator levels (e.g., Run 6 showing a peak), overall signal trending remains stable despite an apparent drop in absorbance in latter runs, supporting the reproducibility of the assay and its suitability for further pooled analysis or 4PL model fitting.

8.5.2.4 Model fit and bias introduced by the 4-parameter logistic curve fitting process

Bias was determined across all calibrator concentrations as the residuals between expected and observed recovery from being back calculated against the aggregated curve model. This is represented below comparing both absorbance dependent and concentration dependent biases.

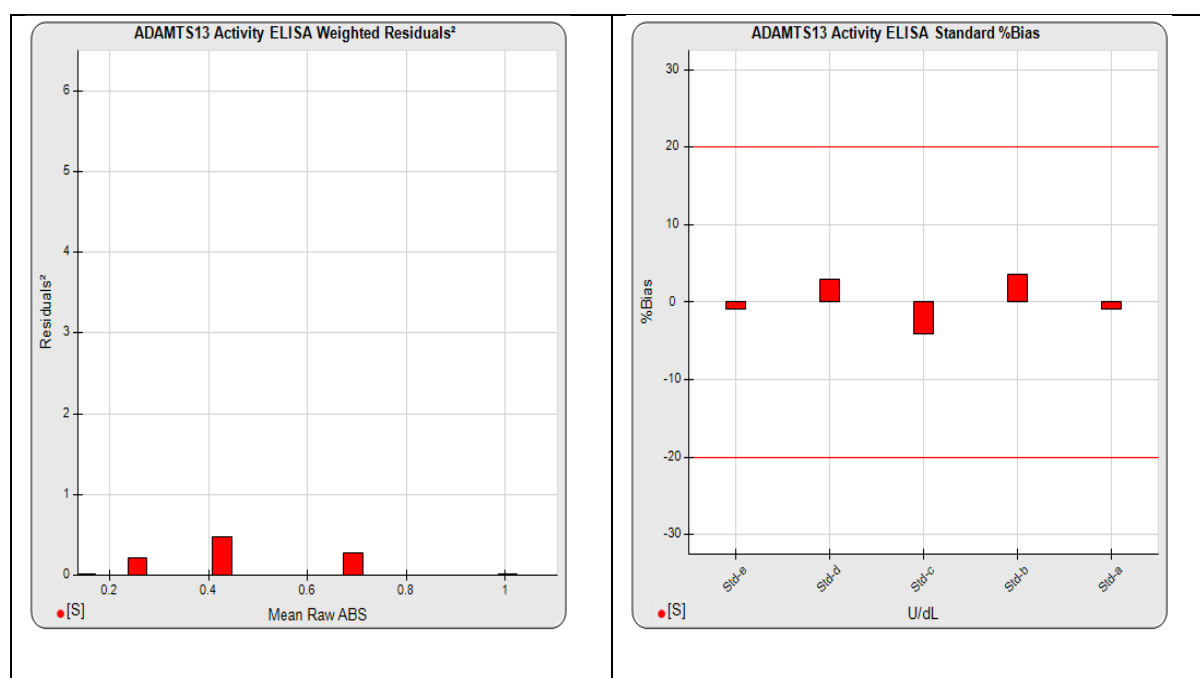


Figure 63: Residual and %Bias from aggregated Technozym[®] runs

(Left) The mean residual² of each standard dilution are taken from the Mean Standard Dilution Residuals² of the pooled reference assays. The individual residuals² are the squared differences between the observed and predicted response of each dilution, divided by its expected variance (weight). (Right) - The mean %Bias of each standard dilution. Large %Bias values can indicate an inadequate curve model fit or that the reaction kinetics are no longer effective at that dose.

Residuals from the 12 curve fitting procedures were uniformly distributed and acceptably low to be able to conclude there was no systematic bias induced by the curve fitting process across the measurement range. This is in contrast to the fitting procedures performed using linear curve fitting procedures where the bow tie shape was seen in a simple linear regression (

Figure 55:), and even increasing the order of polynomial fitted up to a 3rd degree polynomial higher results were impacted differently from lower results.

8.5.2.5 Between and within assay performance using ANOVA

The total sum of squares (SS) and its components confirm that between-run effects account for the majority of the overall variability at every level (Table 42). For example, at calibrator 94, 95% of the total variance (SS = 0.547) arises from between-run variation (SS = 0.522), with only a small portion attributed to within-run variation (SS = 0.025). This pattern is consistent across all calibrators.

Table 42: ANOVA summary statistics for aggregated Technozym[®] ELISA.

Y response (absorbance) across 12 assay runs at each calibrator level. The table shows the distribution of variance components for each calibrator concentration and the blank (Bmin), including total, between-run, and within-run degrees of freedom (DF), sum of squares (SS), mean square (MSS), and variance estimates. Mean absorbance, within-run and between-run standard deviations (SD), and coefficients of variation (%CV) are provided to assess assay precision and reproducibility.

Result	Bmin [1]	94	51	29	14	7
Assay N	12	12	12	12	12	12
Sample N	23	24	24	24	24	24
DF Total	22	23	23	23	23	23
DF Within	11	12	12	12	12	12
DF Between	11	11	11	11	11	11
SS Total	0.00700	0.54720	0.24982	0.09256	0.07169	0.01846
SS Within	0.00249	0.02494	0.00942	0.00410	0.00296	0.00053
SS Between	0.00452	0.52226	0.24040	0.08846	0.06873	0.01793
MSS Total	0.00032	0.02379	0.01086	0.00402	0.00312	0.00080
MSS Within	0.00023	0.00208	0.00079	0.00034	0.00025	0.00004
MSS Btw	0.00041	0.04748	0.02185	0.00804	0.00625	0.00163
Mean	0.03904	1.00454	0.69233	0.42658	0.25563	0.15042
Var Within	0.00023	0.00208	0.00079	0.00034	0.00025	0.00004
Var Between	0.00021	0.02374	0.01093	0.00402	0.00312	0.00082
SD Within	0.01504	0.04559	0.02802	0.01848	0.01570	0.00663
SD Between	0.01464	0.15407	0.10453	0.06341	0.05589	0.02855
%CV Within	38.51173	4.53831	4.04709	4.33309	6.14245	4.40992
%CV Btw	37.49323	15.33777	15.09869	14.86459	21.86595	18.98039

The calculated within-run standard deviations (SD Within) are low across calibrators (ranging from 0.007 to 0.046), indicating strong repeatability. However, between-run standard deviations (SD Between) are notably higher, particularly at higher concentrations—for example, 0.154 at calibrator 94—indicating greater variability across assay runs. This is further reflected in the coefficients of variation: within-run %CVs remain consistently low (around 4–6%) for all calibrators except the blank, which has a %CV of 38.5%, indicating poor precision near the assay’s detection limit. In contrast, between-run %CVs range from ~15% at calibrators 51 and 29 to over 21% at calibrator 14, and reach nearly 19% at calibrator 7, highlighting the impact of run-to-run variability, particularly at lower signal levels.

8.5.2.6 Coefficient of Variation between and within runs

The CV% of mean absorbance responses against absorbance for each calibrator level in the ADAMTS13 activity ELISA are shown in Figure 64. Each data point represents the variability at a given absorbance level (corresponding to the mean of all runs for each calibrator concentration), with filled squares indicating inter-assay %CVs (variation between assay runs), open squares indicating intra-assay %CVs (variation within a single run) and colours corresponding to calibrator levels (all red), including the blank ([Bn], yellow).

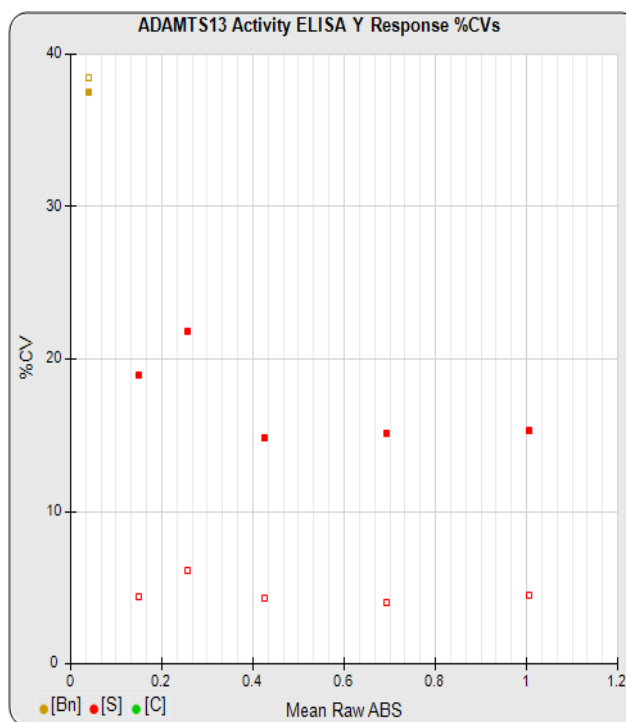


Figure 64: Aggregated reproducibility Technozym[®]

The intra-assay (open squares) and inter-assay (filled squares) %CVs are taken from the within assay and between assay results from the analysis of variation (ANOVA) computed for each individual baseline, standard, and control dilution's Y responses. These can be found with the pooled reference assays in the Y Response ANOVA Statistics table.

At low absorbance levels (below 0.4), both intra- and inter-assay %CVs are elevated, reflecting greater relative variability near the assay's lower detection limit. Inter-assay %CVs in this region reach approximately 20%, while intra-assay %CVs remain below 10%, indicating better within-run precision even at low signal levels.

As absorbance increases beyond 0.4, the inter-assay %CV shows a modest decline, stabilising around 15%. In contrast, the intra-assay %CV is relatively stable at 5%, and remains consistently lower than the inter-assay CV across all levels. This confirms tighter control and less variability within individual plates compared to across different assay runs.

These findings highlight that while the assay demonstrates good intra-assay precision across the full dynamic range, inter-assay reproducibility improves modestly with increasing signal but remains the dominant source of variability, particularly near the lower end of the response curve.

8.5.2.7 Error profiling of the curve fitting process using Monte Carlo Simulation

The plotted concentration errors exhibit a characteristic pattern, where uncertainty increases exponentially at both the lower and upper limits of the assay's measurement range (Figure 65, right). This pattern reflects the fundamental limitations of the calibration curve, with higher error rates at extreme concentrations due to increased signal-to-noise ratio at low levels and substrate saturation effects at high levels.

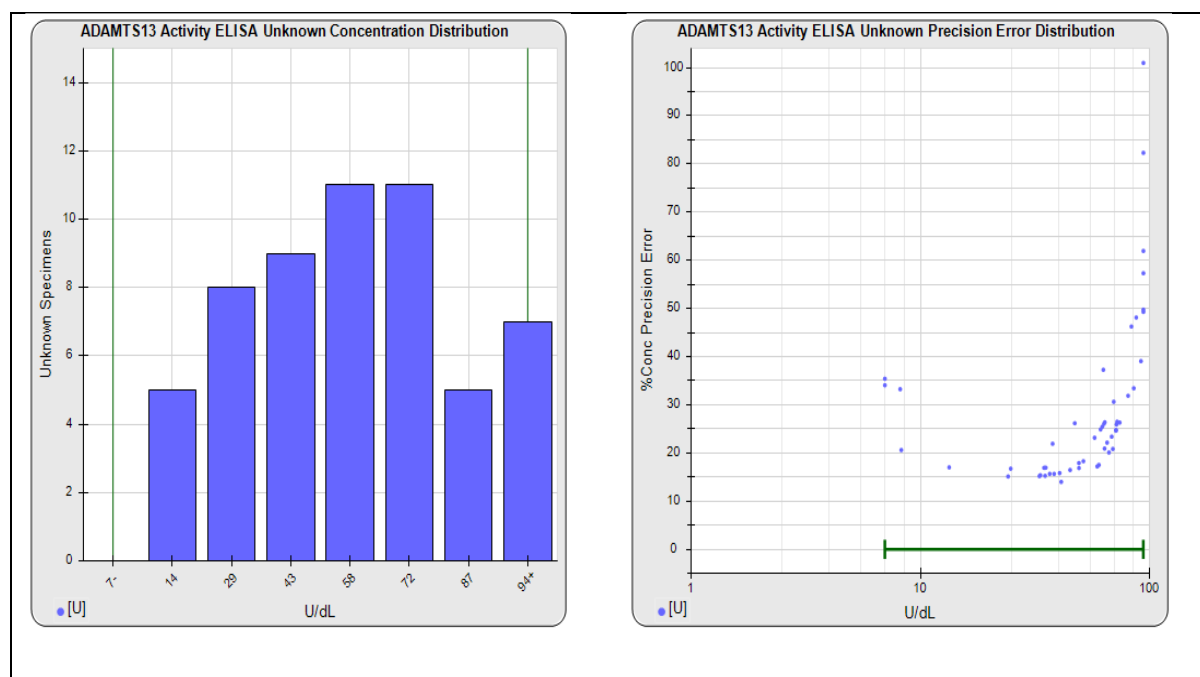


Figure 65: Technozym[®] aggregated unknown sample concentrations.

Left - The distribution of concentrations from each unknown specimen in the pooled reference assays are plotted in this histogram. The green lines are plotted at the lower and upper 0.05 confidence limits. Right - The percent concentration precision errors versus their concentrations are plotted for all of the reference assay unknown specimen results. Gray points are unknown results from critically failed standard curve regressions (Fit Prob < residual threshold). The reportable range specified in assay setup is plotted with the green bar. The data are taken from the Reference Assay Unknown Specimen Conc and Unknown Specimen %Conc Error tables.

By recognising this error profile, a defined reportable range can be established based on APS, ensuring that patient results remain within acceptable precision thresholds. This reportable range is indicated by the green lines above and below the x-axis, delineating the boundaries within which MU is reliably controlled (Figure 66).

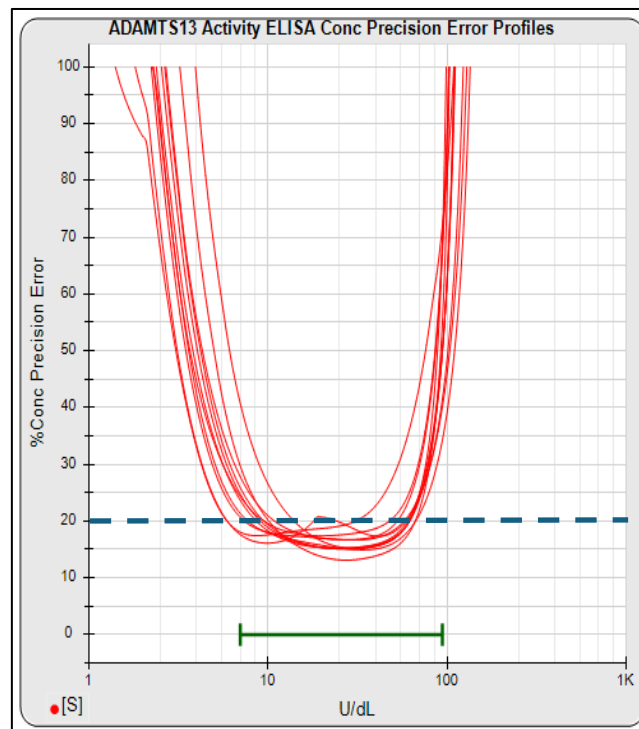


Figure 66: Performance specifications for aggregated Technozym® ELISA runs

The percent concentration precision error versus dilution concentration are plotted for all of the reference assay standard curves. The reportable range specified in assay setup is plotted with the green bar. Reportable ranges should be established that will ensure that all concentration estimates reported will not have a higher %Conc Precision Error than desired. The Limits of Quantitation of each standard curve give the lower and upper concentration bounds at the %Conc Error threshold. The data are taken from the Reference Assay Standard Curve Metrics table.

Error distribution trends further confirm that low-concentration specimens within individual assay runs exhibit greater MU, likely due to the lower signal-to-noise ratio at the detection limit. For instance, Uk-2 in Assay 2 (Table 43) showed a concentration error of 35.41%, reinforcing the trend of higher variability at lower concentrations. At the upper end of the calibration curve, measurement errors also increase, as seen in Uk-4 and Uk-7 in Assay 7, which exhibited concentration errors of 30.63% and 46.24%, respectively. This elevated uncertainty at higher concentrations may be attributed to curve saturation effects, where response values plateau and small fluctuations in signal result in disproportionately large variations in concentration estimates.

8.5.2.8 Determining limit of reporting based on historical data

Performance was assessed using Limits of Quantitation (LLOQ, ULOQ), Limits of Detection (LLOD, ULOD), and the Limit of Blank (LOB) from the aggregated curve fits. The mean LLOQ was 3.54 IU/dL, with a minimum of 2.43 and a maximum of 7 IU/dL, indicating variability in the assay's ability to measure low concentrations consistently, albeit comfortably below the clinical decision limit of 10 IU/dL. The ULOQ ranged from 93.16 to 123.05 IU/dL, suggesting differences in upper concentration range calibration across runs. The LLOD values varied between 0.85 and 2.86 IU/dL, while ULOD exhibited extreme variation, giving results far outreaching what would be physiologically possible, showing the importance of interpreting purely statistical metrics in the context of clinical application. These large differences are again apparently due to substrate depletion at higher concentrations.

8.5.2.9 Applying MU to unknown samples

The percentage concentration errors (from which the absolute errors displayed in Table 43 are derived) provide a measure of the variability associated with the quantification of individual unknown specimens. Uncertainty increases with concentration, particularly in ELISA-based assays where sigmoidal curve fitting can introduce greater variability at higher measurement ranges. Some samples exhibit particularly high errors within their respective runs, indicating greater uncertainty in their measurement. For instance, Uk-2 in Assay 14 showed a concentration error of 82.32%,

reflecting substantial imprecision, while Uk-6 in Assay 5 recorded an exceptionally high error of 145.13%, likely due to poor curve fitting at that specific concentration. Other specimens, such as Uk-3 in Assay 8 and Uk-4 in Assay 8, demonstrated lower and more stable errors of 16.47% and 15.67%, respectively, suggesting greater confidence in their quantification.

Table 43: Unknown samples results with measurement uncertainty.

The measured concentrations of unknown samples across different assay runs along with absolute uncertainty derived from the aggregated model fitting procedure. All measurements are reported in IU/dL.

Assay	Uk-1	Uk-2	Uk-3	Uk-4	Uk-5	Uk-6	Uk-7	Uk-8
1	71.3+/- 17.6	24.1+/-3.6	80.6+/- 25.7					
2	94+/-58.2	7+/-2.5	69.2+/- 14.4	94+/-53.9				
4	91.7+/- 35.8	49.2+/-8.8	85.4+/- 28.5	65.3+/- 14.5	71.4+/- 17.7	68.4+/-16	51.4+/-9.4	13.4+/-2.3
5	59.1+/- 10.2	60.1+/- 10.5	41.2+/-5.8	66.5+/- 13.4		94+/- 136.4		
6	62.9+/- 23.4	47.2+/- 12.3	37.7+/-8.3					
7	63.8+/- 16.8	63.2+/- 16.5	62.2+/- 15.8	69.8+/- 21.4	57.6+/- 13.3	61.1+/- 15.2	83.3+/- 38.5	7+/-2.4
8	8.3+/-1.7	71.7+/- 18.6	45+/-7.4	36.7+/-5.7				
10	38.3+/-6	63.6+/- 13.3	33.4+/-5.1	40.5+/-6.4		49.2+/-8.3		
11	53.5+/-8.8	94+/-13.7	25.2+/-3.3	67+/-10.3				
13	94+/-46.3	24.8+/-4.2	74.1+/- 19.5	94+/-46.8				

Assay	Uk-1	Uk-2	Uk-3	Uk-4	Uk-5	Uk-6	Uk-7	Uk-8
14	33.1+/-5	94+/-77.4	72.2+/- 19.1	87.5+/- 42.1	35+/-5.3	8.2+/-2.7		
15	34.7+/-12	35.2+/- 12.4						

Conversely, lower concentration samples generally have smaller absolute uncertainties. For instance, Uk-8 in Assay 4 had a measured concentration of 13.4 IU/dL with an uncertainty of ± 2.3 IU/dL, while Uk-8 in Assay 7 was recorded at 7 IU/dL with an uncertainty of ± 2.4 IU/dL. Although these values suggest greater measurement stability, even minor errors at lower concentrations may have clinical implications particularly at or around the 10 IU/dL point.

Certain assays exhibit a relatively consistent uncertainty distribution across all unknown samples. For instance, Assay 7 has measurement uncertainties ranging from ± 13.3 IU/dL to ± 38.5 IU/dL, indicating a stable error distribution across different concentration levels. In contrast, other assays, such as Assay 4, demonstrate a wider spread of uncertainties, with some samples exhibiting minimal errors while others, such as Uk-3, have a much larger uncertainty of ± 28.5 IU/dL, albeit again as expected at a higher concentration. This suggests that measurement precision is heavily influenced by where a sample falls within the calibration curve, with mid-range concentrations often displaying the highest uncertainty due to non-linearity effects. This variability is also linked to the individual model fit to the curve.

8.6 Calibration Standardisation and Hidden Uncertainty in Automated Systems

While manual assays rely on repeated calibration, automated systems use master calibration curves adjusted through local bias correction factors. This section explores whether these local adjustments introduce systematic uncertainty within automated platforms, assessing whether calibration standardisation eliminates or merely obscures calibration-related variability.

8.6.1 Methods

8.6.1.1 Data source and extraction

Five successive calibrations performed on the HemosIL® AcuStar ADAMTS13 activity assay were analysed between May and November 2024. This six-month period was selected as it aligns with the typical timeframe used in routine uncertainty evaluations. Each calibration represented a new reagent lot, with an associated lot-specific calibrator, both of which were provided by the manufacturer (Werfen, Bedford, USA). Each reagent lot had a preprogrammed master calibration curve, which was automatically uploaded to the analyser software upon reagent lot installation.

For each calibration, three adjustment calibrators were assayed as part of the standard calibration process on the analyser. These calibrator values were then used to modify the manufacturer-supplied master 4-parameter logistic (4PL) curve, generating a new 4PL curve that served as the local calibration curve for that specific reagent lot.

Upon completion of each calibration, the analyser provided statistical performance metrics, including fitted parameters and calibration data for both the master calibration curve and the locally adjusted curve. As the system does not support electronic data extraction, all calibration data were manually recorded. The data from all five calibrations were then aggregated and analysed to evaluate trends in calibration variability, bias correction, and uncertainty propagation.

8.6.1.2 Analysis Design

The CV of the fit results was used to assess the difference between the assigned values of the adjuster calibrators and the returned values after the adjusted calibration curve was applied.

For each of the five calibrations, summary statistics were derived for the four parameters of the 4-parameter logistic (4PL) local adjusted curves. The difference in relative light units (RLU)—defined as the deviation between the expected result (based on the master curve) and the actual analysed result—was evaluated both across all calibrations and individually to assess whether significant lot-to-lot variation was present.

To further explore the impact of calibration adjustments, average RLU differences were summarised, allowing for an evaluation of how calibrator values influence the transformation of the master calibration curve into the local adjusted curve. This analysis aimed to determine whether systematic trends or inconsistencies exist in the bias correction process.

8.6.2 Results

The calibration fit results across five successive calibrations of the HemosIL AcuStar® ADAMTS13 activity assay revealed distinct patterns of variability depending on the calibrator level. The zero calibrator exhibited no variability, with a mean adjuster fitted value of 0% and a CV of fit consistently at 0% across all calibrations (Table 44). This indicates that the zero calibrator adjustment did not influence baseline values.

Table 44: Calibration fit statistics across five independent calibrations.

The variability in calibration fit across five successive calibrations for the HemosIL AcuStar ADAMTS13 activity assay, comparing the Zero Calibrator (0%), Decision Limit Calibrator (10%), and Normal Level Calibrator (90%). The Mean Adjuster Fitted Value (%) reflects the shift applied to each calibrator during local adjustment, while the Coefficient of Variation (CV) of Fit (%) indicates variability.

	Zero Calibrator	Calibrator at decision limit	Normal level calibrator
n	5	5	5
Mean adjuster fitted value (%)	0	8.3	90.4
Mean CV of fit (%)	0	2.6	3.3
Minimum CV of fit (%)	0	0.3	1.7
Maximum CV of fit (%)	0	8.5	4.2

In contrast, the calibrator at the decision limit demonstrated the most variability, with a mean adjuster fitted value of 8.3%, indicating a slight positive shift in calibration adjustment (Table 44). The mean CV of fit was 2.6%, but variability across calibrations ranged widely from 0.3% to 8.5%, highlighting inconsistencies in how the decision limit calibrator was adjusted. Given the clinical importance of this threshold, where classification of patient results is critical, the observed fluctuations suggest that the adjustment process at this level may introduce additional uncertainty that could impact clinical interpretation.

The normal level calibrator showed a more stable adjustment pattern, with a mean adjuster fitted value of 90.4% and a mean CV of fit of 3.3%. While some degree of variability was present, ranging from 1.7% to 4.2%, the variation was less pronounced than that seen at the decision limit. This suggests that while calibration adjustments introduce some degree of uncertainty, the stability of the normal calibrator fit is better controlled compared to the decision limit calibrator.

Overall, these findings indicate that while the master calibration adjustment process remains stable at baseline and relatively consistent at normal levels, it introduces considerable variability at the decision limit, where accuracy is most critical.

8.6.2.1 Calibration parameters of five successive calibrations

The dataset summarises the fitted four-parameter logistic calibration curve after local adjustment, recorded across five successive calibrations between May and November 2024. The parameters analysed include Log(ED50), Top, Hill, and Bottom, with summary statistics evaluating the consistency and variability of these fitted parameters over time.

8.6.2.1.1 Log(ED50) (Midpoint of the Calibration Curve)

The Log(ED50), representing the inflection point of the calibration curve where the response is 50% of the maximum, had a mean value of 3.997 with an SD of 0.356 and a CV of 8.90% (Table 45, Table 46). While the median Log(ED50) was 3.919, the values ranged from 3.598 to 4.425, indicating notable variability across calibrations. The observed differences suggest that the midpoint of the dose-response curve fluctuates, which could influence the assay's sensitivity and response characteristics.

8.6.2.1.2 Top Parameter (Maximum Response)

The Top parameter, representing the maximum signal response, showed significant variation, with a mean RLU of 28,109,258, a high SD of 17,633,663 RLU (Table 45), and an extremely large CV of 62.73% (Table 46). The values ranged from 12,533,900 to 50,796,788, indicating substantial variability between calibration lots. This suggests that the overall signal intensity fluctuated dramatically, potentially due to differences in reagent lot performance, analyser conditions, or calibration stability. The extreme variability in this parameter could introduce uncertainty in high-concentration measurements, affecting the overall reliability of the assay at the upper end of the calibration curve.

Table 45: Calibration four-parameter logistic fitted parameters.

Key parameters of the fitted calibration curves across five successive calibrations for the HemosIL AcuStar ADAMTS13 activity assay. The parameters include Log(ED50) (midpoint of the curve), Top (maximum response), Hill (slope of the curve), and Bottom (minimum response).

Calibration Date	Calibrator lot number	Trigger lot	Rinse lot	Reagent lot	Log ED50	Top	Hill	Bottom
03/05/2024	6452	6150	3306	6452	3.7416	18129400	0.968083	731.019
06/08/2024	6547	6429	4052	6547	3.5982	15709100	0.987794	588.858
16/09/2024	6705	6429	4117	6705	3.9193	12533900	0.839378	592.344
22/10/2024	6839	6700	4117	6839	4.3014	43377100	0.872455	595.08
19/11/2024	6982	6700	4117	6982	4.4247	50796788	0.88628	662.6752

8.6.2.1.3 Hill Parameter (Slope of the Calibration Curve)

The Hill slope, which defines the steepness of the curve and influences how sharply the response transitions between lower and higher concentrations, had a mean value of 0.911, with a relatively low SD of 0.064 and a CV of 7.03% (Table 46). The range was from 0.839 to 0.988, indicating some variability but within a more controlled range compared to the Top parameter. The fluctuations in the Hill slope suggest minor inconsistencies in the steepness of the dose-response curve, which could slightly affect result interpolation, particularly for concentrations around the midpoint of the curve.

8.6.2.1.4 Bottom Parameter (Minimum Response)

The Bottom parameter, representing the lower asymptote of the calibration curve, had a mean value of 633.99, with a SD of 62.30 and a CV of 9.83% (Table 46). The minimum observed value was 588.86, and the maximum was 731.02, suggesting moderate variability at the lower end of the curve.

This level of fluctuation is not as extreme as the Top parameter, but still indicates potential variability in the assay's sensitivity at low concentrations.

Table 46: Aggregated calibration fit parameter across five independent runs.

Mean, standard deviation (SD), median, standard error (SE), coefficient of variation (CV), minimum, and maximum values for the fitted calibration parameters across five successive calibrations of the HemosIL AcuStar ADAMTS13 activity assay.

	mean	SD	Median	SE	CV	Min	Max
Log_ED50	3.99704	0.355631	3.9193	0.159043	8.897367	3.5982	4.4247
Top	28109258	17633663	18129400	7886014	62.73258	12533900	50796788
Hill	0.910798	0.063996	0.8862795	0.02862	7.026401	0.839378	0.987794
Bottom	633.9952	62.29515	595.08	27.85924	9.825807	588.858	731.019

Overall, these findings highlight notable inconsistencies in calibration curve parameters, particularly in the maximum response (Top) and the calibration midpoint (Log(ED50)). These variations could contribute to MU and potential bias in assay performance, suggesting that the adjustment process may not fully stabilise calibration consistency across reagent lots.

The analysis of key calibration curve parameters revealed several important trends. The log-transformed ED50 values, representing the curve's midpoint, showed notable variability, ranging from 3.598 to 4.425, which could influence assay sensitivity and the reliability of mid-range interpolations. The Top parameter, reflecting maximum signal intensity, exhibited extreme variability with a CV of 62.73%, suggesting significant fluctuations likely attributable to reagent lot differences or calibration inconsistencies. In contrast, the Hill parameter, which determines the steepness of the curve, remained relatively stable (CV = 7.03%), indicating consistent response behaviour across runs. The Bottom parameter, representing the lowest asymptote of the curve, showed moderate variability (CV = 9.83%), implying that lower-end signal responses were generally stable, though not without

some fluctuation. Collectively, these findings highlight the dynamic nature of calibration curve parameters and the importance of monitoring their stability when assessing assay performance.

8.6.2.2 Concentration dependence of %RLU change

The graph in Figure 67 demonstrates a clear concentration-dependent trend in the % RLU change between the master and adjusted calibration curves, with the largest adjustments occurring at low calibrator concentrations (Figure 67). The % RLU change is most pronounced at 1% and 10%, where significant downward adjustments were required, whereas at higher concentrations, the adjusted and master curves align more closely. This is clinically significant, as the highest variability occurs at the critical decision limits for TTP diagnosis, raising concerns about the stability of calibration adjustments in this range.

At the 10% decision threshold, the % RLU change ranged from -23% (Lot 6452) to +98% (Lot 6839), averaging -58.9%, while at 1% calibrator concentration, the change was even greater, averaging -113.1%, with a maximum deviation of -184.9% (Lot 6839). These results suggest that bias correction is most substantial where clinical accuracy is most crucial, while agreement between the master and adjusted curves improves at higher concentrations, where clinical significance is lower.

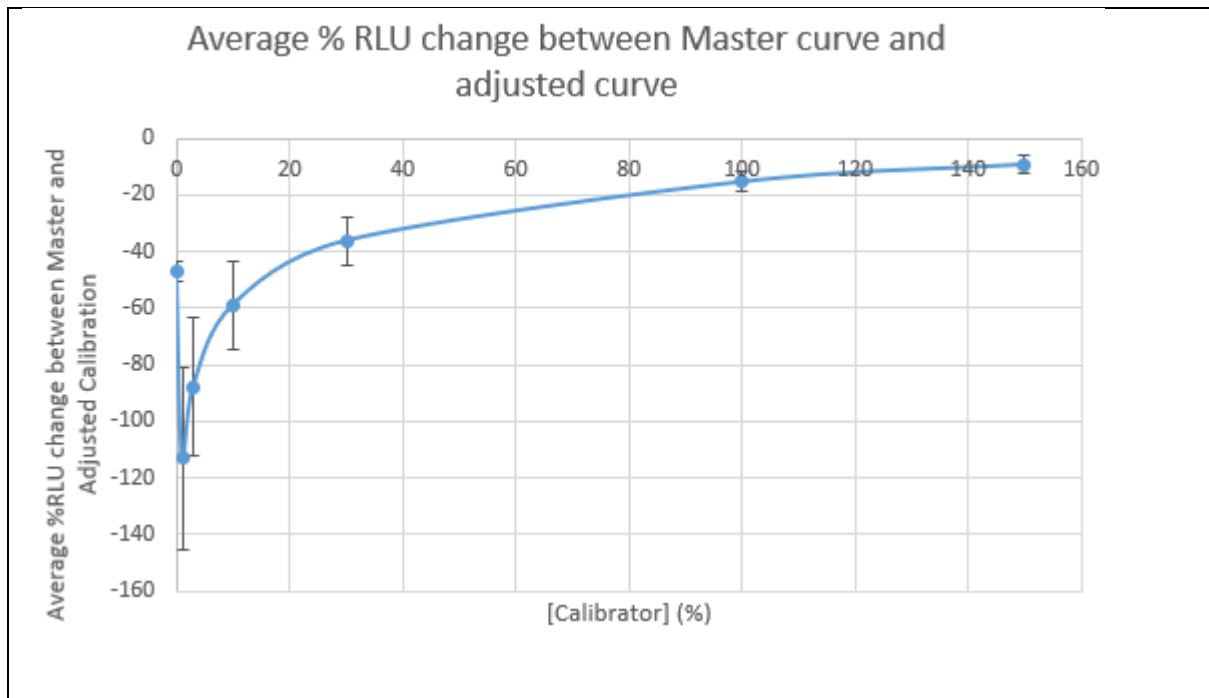


Figure 67: Bias correction using the HemosIL AcuStar® master curve

Average % RLU change between Master Curve and Adjusted Calibration Curve

The relationship between calibrator concentration (%) and the percentage change in Relative Light Units (RLU) between the Master and Adjusted Calibration Curves across five successive reagent lot changes.

There is notable lot-to-lot variability, with no clear pattern predicting how each lot will behave. This suggests that adjustment differences should be assessed more rigorously, as current calibration practices may not sufficiently account for these inconsistencies. Across all calibrations, there was a tendency for the local system to produce higher RLU values at the 10% decision limit, though this effect varied, with Lots 6429 and 6839 showing the most significant deviations.

The findings indicate that calibration adjustments introduce significant variability at the clinical decision threshold, potentially affecting diagnostic reliability. The lack of lot-to-lot consistency suggests a need for closer monitoring of calibration differences, particularly at low analyte concentrations where MU is most critical.

8.6.2.3 *Impact of master calibration adjustment at the clinical decision level*

Figure 68 illustrates a systematic upward shift in RLU values following calibration adjustment at 10% ADAMTS13 activity, with the Calibrated RLU consistently higher than the Master RLU across all reagent lots. The median RLU for the calibrated curve is significantly elevated, indicating a consistent bias introduced by the local calibration process.

The greater variability in the calibrated RLU values, as shown by the wider interquartile range and extended whiskers, suggests that lot-to-lot fluctuations increase after calibration adjustment, whereas the Master RLU remains more consistent. This aligns with previous findings that bias correction at the clinical decision threshold is not uniform across reagent lots, raising concerns about calibration stability and its potential impact on diagnostic accuracy for TTP.

These results highlight the need for closer monitoring of calibration performance, particularly at the 10% ADAMTS13 decision limit, where systematic shifts and increased variability could contribute to MU.

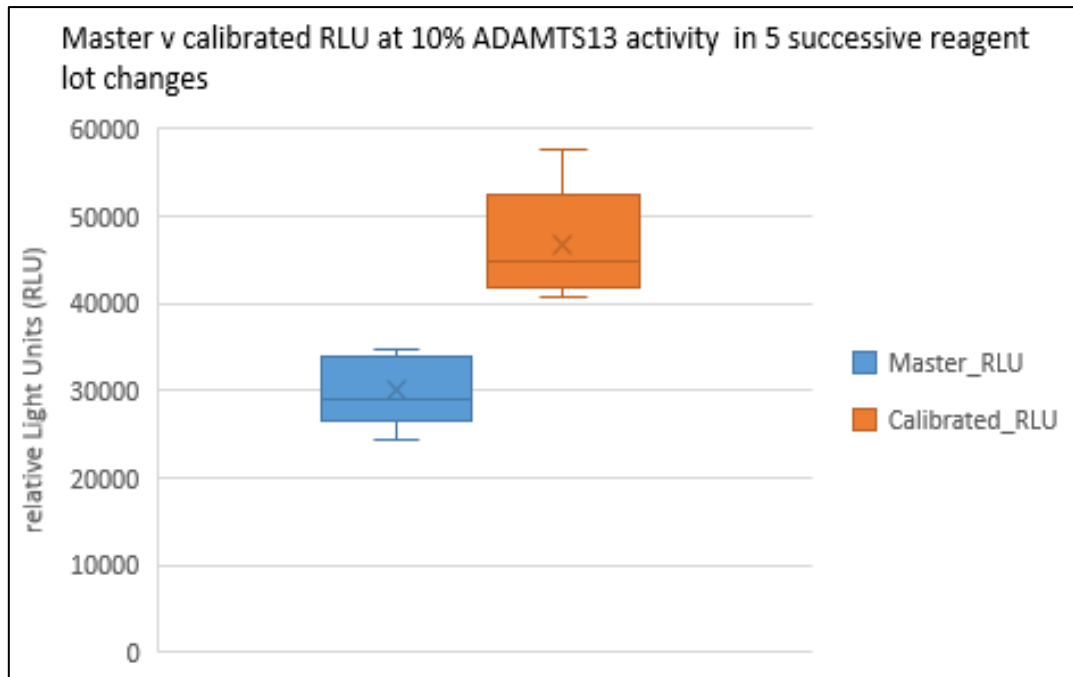


Figure 68: Comparison of Master and Calibrated RLU values at 10% ADAMTS13 activity.

Boxplot showing the distribution of Master RLU (blue) and Calibrated RLU (orange) at 10% ADAMTS13 activity across five reagent lots.

The line chart below (Figure 69) illustrates the RLU changes at 10% ADAMTS13 activity across five successive reagent lot changes, with individual lines representing each lot. In all cases, the Calibrated RLU values are consistently higher than the Master RLU values, confirming a systematic upward shift following calibration adjustment.

The slope of each line indicates the degree of adjustment required for each reagent lot. While some lots exhibit only a moderate increase, others show a steeper rise in RLU post-adjustment, suggesting variability in the extent of calibration correction between different lots. This further supports previous findings that calibration bias correction is inconsistent, particularly at clinically significant levels.

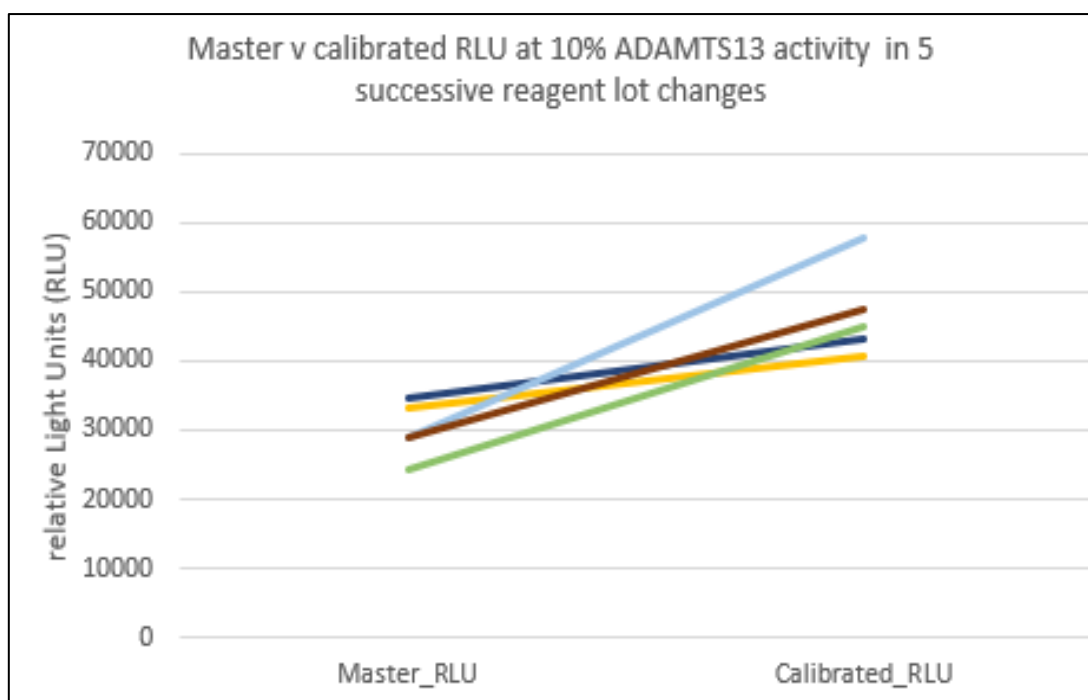


Figure 69: Lot-to-lot variability in RLU adjustment at 10% ADAMTS13 activity.

Line plot illustrating Master vs. Calibrated RLU values for five successive reagent lot changes. Each line represents an individual lot.

The spread of lines highlights lot-to-lot variability in calibration behaviour, with some reagent lots experiencing a more pronounced deviation than others. This reinforces concerns about the unpredictability of calibration adjustments at the clinical decision threshold, suggesting that current calibration processes may not fully standardise assay performance across reagent changes.

These findings confirm that systematic upward bias is introduced at 10% ADAMTS13 activity, but the degree of correction varies, potentially impacting MU and clinical interpretation. More rigorous evaluation of calibration consistency across reagent lots is needed to mitigate these effects.

The SEM can be considered as the standard uncertainty associated with the bias present in the calibration process, reflecting the extent of correction applied through local adjustment.

Consequently, the standard uncertainty due to bias (u_{bias}) varies across the calibration range. At the

10% clinical decision threshold, the average bias between the local and master calibrations over five calibrations across six months was 15%. This indicates that, without local adjustment, the bias would be clinically significant, supporting the necessity of applying local corrections to the master calibration to ensure accurate measurement and clinical reliability.

8.7 Chapter 8 Discussion

This chapter demonstrated that calibration processes, particularly model selection and calibration adjustment strategies, act as unrecognised contributors to measurement variability in ADAMTS13 activity assays. Traditionally considered a neutral background step within measurement uncertainty frameworks, calibration was shown here to directly influence both systematic error and between-run variability, supporting its reconsideration as a structural contributor to assay uncertainty.

The evaluation of single-run calibration models revealed that the choice of curve-fitting method significantly affects interpolation behaviour. Linear regression introduced consistent systematic bias across the calibration range, particularly underestimating mid-range results and overestimating both low and high concentrations. Higher-order polynomial models reduced interpolation uncertainty but introduced overfitting at higher concentrations. In contrast, four-parameter logistic and five-parameter logistic models provided the most reliable interpolation, minimising bias and instability across the dynamic range. This analysis established calibration model choice as a source of systematic error, particularly at clinically significant thresholds.

Extending this analysis to multiple calibration runs demonstrated that model-dependent differences are not limited to systematic shifts within single assays. Instead, the choice of calibration model influenced run-to-run variability, as reflected in the standard deviation of absolute bias across calibrations. Models with poor curve stability, such as spline fitting and linear regression, exhibited greater between-run variability, while 4PL and 5PL models offered more consistent calibration

performance. These findings indicate that structural variability introduced by calibration modelling is currently absorbed into within-laboratory reproducibility (u_{RW}) and misclassified as random variability in conventional measurement uncertainty estimation.

Monte Carlo modelling further demonstrated that calibration-derived uncertainty accumulates across the assay range. Using pooled calibration data and 4PL modelling, cumulative interpolation uncertainty was quantified across clinically relevant decision points. The analysis showed that measurement uncertainty is non-uniform, with greatest imprecision at the curve asymptotes, driven by inherent limitations in signal stability and curve geometry. These findings support the use of aggregated calibration and simulation-based approaches to quantify calibration-derived uncertainty, particularly where traditional single-run estimates fail to capture inter-assay variability.

Finally, evaluation of automated master calibration systems confirmed that calibration uncertainty is not eliminated by local corrections. Local bias correction factors, routinely applied to master curves in the HemosIL AcuStar® assay, introduced systematic shifts in calibration performance, particularly at the critical 10 IU/dL threshold. Lot-to-lot variability in these adjustments further highlighted that calibration correction introduces unrecognised uncertainty, which remains unaccounted for within current top-down MU estimation frameworks.

In summary, this chapter established that calibration model choice, cumulative interpolation variability, and local bias correction procedures all contribute structurally to measurement variability in ADAMTS13 activity assays. Calibration is not a neutral process: it introduces both systematic error and structural variability across runs, directly contributing to measurement uncertainty. Recognition of calibration modelling as a contributor to MU is essential for accurate uncertainty estimation in both manual and automated assay systems.

9 Chapter 9 – Interpreting ADAMTS13 Results: A

Clinically Risk-Based Framework

9.1 Introduction

This chapter presents a series of linked analyses to investigate the impact of MU on the clinical interpretation of ADAMTS13 activity results.

The first section uses the same post implementation verification data for the HemosIL AcuStar[®] that was used to optimise clinically relevant decision thresholds using likelihood ratios (LR) in chapter 5. MU is then added to simulate the range of LRs that could be observed around these thresholds. This approach quantifies how MU affects the strength of diagnostic evidence, expressed as variation in LR.

The second section extends this analysis by converting LR intervals into post-test probabilities, using pre-test probability estimates derived from the PLASMIC score. This allows evaluation of how MU affects diagnostic interpretation across different levels of clinical risk. Simulated data are used to demonstrate the range of post-test probabilities associated with a given result, incorporating both analytical and clinical sources of uncertainty.

The final section introduces the concept of interpretive robustness, defined as the range of post-test probabilities resulting from LR variation due to MU. This analysis explores how the impact of uncertainty differs depending on the pre-test probability, and whether certain clinical scenarios are more vulnerable to interpretive ambiguity. The findings are used to propose risk-based performance specifications that reflect not only analytical variation, but the consequences of that variation on clinical decision-making.

Together, these analyses aim to bridge the gap between analytical method performance and the clinical utility of results. The chapter develops a framework for quantifying and visualising interpretive risk, supporting more clinically aligned validation strategies.

9.2 Aims

This chapter aims to evaluate how MU affects the clinical interpretation of ADAMTS13 activity results, and to determine whether this impact can inform performance specifications based on diagnostic risk.

9.3 Quantifying Diagnostic Evidence from Uncertainty at the Decision

Threshold

9.3.1 Methods

9.3.1.1 MU Determination

The MU applied in this analysis was determined based on prior work detailed in Chapter 4 of this thesis, where a clinical outcome model was used to establish acceptable analytical imprecision for this assay. Specifically, the maximum allowable MU was set at $\pm 20\%$ relative to the decision threshold, corresponding to ± 2 IU/dL at the diagnostic cutoff of 10 IU/dL. This value was derived by modelling the potential clinical impact of analytical variability on patient outcomes, ensuring that the defined uncertainty limit remained clinically acceptable. The same MU estimate was applied in this section to illustrate how analytical imprecision influences the interpretation of LR^+ around the decision threshold.

9.3.1.2 LR^+ and MU Framework

To illustrate the effect of MU on diagnostic interpretation, a theoretical framework was developed in which LR^+ values were evaluated within the MU range. Specifically, for a representative result of 10 IU/dL, the corresponding range of LR^+ values for equivalent results between 8 and 12 IU/dL was identified. This approach allowed for the assessment of how analytical imprecision may influence the strength of diagnostic evidence provided by a positive test result.

9.3.2 Results

Figure 70 displays the relationship between LR^+ values and equivalent % results within the MU interval for a representative test result of 10 % on the HemosIL AcuStar[®]. The grey shaded area in the figure represents the MU range of ± 2 IU/dL (i.e. 20% MAU), corresponding to a result interval of 8–12 IU/dL with 95% confidence. The LR^+ values within this interval were derived from local post-implementation verification data (See chapter 5.4).

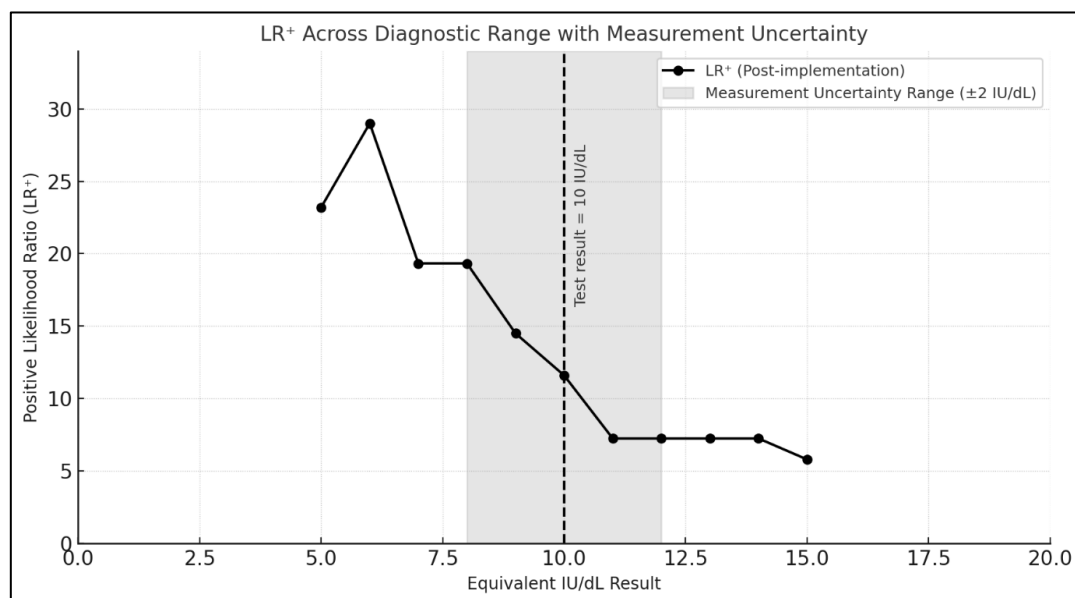


Figure 70: Measurement Uncertainty impact on positive likelihood ratio.

For a result of 10 IU/dL (or % on the AcuStar[®] the range of positive likelihood ratios possible can be used to represent the measured result $\pm$ MU.

A result of 8 IU/dL corresponds to a approximate LR^+ of 19.33, representing strong diagnostic evidence, while a result of 12 IU/dL corresponds to an approximate LR^+ of 7.25, indicating only moderate diagnostic strength. The central result of 10 IU/dL yields an approximate LR^+ of 11.60, near the commonly accepted threshold of $LR^+ \geq 10$ for strong evidence. This demonstrates the diagnostic informativeness of an apparently single numeric result is not fixed but instead varies within the MU interval. It highlights how small fluctuations around the decision threshold can lead to meaningful changes in LR^+ , thereby affecting the post-test probability of disease.

9.4 Bayesian Updating of ADAMTS13 Results Using PLASMIC-

Derived Risk

Having established how MU translates into a range of plausible likelihood ratios at clinically relevant cut-offs, the next section explores how these likelihood ratio ranges influence diagnostic interpretation. By applying Bayesian modelling, the effect of this uncertainty is assessed across different pre-test probability scenarios reflective of real-world clinical risk.

9.4.1 Methods

9.4.1.1 Pre-test Probabilities and Clinical Risk Groups

Pre-test probabilities were derived from the PLASMIC score, a validated clinical tool used to estimate the probability of severe ADAMTS13 deficiency (<15%) in patients with suspected TTP. Based on published validation data, three representative risk categories were selected at low risk: 0.1%, Intermediate risk: 6%, High risk: 72% (9, 13, 228).

These values were used to define three distinct clinical pre-test probability scenarios, each of which was carried forward into post-test probability modelling.

9.4.1.2 Likelihood Ratio Range and Simulation of Post-test Probability Curves

Post-test probabilities were calculated using Bayesian inference based on Pretest probabilities from the PLASMIC score as follows:

$$Pre\ Test\ odds = \frac{Pre\ Test\ Probability}{1 - Pre\ test\ Probability}$$

Equation 23: Calculating Pretest odds from Pre Test Probability (provided by the PLASMIC score)

$$Post\ Test\ odds = Pre\ Test\ odds \times Likelihood\ Ratio$$

Equation 24: Converting Pre Test odds to Post test Odds using the Likelihood ratio

$$\text{Post Test Probability} = \frac{\text{Post Test odds}}{1 + \text{Post Test odds}}$$

Equation 25: Conversion of Post test odds back to Post test probability

To illustrate the full relationship between LR and post-test probability for each clinical risk group, continuous curves were simulated across a log-scale range of LRs from 0.01 to 100. This range was chosen to cover the full range of LR values from weak to very strong diagnostic evidence, and to allow visualisation of both low and high post-test probability behaviour for each pre-test probability.

For each LR point, the Bayesian formula was applied to calculate post-test probabilities corresponding to each of the three pre-test risk groups. This allowed the generation of smooth sigmoid-like curves reflecting how diagnostic confidence changes with varying test performance.

9.4.1.3 Integration of Measurement Uncertainty

The shaded region in the resulting plot corresponds to an LR interval of 7.25 to 19.33, representing the simulated effect of MU at the clinical cut-off of 10 IU/dL. This interval was derived from prior modelling of local assay verification data, where a $\pm 20\%$ maximum allowable uncertainty (MAU) was applied around the cut-off (see sections 4.3 and 5.4) .

This LR range was then overlaid onto the full sigmoid curves for each pre-test probability. Vertical dashed lines and annotations indicate the post-test probability range attributable to the LR uncertainty at each clinical risk level, providing a visual representation of how MU may affect diagnostic interpretation.

9.4.1.4 Computational Implementation

All post-test probability calculations and figure generation for this section were performed using Python (v3.10), with the following libraries: NumPy (v1.24.4) for numerical operations and probability transformations, and Matplotlib (v3.7.2) for plotting.

9.4.2 Results

Figure 71 displays the relationship between likelihood ratio (LR) and post-test probability for three clinical scenarios, defined by pre-test probabilities of 0.1%, 6%, and 72%, corresponding to low, intermediate, and high-risk groups from the PLASMIC score. Each pre-test probability is represented by a sigmoid curve, generated from LR values ranging from 0.01 to 100. The shaded region highlights the LR interval of 7.25 to 19.33, derived from MU applied to a result at the 10 IU/dL clinical decision limit.

For the 72% pre-test probability group, the curve is positioned in the upper portion of the graph, with post-test probabilities ranging from 94.9% to 98.0% across the shaded LR interval. The corresponding vertical span is narrow, indicating relatively little change in post-test probability within this LR range.

The 6% pre-test group shows a broader curve centred in the middle of the graph. In this case, post-test probabilities range from 31.6% to 55.2% across the same LR interval, representing a noticeably wider vertical span.

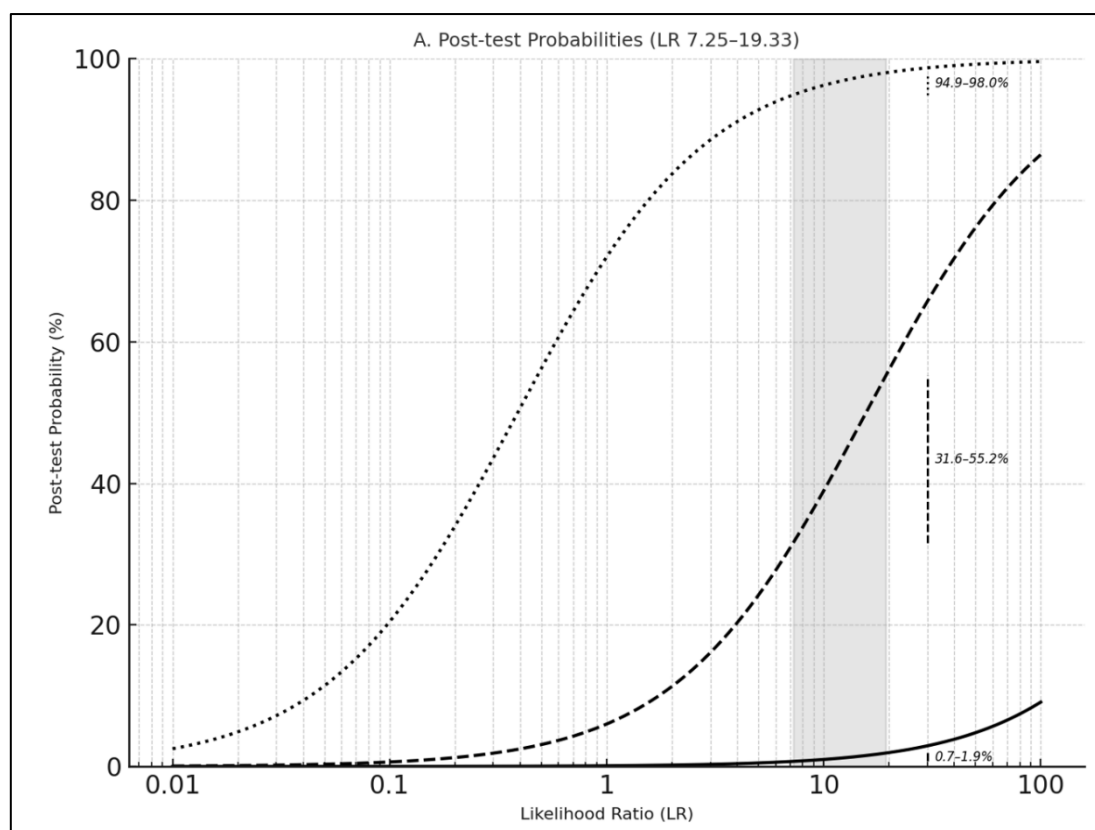


Figure 71: Post test probabilities using the PLASMIC score and simulated test performance.

Three pre test probabilities were used to demonstrate the risk ratings from the PLASMIC score. 72% (High risk – dotted line), 6% (intermediate risk – Dashed line) and 0.1% (low risk – Solid line).

For the 0.1% pre-test group, the curve lies in the lower region of the graph, with post-test probabilities rising modestly from 0.7% to 1.9% across the LR range. As with the high pre-test group, the vertical span here is limited, with minimal variation in post-test probability.

The figure illustrates that, although the LR interval is fixed, the extent of variation in post-test probability differs between pre-test risk groups. The range is narrowest at the extremes and widest in the intermediate-risk setting. This implies a degree of interpretive robustness that is both pretest probability and analytical performance related. This is explored further in the following section.

In addition to the primary LR band, a second interval was modelled based on an alternative diagnostic threshold of 8 IU/dL, derived from prior local optimisation work. Applying the same 20% MU yielded a result range of 6.4 to 9.6 IU/dL, corresponding to an LR interval of 12.0 to 26.2. This range, positioned further to the right on the x-axis, falls within a region of higher diagnostic strength. For each pre-test probability, the resulting post-test probability span was narrower than the 7.25–19.33 interval, indicating increased interpretive stability. This effect is explored in greater detail in the subsequent section on interpretive robustness.

9.5 Interpretive Robustness: Quantifying the Clinical Consequences of MU

While the preceding section demonstrates how likelihood ratio variation affects post-test probability in different risk groups, it also reveals that the magnitude of this effect is not uniform. The following section investigates this more directly by introducing the concept of interpretive robustness, quantifying how susceptible each clinical scenario is to diagnostic ambiguity arising from MU.

9.5.1 Methods

9.5.1.1 Objective and Conceptual Framework

This section evaluates interpretive robustness — the susceptibility of diagnostic interpretation to uncertainty in likelihood ratio, propagated through to post-test probability. The goal is to quantify how wide the range of possible post-test probabilities is for a given LR value, under simulated uncertainty, and to assess whether this variability is acceptable in a clinical context.

Interpretive robustness is operationally defined here as the vertical spread between the maximum and minimum post-test probabilities resulting from a specified LR uncertainty range. This was calculated for multiple pre-test probability levels, allowing comparison of robustness across different clinical risk settings.

9.5.1.2 Pre-Test Probabilities

Pre-test probabilities were the same as those used in the prior section, based on PLASMIC score-derived clinical risk categories i.e. low: 0.1%, intermediate: 6%, high: 72% as per the validation data for the PLASMIC score (9, 13, 228).

These values represent realistic starting probabilities for a patient being tested for severe ADAMTS13 deficiency, and each was modelled independently.

9.5.1.3 Central Likelihood Ratio Values and Simulation Range

Interpretive robustness was evaluated across a range of central LR values, spanning from 1 to 100.

Central LR values represent the LR that would be returned by a test, based on its reported result in the absence of uncertainty, using the post implementation dataset used in section 5.4.

To simulate the effect of MU on this LR, a symmetric interval of $\pm 50\%$ around each central value was used, creating LR bands such as, for LR = 10 the band ranges from 5 to 15, for LR = 20 the band ranges from 10 to 30. The simulation used a fixed $\pm 50\%$ range around each central likelihood ratio value to model uncertainty in post-test probability interpretation. While the underlying measurement uncertainty was defined as $\pm 20\%$, the non-linear relationship between test value and LR — particularly near clinical thresholds — results in LR variation that often exceeds $\pm 20\%$. For example, a $\pm 20\%$ range around a 10 IU/dL result (i.e. 8–12 IU/dL) corresponded to an observed LR range of 7.25–19.33, which approximates $\pm 50\%$ from the central LR. Thus, the $\pm 50\%$ modelling approach more accurately reflects the real-world uncertainty in diagnostic evidence introduced by measurement variability, and enables consistent simulation of interpretive robustness across the LR spectrum.

9.5.1.4 Calculation of Post-Test Probability Range Widths

For each pre-test probability (0.1%, 6%, and 72%) and for each central LR value the lower and upper LR values (central $\times 0.5$ and central $\times 1.5$) were used to calculate corresponding post-test probabilities using Bayes' theorem (see equations Equation 23, Equation 24 and Equation 25 from Chapter 9.4).

The range width was then calculated as $\text{Post-test probability range} = \text{Post-test}_{\text{upper}} - \text{Post-test}_{\text{lower}}$

The result was expressed as a percentage and plotted as a function of the central LR. This process yielded a separate interpretive robustness profile for each clinical risk group.

9.5.1.5 Analytical Performance Specification zones

To represent an acceptable amount of diagnostic uncertainty as an APS, limits of 1%, 5% and 10% were chosen as ranges to equate to optimum, desirable and minimum performance to align with previously described limits . These numbers represent the width of the uncertainty (diagnostic uncertainty) induced at a given LR, in the presence of MU, under a pre-defined clinical pre-test probability.

9.5.1.6 Visualisation and Performance Thresholds

The resulting graph (Figure 72) plots post-test probability range width on the Y-axis against the central likelihood ratio on the X-axis, using a logarithmic scale. Each line represents a different pre-test probability group in that the solid line is high pre-test probability (72%), the dashed line is intermediate pre-test probability (6%) and the dotted line is low pre-test probability (0.1%).

To support interpretation, horizontal coloured reference lines were added to define performance specification zones according to Table 47, below.

Table 47: Performance thresholds and associated limits used for interpretive robustness chart

Performance Threshold (Robustness)	Range Width	Colour
Optimal	< 1%	Green
Desirable	< 5%	Blue
Minimum	< 10%	Orange

These thresholds were chosen to reflect clinically meaningful boundaries of interpretive variability, acknowledging that uncertainty has a greater clinical impact in some contexts than others.

Each data point on the curves corresponds to a simulated diagnostic scenario, and the vertical distance from each curve to the performance thresholds allows direct assessment of whether a given LR, in a given clinical context, maintains acceptable interpretive stability under uncertainty.

9.5.2 Results

Figure 72 illustrates the change in interpretive robustness — expressed as the range of post-test probabilities (%) — across a series of increasing central likelihood ratios, for three distinct pre-test probability scenarios.

For the intermediate pre-test probability group (6%, Figure 72 dashed line), the post-test probability range was widest overall. The curve peaked at a central LR value of approximately 20, where the post-test probability span exceeded 27%. This group consistently demonstrated the greatest variation in diagnostic interpretation arising from LR uncertainty, particularly across central LR values between 5 and 50. The curve traversed all three interpretive robustness performance zones, crossing above the minimum acceptable threshold (>10%) for much of the LR range.

The low pre-test probability group (0.1%, Figure 72, dotted line) displayed a steadily decreasing range of post-test probabilities as the LR increased. At low LRs, variability was high (above 20%), but the curve declined sharply, falling below the minimum threshold around LR = 10 and reaching within the optimal zone (<1%) by LR = 50. This indicates that interpretive robustness improves rapidly for this group as the strength of diagnostic evidence increases.

The high pre-test probability group (72%, Figure 72, solid line) showed the opposite trend. Post-test probability variation remained very low across most of the LR range but increased with higher LR values, peaking modestly above the desirable zone (>5%) near LR = 100. However, at central LRs below 20, this group consistently remained within the optimal or desirable performance bands, indicating stable interpretation across a broad range of plausible LR values.

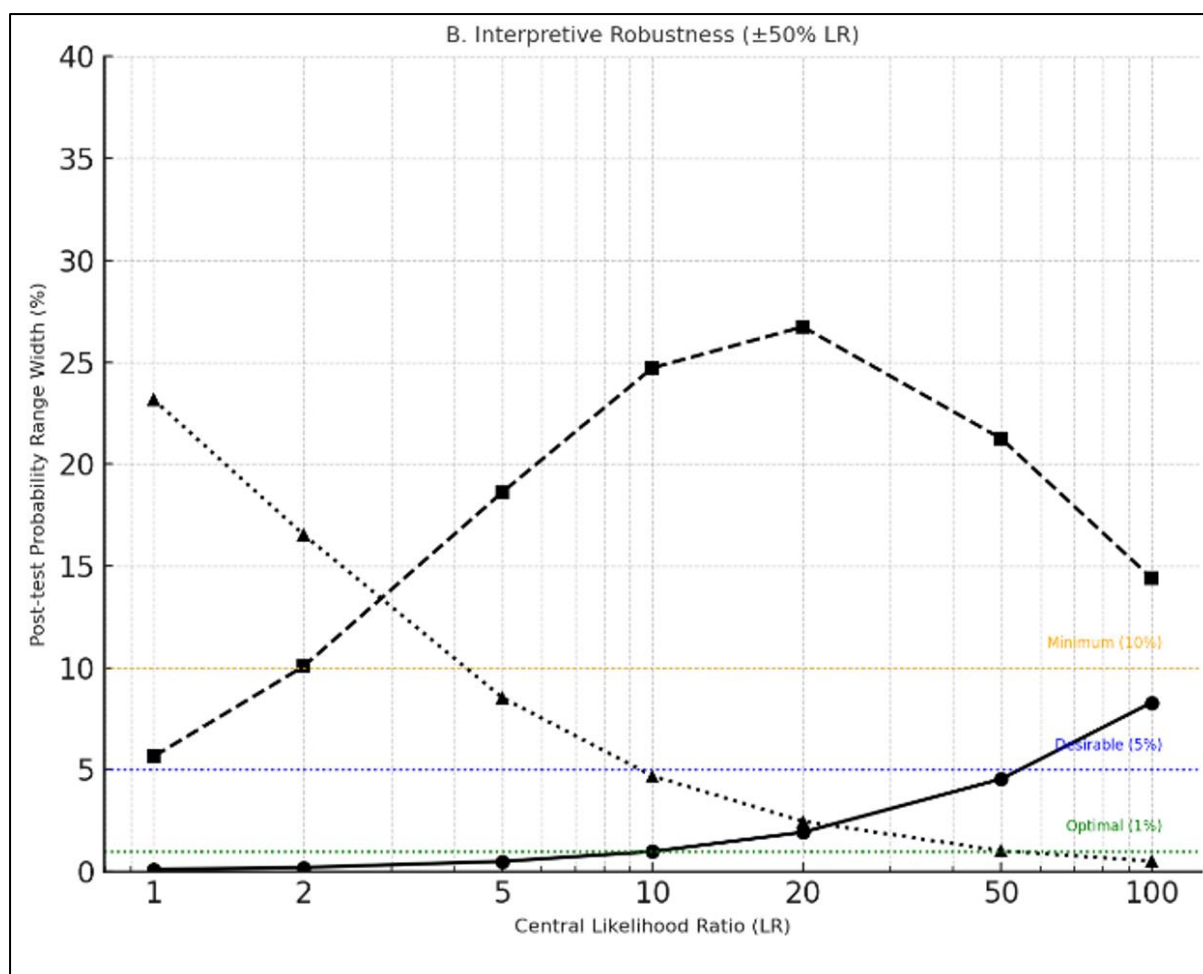


Figure 72: Interpretation robustness of low, intermediate and high pre test probability as a function of analytical performance

Collectively, the three curves demonstrate that interpretive robustness varies systematically by clinical context. The widest ranges — and therefore greatest susceptibility to diagnostic instability — occurred in the intermediate-risk group.

9.6 Chapter 9 Discussion

This chapter has presented a novel, integrated framework that links local assay verification, MU, LR modelling, and Bayesian interpretation to support clinically meaningful diagnostic decision-making. The work builds on earlier thesis sections by advancing beyond the analytical domain, embedding variability directly into diagnostic interpretation, and proposing tools to bridge the gap between laboratory performance and clinical risk.

In Section 9.3, MU was applied to the guideline recommended cut off of 10 IU/dL (3, 33) generating a range of plausible positive likelihood ratios (LR^+) that might result from a single test value. The resulting range of LR^+ (7.25–19.33) was clinically meaningful, spanning moderate to strong diagnostic evidence. This demonstrated that even a well-defined result can yield a spectrum of diagnostic interpretations due to analytical imprecision. The analysis reinforced the limitations of binary thresholds and underscored the value of representing uncertainty probabilistically. The broader implication is that MU, often relegated to a technical afterthought, has direct consequences for the communication of diagnostic confidence and must be actively accounted for when defining assay performance.

Section 9.4 extended this approach by translating LR intervals into post-test probabilities using pre-test estimates derived from the PLASMIC score. This created simulated diagnostic scenarios across three risk groups: low (0.1%), intermediate (6%), and high (72%). The Bayesian curves revealed how the same LR range produced different shifts in post-test probability depending on the clinical context. In particular, the intermediate-risk group demonstrated the widest span of possible post-test probabilities — ranging from indeterminate to moderately high probability — as a result of LR variation. This finding highlighted that uncertainty in test performance does not affect all patients equally. Where pre-test risk is low or high, post-test shifts are more constrained and less vulnerable to variability. In contrast, intermediate-risk cases may be pushed across key diagnostic boundaries depending on the degree of uncertainty present. This observation links directly to clinical risk

tolerance and raises important questions for laboratories about how and when MU must be explicitly communicated.

Section 9.5 developed this observation further by formalising the concept of interpretive robustness. This new term describes the range of post-test probabilities that can result from a given LR subject to measurement uncertainty. The robustness curves presented here quantify this spread across a full spectrum of LR values, for each pre-test probability. The analysis revealed that the intermediate-risk group again exhibited the greatest susceptibility to diagnostic instability, particularly at central LR values between 10 and 30. These results were contextualised using performance specification zones (<1%, <5%, <10%, >10% post-test probability range), adapted to reflect a proposed framework for diagnostic interpretive thresholds. Although the development of this framework remains exploratory, the results suggest that interpretive robustness could serve as a clinically aligned performance specification — one that directly incorporates both MU and the consequences of diagnostic ambiguity.

The introduction of this performance specification is novel. While ISO/TS 20914 and related guidance emphasise the importance of uncertainty in reporting, the direct translation of MU into post-test probability variability, and the formalisation of this as a measurable specification of diagnostic quality, has not previously been described. This approach also allows for local adaptation because robustness depends not only on test precision but also on LR calibration and pre-test probability, laboratories can tailor specifications to reflect their patient population and verification data.

A further strength of this study lies in its use of visual simulation as a diagnostic communication tool. The modified Fagan plots and interpretive robustness curves provide intuitive representations of uncertainty that move beyond traditional statistical descriptors. These visualisations offer an opportunity to inform clinical decision-making in real time, especially in equivocal scenarios. Although not yet suitable for routine reporting, their conceptual clarity supports the wider aim of making laboratory uncertainty both visible and actionable.

Finally, by including two different LR intervals — one based on the traditional 10 IU/dL cut-off and the other based on an assay-optimised 8 IU/dL threshold — this work links back to prior analysis in the thesis concerning cut-off optimisation and clinical calibration. The same $\pm 20\%$ MU applied to different thresholds yielded distinct LR bands and therefore different implications for post-test probability. This illustrates how even small changes in cut-off strategy interact with uncertainty to shift diagnostic meaning — further justifying the need for dynamic, probabilistically informed approaches to assay evaluation.

Together, these findings offer a unified model that connects the analytical verification process with clinical reasoning. MU is no longer confined to abstract statistical limits but is embedded directly in the diagnostic pathway. The concept of interpretive robustness offers a tangible, clinically weighted endpoint by which assay performance — and its limitations — can be evaluated and communicated. While further development and standardisation are required, particularly in determining acceptable levels of post-test probability range in different contexts, this chapter provides the first demonstration of a fully integrated model linking method verification, MU quantification, diagnostic predictive power, diagnostic uncertainty and clinical risk in a single, clinically interpretable framework.

10 Thesis Discussion

10.1 Introduction

This thesis has explored the complex and under-addressed topic of MU in the context of ADAMTS13 activity assays, which play a critical role in the diagnosis and management of TTP (3, 33) . The accurate measurement of ADAMTS13 activity is fundamental to differentiating between TTP and other thrombotic microangiopathies, guiding urgent treatment decisions that can significantly impact patient outcomes (30, 227). However, despite the centrality of these assays in clinical decision-making, current international guidance documents—such as ISO 15189:2022, ISO/TS 20914:2019, and CLSI EP29-A—provide only broad principles on MU estimation, offering limited practical guidance for its implementation in specialised assay contexts like ADAMTS13 testing.

The research presented in this thesis was motivated by the recognition that MU is often underutilised in laboratory practice, and that its application is particularly challenging in assays characterised by low concentrations, tight clinical decision thresholds, and complex calibration models. The potential impact of uncertainty on result interpretation, method verification, and patient management in this setting remains poorly defined.

Accordingly, the overarching aim of this thesis was to develop and evaluate a structured, clinically informed, and statistically robust framework for MU estimation and application in ADAMTS13 activity assays. Through a series of interrelated studies, the thesis investigated the methodological, analytical, and clinical dimensions of MU, covering derivation of APS (Chapter 4), method comparison practices (Chapter 5), IQC-based uncertainty estimation (Chapter 6 and 7), calibration model evaluation (Chapter 8), and advanced interpretative strategies (Chapter 9).

Despite much progress in the clinical chemistry arena it is recognised harmonisation is needed in all disciplines (258). It ensures laboratory results can be interpreted reliably and consistently (259).

Inconsistencies in how assays define, calibrate, and report ADAMTS13 activity have direct consequences for clinical interpretation. Despite all assays quantifying ADAMTS13 proteolytic activity, substantial variability existed in their measurand components (Summarised in Appendix A). Differences were identified in calibration traceability, detection methodology, sample type specification, and unit assignment, reflecting a lack of standardisation in assay design and documentation, influencing many of the findings in later chapters.

This lack of harmonisation is not unique to ADAMTS13 activity testing. It reflects a broader challenge observed across haemostasis assays, where the definition of the measurand is often complex and assay-specific. For example, in Factor VIII activity assays, significant variability exists in how different assays define and measure Factor VIII. Some assays quantify Factor VIII clotting activity using one-stage clotting methods, while others use chromogenic substrate-based approaches (167). The evolution of Factor VIII treatments, from plasma-derived to recombinant, extended half life, and more recently, gene therapy and non-replacement therapies such as emicizumab, has further complicated the measurand definition (192). Therapeutic innovations alter the molecular form or function of the analyte, meaning that different assays may detect, under-recover, or fail to measure certain therapeutic agents depending on the assay design (260-262). This heterogeneity in assay behaviour is rarely fully captured in the measurand definition and poses significant challenges for comparability, verification, and uncertainty estimation.

The lack of transparency from some assay manufacturers, particularly regarding calibration traceability and performance characteristics, constrained the ability to fully evaluate measurand components. Additionally, the absence of regulatory requirements mandating disclosure of measurand definition details represents an ongoing barrier to harmonisation efforts recognised in other disciplines of significant importance in haemostasis (68, 259, 263).

Future work should focus on engaging with assay manufacturers, regulatory bodies, and professional organisations to promote the routine adoption of transparent, standardised measurand definitions.

Such harmonisation is essential to support the broader application of MU principles in laboratory medicine and to ensure that analytical verification processes are robust, clinically meaningful, and aligned with international quality standards.

10.2 Summary of findings

10.2.1 Analytical Performance Specifications for ADAMTS13 Activity Assays

The first study (Chapter 4) examined the derivation of APS for ADAMTS13 activity assays. APS ensure analytical methods deliver results that support clinical decision-making. However, this study highlighted significant limitations in the current approaches used to define APS, particularly when applied to complex haemostasis assays such as ADAMTS13 activity measurements. In particular recognising that ADAMTS13 activity measurements are used in a variety of clinical context directly challenges the accepted approach of using single APS for all uses. In ADAMTS13 assays results here show that defining different APS models depending on the use of the result is a pragmatic, albeit potentially complicated solution to provide better transparency on performance.

The evaluation of biological variation-based models, which form the basis of many APS derivation strategies, demonstrated that these models were of limited utility for ADAMTS13 activity. Crucially, the available biological variation data for ADAMTS13 are derived from two published studies (108, 112). Moreover, BV studies inherently focus on variability in health, capturing the biological fluctuation of measurand concentrations in healthy individuals under stable physiological conditions (104, 110, 180, 205, 264). This approach fails to capture the increased variability and altered measurand behaviour in acute clinical settings, such as the presentation of TMA, where ADAMTS13 activity may be critically reduced and biological systems are dysregulated.

The reliance on BV-derived specifications in this context is therefore fundamentally flawed. There is no evidence to support the assumption that the biological variation of ADAMTS13 activity in health is representative of variability during acute disease presentations, nor that performance specifications

derived from health-based data adequately reflect the analytical requirements needed to support urgent clinical decision-making.

To address these limitations, the study employed simulation-based modelling to derive APS that incorporated clinical decision thresholds, patient misclassification risk, and MU (section 4.3). This approach aligned with the Milan criteria's Model 1b framework (96, 100) but extended it by applying Monte Carlo simulation to evaluate platform-specific MAU values. Importantly, this study represents the first application of clinical outcome simulation in the derivation of APS for ADAMTS13 activity assays, and one of the first for haemostasis in general. By linking analytical performance to patient misclassification risk at decision thresholds, the simulation-based framework provides a performance specification model that is both clinically relevant and operationally meaningful (100).

The findings revealed that the derived MAU thresholds varied across assay platforms, reflecting differences in analytical performance and calibration models, further studied later in the thesis. The simulation-based approach moved beyond static and somewhat arbitrary thresholds generated by BV-based models.

This Chapter also looked at using patient centric biological variation as an alternative method of determining personalised interpretation of results, in particular in patients in relapse (section 4.4). Individual reference ranges representing stable remission phases for patients who, in some cases may not return baseline ADAMTS13 into the traditional reference ranges demonstrated a more nuanced interpretation of changes in longitudinal results. Including local population derived reference ranges using the refineR algorithm also demonstrated that there are distinct categories of sub populations of results, whether for all samples or first time presentation samples. Additionally, these distinct categories, akin to those described for sodium previously (229) are platform dependent and influence interpretation of abnormality.

The study also provided the first investigation of APS in the context of haemostasis treatment monitoring. By incorporating pharmacokinetic modelling of ADAMTS13 replacement therapies (213,

214), including recombinant ADAMTS13 products, the analysis evaluated how APS thresholds may be adjusted to reflect expected treatment regimes. However, this area of APS derivation introduces further complexity. The therapeutic products used to treat ADAMTS13 deficiency may differ structurally and functionally from native, patient-derived ADAMTS13 (210, 265, 266) . Recombinant ADAMTS13, while purified and standardised, is synthetically manufactured and may not behave identically to plasma-derived ADAMTS13, future provision of real world, in patient data will help to understand this better.

The challenges posed by novel or artificially adapted ADAMTS13 molecules parallel issues already seen in Factor VIII activity measurement (260, 261). As animal model-derived or synthetically engineered ADAMTS13 products continue to emerge in research and therapeutic settings (224, 267), similar challenges are likely to arise. MU and APS frameworks will need to account for the potential variability introduced by these novel treatments, ensuring that assay performance remains clinically meaningful.

Preanalytical factors were also evaluated in the context of APS derivation, specifically sample stability by the CRESS framework (268, 269). The findings confirmed that ADAMTS13 activity measurements remained stable for up to 24 hours at room temperature, with degradation exceeding clinically acceptable thresholds beyond 48 hours (270). This highlighted the need to integrate preanalytical variables into APS development, ensuring that sample handling and transport conditions do not undermine assay performance.

The approach to APS derivation presented in this thesis has broader relevance across haemostasis assays. The limitations of BV-derived specifications are similarly evident in other tightly regulated haemostasis measurands, such as antithrombin, protein C, or Factor VIII activity. For example, in Factor VIII assays, biological variation data cannot adequately account for the analytical variability introduced by different assay formats or the pharmacodynamic effects of EHL factor VIII products and non-replacement therapies such as emicizumab. The simulation-based, risk-informed APS

framework developed in this thesis therefore offers a transferable methodology for deriving clinically meaningful performance criteria in other complex haemostasis assays, particularly where BV data are sparse, outdated, or irrelevant to acute disease states.

However, the study identified practical limitations. The simulation-based approach relies on accurate estimation of patient misclassification risk and requires detailed knowledge of assay performance characteristics and clinical decision limits. These data may not always be readily available in routine laboratory settings. Furthermore, while pharmacokinetic-informed APS derivation is conceptually valuable, its implementation may be constrained by variability in individual patient responses, and the introduction of novel therapeutic products with uncertain behaviour in laboratory assays.

Future work should focus on validating this multifaceted APS framework across multiple laboratories and patient populations and on developing practical tools to support its routine application.

10.2.2 Method Comparison and Clinical Risk Evaluation

Chapter 5 evaluated the comparability of ADAMTS13 activity assays and examined how analytical differences between methods translate into clinical risk. This study demonstrated that conventional statistical approaches to method comparison are frequently insufficient to capture the analytical and clinical complexities associated with ADAMTS13 testing (section 5.3).

A critical appraisal of published method comparison studies (section 5.1, Appendix I) revealed consistent methodological shortcomings. Many studies relied heavily on correlation analyses, reporting high correlation coefficients without adequately evaluating systematic bias or assessing clinical implications. Correlation coefficients do not assess agreement between laboratory methods, as they are sensitive only to random error and fail to detect systematic bias (128, 242, 271).

Regression methods such as Passing-Bablok and Deming regression were inconsistently applied and, even when used, were rarely interpreted in a clinically meaningful way. Bland-Altman analysis, which more effectively assesses agreement and bias across the measurement range, was infrequently

reported. More fundamentally, many published studies failed to define APS or clinical acceptability criteria in advance of the comparison exercise, limiting the utility of their findings for laboratory verification or procurement decisions.

This deficiency is particularly problematic in the early implementation of new assay methodologies, when laboratories and clinical services require robust, transparent data to support procurement, verification, and patient management decisions. It is essential that method comparison studies are designed and conducted with scientific rigour, including pre-definition of performance specifications, appropriate sample selection across the full measurement range, and clear reporting of compliance (or otherwise) with those specifications. This requirement is heightened when the clinical risks associated with incorrect classification or interpretation of results are high, as in the diagnosis and monitoring of TTP.

Conventional statistical metrics, particularly point estimates such as correlation coefficients, slope, and mean bias, fail to fully capture the clinical implications of assay differences (127, 246, 249). Confidence intervals, when reported, reflect sampling uncertainty but do not account for analytical variation or the practical consequences of result discrepancies. Importantly, these point estimates obscure the clinical significance of outlier or borderline results, which often carry the greatest patient risk. The practice of highlighting overall assay agreement while overlooking clinically meaningful disagreement in individual cases represents a critical deficiency in many published studies (239, 251, 272).

To address these gaps, this study simulated MU in method comparisons. Increasing analytical imprecision significantly degraded correlation coefficients and regression parameters (section 5.3.2). However, while MU modelling improved the analytical evaluation, it remained limited in its ability to fully inform clinical interpretation.

To overcome these limitations, the study introduced a novel ADAMTS13 error grid framework, adapted from clinical risk assessment methodologies previously used in glucose and coagulation

testing (129-131). The error grid incorporated clinically relevant decision thresholds, including the critical 10 IU/dL threshold used in the diagnosis of iTTP (3, 33, 201), and stratified the potential clinical impact of assay discrepancies into low-, moderate-, and high-risk zones. Guard bands were incorporated to account for MU, ensuring that classification of discrepancies reflected both analytical variability and clinical significance.

The application of the error grid revealed that even small analytical differences between assays could translate into meaningful clinical risk, particularly in borderline cases near decision thresholds. For example, assay bias or increased imprecision at low activity levels could result in patient misclassification, influencing diagnosis and treatment decisions (section 5.5). By contrast, larger discrepancies at high activity levels, outside clinical decision thresholds, were less likely to affect patient care. This patient-centred, risk-based approach provided an intuitive and clinically meaningful framework for method comparison, shifting the focus from purely statistical agreement to clinical consequence.

The implications of this approach extend beyond ADAMTS13 activity testing. In haemostasis assays more broadly, method comparison studies frequently fail to assess whether observed assay differences matter in clinical terms. For example, in Factor VIII activity assays, biases of up to 20% have historically been accepted, and more recently, acceptance criteria have extended to 25% in the case of novel products such as efanesoctocog alfa (Altuviiiio®) (273). However, these bias thresholds are arbitrary and have not been systematically linked to clinical risk. The application of an error grid framework to Factor VIII assays would provide a more robust means of assessing whether observed differences in assay results are clinically significant. Notably, the presence of known assay bias, as seen with certain Factor VIII products such as Afstyla® (191), has required assay-specific correction factors. These corrections directly impact uncertainty estimation and demonstrate the interconnectedness of bias, MU, and clinical decision-making.

More broadly, this work identifies an urgent need for higher standards of study design, analysis, and reporting in method comparison studies. The current tolerance of poor statistical reporting, absence of *a priori* performance criteria, and over-reliance on inadequate summary statistics undermines the quality and clinical utility of these studies. A concerted effort is required from journals, peer reviewers, and the wider scientific community to demand greater scientific rigour and reject method comparison studies that fail to meet these essential requirements.

Future work should focus on prospective validation of assay-specific error grids, engagement with clinical stakeholders to evaluate their practical utility, and incorporation of these frameworks into laboratory verification protocols and regulatory guidance. Such work will help ensure that method comparison studies are designed and reported with the scientific rigour, transparency, and clinical relevance necessary to support laboratory practice and patient safety.

10.2.3 MU Estimation Using Internal Quality Control Data and Calibration

Modelling

The following three studies presented in this thesis critically evaluated underexplored but essential contributors to MU estimation in ADAMTS13 activity assays: the use of IQC data and the influence of calibration modelling strategies.

The use of IQC data for random uncertainty estimation represents a pragmatic and accessible approach, consistent with international recommendations such as ISO 15189:2022 and ISO/TS 20914:2019 (103, 149, 151). However, this thesis demonstrated that while the top-down approach offers operational convenience, its validity depends on appropriate application and rigorous statistical treatment. Specifically, the study revealed that the conventional use of long-term IQC standard deviation systematically underestimating random uncertainty (u_{RW}) in assays employing replicate measurements, such as the Technozym® ADAMTS13 ELISA method. This overestimation resulted from the use of summary statistics derived from replicate mean values, which failed to

account for the precision gain achieved by averaging replicates and conflated random and systematic sources of variability (section 6.3.2).

To address this limitation, variance components analysis was applied, in line with CLSI EP29-A guidance, allowing for the partitioning of within-run and between-run variability. This produced a more representative estimation of u_{RW} . A further important finding was that overly simplifying IQC data, particularly by treating replicate mean values as singular measurements without stratification, risks underestimating MU and may lead to inappropriate conclusions regarding assay performance.

A novel and significant methodological contribution of this thesis was the application of time series analysis to IQC data—an approach not previously described in the context of MU estimation in laboratory medicine. Autocorrelation assessment, runs testing, and decomposition of time series data revealed the presence of transient systematic variation in automated assay platforms, particularly in the HemosIL AcuStar[®] system.

This application of time series analysis provided valuable insight into the ongoing debate between MU and total error frameworks, particularly concerning the treatment of small, transient biases (102, 103, 203, 274). The EFLM contend that such biases are incorporated into intermediate precision estimates as they average out over time (103, 154, 188). TEa proponents, by contrast, argue that these biases require separate consideration (203).

However, the study also identified a significant gap in current guidance: the absence of defined performance specifications or thresholds against which preliminary MU estimates can be evaluated. There is a risk that preliminary MU estimation, particularly when based on short-term IQC data, may underestimate real-world uncertainty by failing to capture intermediate precision and long-term variability. This limitation could result in laboratories accepting analytical methods that initially appear acceptable but later demonstrate excessive MU once implemented in routine practice.

A further critical component of MU estimation explored in this thesis was the influence of calibration modelling strategies. In highly complex assays such as ADAMTS13 activity testing, the calibration model itself introduced interpolation uncertainty (section 8.3). Conventional single-run strategies, particularly those employing linear or polynomial regression, inadequately characterised the sigmoidal dose-response relationship of ELISA-based assays. Linear regression, while convenient, introduced systematic bias, particularly at low and mid-range concentrations, and higher-order polynomial models risked overfitting at the upper range (section 8.3.2.1). Overfitting of calibration data may create false assurance about the acceptability of calibration activities, with inflated r^2 statistics masking the poor biological fit of the model. This practice has become increasingly common, deviating significantly from the traditional linear calibration modelling approaches historically applied in coagulation factor assays where parallelism is maintained.

In biological assays such as ADAMTS13 activity measurement, it is essential that manufacturers provide a sufficient number of calibrators, at appropriate levels across the measurement range, to permit the use of flexible and biologically representative models—ideally five-parameter logistic (5PL) regression (74). However, current commercial assays often lack adequate calibrator points, limiting the ability to adopt such models and constraining calibration accuracy (section 8.5.2.2).

A further important consideration identified in this study is the influence of including a zero calibrator in calibration models. While conceptually attractive, inclusion of a blank calibrator artificially constrains the lower asymptote of the calibration curve, adversely affecting the resolution and uncertainty at the clinically critical low end of the ADAMTS13 measurement range (74). This practice can directly impact result interpretation around decision thresholds such as 10 IU/dL, with the potential to increase patient misclassification risk.

The aggregation of calibration data across multiple assay runs provided a novel and practical refinement of MU estimation (section 8.5.2). By pooling data from twelve consecutive Technozym® ADAMTS13 ELISA calibration curves and applying Monte Carlo simulation, the study incorporated

both within-run and between-run variability into a unified calibration model. This approach, extending previous work applied to coagulation factor assay calibration (150), demonstrated that aggregated calibration curves produced more reliable and clinically meaningful uncertainty estimates than single-run models and better reflected long-term assay performance.

Evaluation of the HemosIL AcuStar[®] rapid assay revealed substantial variability in master calibration adjustments across five reagent lots. While baseline calibration remained relatively stable, local lot-specific systematic bias was introduced requiring correction, particularly at the clinically critical decision limit of 10 IU/dL (section 8.6.2.1). This finding suggests incorporating bias uncertainty (ubias) into MU estimates and highlighted the potential for lot-to-lot variability to influence patient classification may be required.

An important distinction that emerged from this analysis was the difference in uncertainty contributors between IQC data and patient result interpolation. IQC materials may be analysed in singlicate while patient samples are run in replicate and reported as mean values, creating a mismatch in analytical process and associated uncertainty contributors. Consequently, the calibration process may contribute differently to IQC-derived MU and patient result uncertainty, necessitating separate evaluation of these components.

Collectively, these studies demonstrated that MU estimation in ADAMTS13 activity assays cannot rely on simplistic or uncritical application of top-down approaches as is commonplace in other disciplines (148, 149, 151, 253). The validity of IQC-based uncertainty estimates depends on recognising the statistical structure of IQC data, the influence of replicate measurement design, and the distinct uncertainty contributors introduced by calibration modelling. This further extends the existing movement to more complex assessment of IQC data for MU estimation (148, 154). It also requires recognition of the responsibilities of assay manufacturers in providing appropriate calibration models (if built into software) and materials to support uncertainty estimation.

Future work should focus on formalising guidance for appropriate IQC data analysis and calibration modelling in MU estimation, developing accessible tools to support advanced statistical techniques such as variance components analysis and time series decomposition, and engaging with professional bodies and manufacturers to promote the adoption of these practices. Ultimately, ensuring the clinical appropriateness of laboratory results depends on the accurate, transparent, and statistically robust estimation of uncertainty across all stages of the analytical process.

The final study presented in this thesis addressed a fundamental challenge in the clinical application of ADAMTS13 activity testing: how to interpret results accurately and consistently in the context of analytical uncertainty and clinical risk. While considerable effort has been made in laboratory medicine to standardise analytical processes and quantify MU, much less attention has been paid to how uncertainty information is incorporated into clinical interpretation frameworks. This study sought to bridge that gap, developing and evaluating advanced statistical methods for interpreting ADAMTS13 activity results that explicitly account for uncertainty and align with clinical decision-making processes.

A central focus of this work was the application of likelihood ratios and their graphical representation as a tool to contextualise ADAMTS13 activity results (section 9.3, 9.4 and 9.5). Likelihood ratios offer a clinically intuitive and standardised means of assessing the weight of evidence provided by a test result, reflecting how much more (or less) likely a result is to be observed in patients with TTP compared to those without (275, 276). By incorporating MU-derived guard bands into LR calculations, the study demonstrated a patient-centred approach to result interpretation that moves beyond the rigid application of fixed decision limits.

The development of graphical LR plots represented a significant advancement in this regard. These plots provided a visual framework for clinicians to understand how individual results, particularly those falling near clinical decision thresholds, contribute to diagnostic certainty. Importantly, the inclusion of uncertainty bands around the LR curve communicated the degree of confidence

associated with specific results, supporting nuanced clinical interpretation rather than binary classification.

Likelihood ratios help link laboratory results to clinical decisions. They provide a consistent way to interpret results across different methods and platforms (277-279)

Unlike positive/negative cut-offs that may vary by method, likelihood ratios use a continuous scale. This allows test results to be understood in the same way, regardless of how the test is done (277, 280, 281). It improves how results are communicated and helps support clinical decision-making. But these ratios must be based on local data. Manufacturer or published values may not apply to a specific lab's patients or testing process.

To create valid likelihood ratio curves, local verification studies are needed. These must include real patient samples and produce a receiver operating characteristic (ROC) curve (276, 282). The ROC curve depends on the test performance and the local patient population.

The study also clarified the intended use and scope of the LR⁺ and MU integration framework developed. It was conceptualised as a theoretical and exploratory model, designed to demonstrate how incorporating MU into LR interpretation can enhance diagnostic transparency and provide a more nuanced understanding of assay performance. The approach was not intended for immediate clinical implementation without further work, including engagement with laboratory users and clinical stakeholders, prospective evaluation of its impact on clinical decision-making, and development of standardised reporting formats. In particular, the practical implications of communicating LR⁺ ranges in laboratory reports would require careful consideration to ensure clarity, prevent misinterpretation, and support effective clinical use.

The goal is not to replace routine positive/negative reporting but to improve interpretation where uncertainty matters most—especially around the diagnostic cut-off. For results well above or below

this threshold, adding a likelihood ratio range adds little. But for borderline results, showing an LR⁺ range helps clinicians see the diagnostic strength more clearly and make better decisions.

For example, a result of 10 IU/dL with an MU of ± 2 IU/dL gives an LR⁺ range of ~ 7.3 – 19.3 . An LR⁺ of ~ 7 gives moderate evidence, while ~ 20 gives strong evidence, potentially triggering treatment without more tests (276, 282). That gap creates real ambiguity, especially near the cut-off. Displaying the LR⁺ range makes that uncertainty visible and supports transparent decision-making.

The derivation of a novel analytical performance specification framework allowed Bayesian interpretation of individual results, in the context of MU, using local verification data and local requesting patterns to inform the clinician of the degree of uncertainty in the post test probability. This is a novel approach to APS derivation and allows a fluid interpretation of any method, and its associated uncertainty, stratified here by pre test probability based on validated clinical scoring systems. The differences in interpretation shown here based on assay and clinical integrated knowledge is a step forward in the application of MU in a clinical setting. These methods support risk-based decisions in haemostasis tests, where precision matters most. Other assays like D-dimer or factor assays around known decision limits could also benefit. But when tests serve multiple purposes—for example, factor VIII for haemophilia diagnosis and therapy monitoring—interpretive frameworks must adapt to each context.

10.3 Integration of Findings and Contribution to Knowledge

The studies presented in this thesis collectively address several longstanding challenges in the measurement, verification, and clinical interpretation of ADAMTS13 activity assays, providing a coherent body of work that advances both the scientific and practical understanding of assay performance in haemostasis laboratory medicine. Each individual study explored a distinct aspect of

the analytical and interpretative process, yet when considered together, they provide an integrated and clinically meaningful framework for the evaluation and reporting of complex laboratory assays.

A key contribution of this thesis has been the identification and characterisation of analytical variability inherent in ADAMTS13 activity measurement across different assay platforms. Significant variability exists at clinically critical decision points, particularly at low activity levels (sections 4.3, 5.3, 5.4, 5.5). These findings highlighted the limitations of conventional verification studies, which often rely solely on correlation analysis and fail to adequately evaluate clinical implications of assay differences. This work underscored the need for more rigorous, clinically informed verification strategies and provided the foundation for the subsequent studies.

The evaluation of utility specific APS addressed a fundamental gap in current guidance and practice. By critically appraising the limitations of biological variation-derived APS models and introducing simulation-based modelling linked to patient misclassification risk, the study demonstrated the value of a clinically driven approach to APS derivation. This methodology moved beyond theoretical statistical models and instead aligned APS with real-world clinical decision-making processes, providing laboratories with meaningful criteria against which to evaluate assay performance. Additionally, applying local specific criteria for reference intervals, identifying distinct clinical categories, and personalised reference ranges for the first time allows a patient centred interpretation of longitudinal monitoring.

Building on this, the second chapter introduced a structured, risk-based framework for method comparison and clinical risk evaluation. By developing and applying an ADAMTS13-specific error grid, the study provided a novel tool for quantifying the clinical impact of assay differences and explicitly linking analytical variability to patient risk (section 5.5). This approach advanced the field by moving beyond simplistic summary statistics, such as correlation coefficients, towards clinically meaningful metrics of comparability and patient safety.

This analytical focus was extended in chapters 6, 7 and 8 by critically evaluating the estimation of MU using IQC data and calibration modelling strategies. Together, these studies demonstrated that while top-down approaches to uncertainty estimation offer operational convenience, they must be applied with statistical rigour and an understanding of underlying data structures. The novel application of time series analysis to IQC data provided important insight into the contribution of transient biases to overall uncertainty and addressed a topical debate within the laboratory medicine community regarding the treatment of small, systematic variations. Furthermore, the evaluation of calibration modelling strategies and the demonstration of differences in uncertainty contributors between IQC data and patient result interpolation provided important clarification on the limitations of conventional MU estimation approaches, particularly in assays employing complex calibration procedures.

Finally, the critical question of how uncertainty information should be incorporated into clinical interpretation frameworks was presented. By developing and evaluating advanced statistical methods including the integration of likelihood ratios the study provided a patient-centred, clinically meaningful approach to result interpretation. Importantly, this work moved beyond uncertainty quantification to address how uncertainty information can be communicated effectively to clinicians and incorporated into risk-based decision-making processes.

Taken together, the findings of this thesis represent a significant contribution to knowledge in the field of laboratory medicine, specifically in the domain of haemostasis assay evaluation and interpretation. The research advances the application of MU principles from a narrow, technical exercise to a broader, clinically integrated framework that links analytical performance to patient outcomes and risk management.

Furthermore, the methodological innovations developed throughout this work—including the use of simulation-based APS derivation, error grid frameworks, time series decomposition of IQC data, calibration model aggregation, and LR-based result interpretation—are transferable beyond

ADAMTS13 activity testing. These approaches have broad applicability to other complex haemostasis assays and diagnostic tests characterised by clinical decision thresholds and analytical variability. In particular, the findings have direct relevance to factor VIII and factor IX activity assays, D-dimer testing, and other assays where accurate, transparent communication of result uncertainty is essential to patient safety and clinical decision-making.

Beyond the specific methodological contributions, this thesis also identifies important gaps in current guidance, standards, and laboratory practice. It highlights the need for improved guidance on the appropriate use of IQC data in MU estimation, clearer expectations regarding preliminary MU estimation and long-term performance specifications, and the systematic incorporation of uncertainty information into laboratory reporting and clinical communication. The work also underscores the critical role of assay manufacturers in providing appropriate calibration models and materials, and the responsibility of laboratories to undertake robust, locally informed verification and validation of assay performance.

In summary, the integrated body of work presented in this thesis offers a comprehensive, clinically meaningful, and practically applicable framework for the evaluation, verification, and interpretation of complex haemostasis assays. It addresses longstanding methodological challenges, advances the field's understanding of MU, and provides a pathway towards more transparent, patient-centred laboratory practice.

10.4 Recommendations for Future Research and Practice

The body of work presented in this thesis has identified several areas where further research and development are warranted to improve the evaluation, verification, and interpretation of complex haemostasis assays, including ADAMTS13 activity measurement. While the studies conducted have made significant contributions to the field, they have also highlighted methodological limitations,

gaps in current practice, and opportunities to advance laboratory medicine and patient care further.

This section outlines key recommendations for future research and practical implementation.

10.4.11. Validation and Clinical Implementation of Risk-Based Interpretative

Frameworks

A primary recommendation arising from this thesis is the need for formal validation of the advanced interpretative frameworks developed, particularly the integration of likelihood ratios with MU, RCVs, and personalised reference intervals. While this thesis has demonstrated the theoretical utility of these approaches, prospective studies are required to evaluate their practical impact on clinical decision-making, diagnostic accuracy, and patient outcomes.

Future research should include engagement with laboratory users and clinical stakeholders to assess the usability and interpretive clarity of LR⁺ range reporting and uncertainty-informed result communication. Additionally, implementation studies are needed to determine how these frameworks can be effectively integrated into laboratory information systems and reporting processes without introducing unnecessary complexity or cognitive burden for clinicians.

10.4.22. Development of Condition-Specific and Indication-Specific

Interpretative Models

The research presented in this thesis has primarily focused on the application of advanced interpretative strategies in the context of iTTP. However, as the clinical use of ADAMTS13 activity assays expands into other thrombotic and inflammatory conditions, the relevance and applicability of the interpretative frameworks developed will require further evaluation.

Future work should focus on developing condition-specific and indication-specific LR models, RCV thresholds, and prRI boundaries. Particular attention should be paid to the challenges posed by assays with multiple clinical utilities. For example, factor VIII activity assays are used both to diagnose

haemophilia A and to monitor treatment efficacy. The clinical significance of analytical variability differs between these contexts, necessitating distinct verification processes and interpretative frameworks. Systematic assessment of how interpretative strategies can be standardised across different clinical utilities is essential.

10.4.33. Standardisation and Oversight of Verification Study Design

This thesis has demonstrated the critical importance of robust verification study design in supporting meaningful MU estimation and interpretative framework development. In particular, the need for locally conducted verification studies, based on consecutively collected and representative samples, was highlighted as essential for the valid derivation of ROC curves and LR estimates.

Future research should explore the development of standardised protocols for verification studies, ensuring that laboratories undertake verification processes that capture relevant population characteristics, reflect local clinical practice, and account for potential shifts in testing patterns following the implementation of new assay methodologies. Additionally, increased oversight by regulatory bodies and professional organisations is recommended to ensure that verification studies are conducted with sufficient scientific rigour and transparency.

10.4.44. Clarification of Guidance on MU Estimation

The findings of this thesis have highlighted gaps in current guidance concerning the estimation and application of MU in laboratory practice. Specifically, there is a need for clearer expectations regarding the relationship between preliminary MU estimates derived from IQC data and long-term performance specifications. Current guidance does not adequately address the limitations of preliminary MU estimation, particularly in the context of complex assays employing duplicate measurements, calibration interpolation, and long-term procedural variability.

Professional bodies and standards organisations should seek to formalise guidance on appropriate IQC data analysis, incorporating stratified statistical approaches, variance components analysis, and

time series decomposition techniques. Further research is recommended to develop accessible tools and resources to support laboratories in adopting these advanced methodologies.

10.4.55. Manufacturer Responsibilities and Calibration Modelling

The thesis has demonstrated that calibration model selection is a critical determinant of MU estimation accuracy and clinical reliability. Manufacturers of haemostasis assays should be required to provide calibration materials and models that reflect the biological behaviour of the measurand and support robust uncertainty estimation. This includes providing a sufficient number of calibrators at appropriate levels to permit the use of flexible calibration models, such as five-parameter logistic (5PL) regression, where appropriate.

Future collaboration between manufacturers, laboratories, and regulatory bodies is recommended to establish minimum requirements for calibration model validation, documentation of calibration uncertainty, and ongoing monitoring of calibration performance.

10.4.66. Broader Application of Methodological Innovations

The methodological innovations developed in this thesis, including simulation-based APS derivation, error grid frameworks, time series decomposition of IQC data, calibration model aggregation, and LR-based result interpretation, have broad applicability beyond ADAMTS13 activity testing. Future research should explore the transferability of these approaches to other haemostasis assays and diagnostic tests characterised by complex analytical performance and defined clinical decision limits.

In particular, assays such as factor VIII and factor IX activity tests, D-dimer testing, and emerging biomarker assays would benefit from the application of risk-based verification, uncertainty estimation, and interpretative frameworks developed in this thesis. Prospective studies to validate these methodologies in diverse assay contexts are recommended.

10.4.77. Enhancing Transparency and Communication of Uncertainty

Finally, this thesis has highlighted the need for laboratories to systematically incorporate uncertainty information into result reporting and clinical communication. Current practice often omits or obscures MU data, limiting the ability of clinicians to make fully informed decisions. Future work should focus on developing standardised reporting formats and communication strategies that effectively convey uncertainty information without overburdening or confusing end users.

Engagement with clinical stakeholders will be essential to ensure that uncertainty information is presented in a way that supports, rather than hinders, clinical decision-making. Professional bodies and accreditation organisations should consider incorporating explicit expectations regarding uncertainty communication into laboratory quality standards.

10.4.8 Conclusion

And finally, returning to the clinical vignette in section 1.1, “what’s the likelihood of the result being less than 10?” The answer is, as is often the case – it depends. It depends on the performance of the method, both in implementation and in clinical use. It depends on how that performance is monitored, and against what standard it is determined to be performing appropriately. All of these things are determined locally, and not through universally ordained standards. Most importantly, this thesis has shown that in fact this is the wrong question anyway, using the methods described here for the first time, and applying these novel approaches at an individual patient level, we can, for the first time, provide an answer to what is actually being asked:

“What does this result mean?”

11 References

1. Zheng XL, Vesely SK, Cataland SR, Coppo P, Geldziler B, Iorio A, et al. ISTH guidelines for treatment of thrombotic thrombocytopenic purpura. *J Thromb Haemost*. 2020;18(10):2496-502.
2. Scully M, Hunt BJ, Benjamin S, Liesner R, Rose P, Peyvandi F, et al. Guidelines on the diagnosis and management of thrombotic thrombocytopenic purpura and other thrombotic microangiopathies. *Br J Haematol*. 2012;158(3):323-35.
3. Scully M, Rayment R, Clark A, Westwood JP, Cranfield T, Gooding R, et al. A British Society for Haematology Guideline: Diagnosis and management of thrombotic thrombocytopenic purpura and thrombotic microangiopathies. *Br J Haematol*. 2023;203(4):546-63.
4. Arnold DM, Patriquin CJ, Nazy I. Thrombotic microangiopathies: a general approach to diagnosis and management. *Cmaj*. 2017;189(4):E153-e9.
5. Coppo P, Schwarzwinger M, Buffet M, Wynckel A, Clabault K, Presne C, et al. Predictive features of severe acquired ADAMTS13 deficiency in idiopathic thrombotic microangiopathies: the French TMA reference center experience. *PLoS One*. 2010;5(4):e10208.
6. Zhao N, Zhou L, Hu X, Sun G, Chen C, Fan X, et al. A modified PLASMIC score including the lactate dehydrogenase/the upper limit of normal ratio more accurately identifies Chinese thrombotic thrombocytopenic purpura patients than the original PLASMIC score. *J Clin Apher*. 2020;35(2):79-85.
7. Wells PS, Anderson DR, Rodger M, Forgie M, Kearon C, Dreyer J, et al. Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. *N Engl J Med*. 2003;349(13):1227-35.
8. Lo GK, Juhl D, Warkentin TE, Sigouin CS, Eichler P, Greinacher A. Evaluation of pretest clinical score (4 T's) for the diagnosis of heparin-induced thrombocytopenia in two clinical settings. *J Thromb Haemost*. 2006;4(4):759-65.
9. Bendapudi PK, Hurwitz S, Fry A, Marques MB, Waldo SW, Li A, et al. Derivation and external validation of the PLASMIC score for rapid assessment of adults with thrombotic microangiopathies: a cohort study. *Lancet Haematol*. 2017;4(4):e157-e64.
10. Tufano A, Polimeno M, Matani B, Cardillo G, Capasso F, Mormile R, et al. Validation of PLASMIC score in a cohort of patients with suspected thrombotic microangiopathy in an academic medical centre. *Blood Transfus*. 2023;21(4):345-9.
11. Gupta GK, Hendrickson JE, Tormey CA. Application of PLASMIC score in prediction of ADAMTS13 deficiency in a pediatric case of acquired thrombotic thrombocytopenic purpura. *J Clin Apher*. 2020;35(2):140-1.
12. Miller DD, Krenzer JA, Kenkre VP, Rose WN. Sequential Immune Thrombocytopenia (ITP) and Thrombotic Thrombocytopenic Purpura (TTP) in an Elderly Male Patient with Primary Sjogren's Syndrome: When in Doubt, Use the PLASMIC Score. *Case Rep Med*. 2021;2021:6869342.
13. Lee CH, Huang YC, Li SS, Hsu YT, Chen YP, Chen TY. Application of PLASMIC Score in Risk Prediction of Thrombotic Thrombocytopenic Purpura: Real-World Experience From a Tertiary Medical Center in Taiwan. *Front Med (Lausanne)*. 2022;9:893273.
14. Lee JA, Lin MH, Kang CM, Chuang MK, Fung CKB, Lo SC. A validation and modification of PLASMIC score by adjusting the criteria of mean corpuscular volume and international normalized ratio. *J Clin Apher*. 2023;38(5):582-9.
15. Bentley MJ, Lehman CM, Blaylock RC, Wilson AR, Rodgers GM. The utility of patient characteristics in predicting severe ADAMTS13 deficiency and response to plasma exchange. *Transfusion*. 2010;50(8):1654-64.
16. Nishimura N, Yoshimoto K, Yada N, Kakiwaki A, Sawa A, Senzaki S, et al. The Combination of the Lactate Dehydrogenase/Hemoglobin Ratio with the PLASMIC Score

Facilitates Differentiation of TTP from Septic DIC Without Identification of Schistocytes. *Clin Appl Thromb Hemost*. 2023;29:10760296231207629.

17. Moschcowitz E. An acute febrile pleiochromic anemia with hyaline thrombosis of the terminal arterioles and capillaries. An undescribed disease. *Thromb Haemost*. 1978;40(1):4-8.
18. Lämmle B, Vanhoorelbeke K, Kremer Hovinga JA, Knöbl P. 100 Years of Thrombotic Thrombocytopenic Purpura: A Story of Death and Life. *Hamostaseologie*. 2024;44(1):59-73.
19. Matsumoto M, Miyakawa Y, Kokame K, Ueda Y, Wada H, Higasa S, et al. Diagnostic and treatment guidelines for thrombotic thrombocytopenic purpura (TTP) in Japan 2023. *International Journal of Hematology*. 2023:1-18.
20. Sukumar S, Gavrilaki E, Chaturvedi S. Updates on thrombotic thrombocytopenic purpura: Recent developments in pathogenesis, treatment and survivorship. *Thrombosis Update*. 2021;5:100062.
21. Sartain S, Shubert S, Wu MF, Srivaths P, Teruya J, Krance R, et al. Therapeutic Plasma Exchange does not Improve Renal Function in Hematopoietic Stem Cell Transplantation-Associated Thrombotic Microangiopathy: An Institutional Experience. *Biol Blood Marrow Transplant*. 2019;25(1):157-62.
22. Long B, Bridwell RE, Manchanda S, Gottlieb M. Evaluation and Management of Thrombotic Thrombocytopenic Purpura in the Emergency Department. *J Emerg Med*. 2021;61(6):674-82.
23. Thomas MR, Robinson S, Scully MA. How we manage thrombotic microangiopathies in pregnancy. *Br J Haematol*. 2016;173(6):821-30.
24. Kangro K, Roose E, Schelpe AS, Tellier E, Kaplanski G, Voorberg J, et al. Generation and validation of small ADAMTS13 fragments for epitope mapping of anti-ADAMTS13 autoantibodies in immune-mediated thrombotic thrombocytopenic purpura. *Res Pract Thromb Haemost*. 2020;4(5):918-30.
25. Ferrari S, Palavra K, Gruber B, Kremer Hovinga JA, Knobl P, Caron C, et al. Persistence of circulating ADAMTS13-specific immune complexes in patients with acquired thrombotic thrombocytopenic purpura. *Haematologica*. 2014;99(4):779-87.
26. Bendapudi PK, Foy BH, Mueller SB, Liu J, Feingold LM, Burke KE, et al. Recombinant ADAMTS13 for Immune Thrombotic Thrombocytopenic Purpura. *N Engl J Med*. 2024;390(18):1690-8.
27. Lämmle B. ADAMTS13 recovery in caplacizumab-treated TTP. *Blood*. 2024;143(18):1785-6.
28. Graça NAG, Joly BS, Voorberg J, Vanhoorelbeke K, Béranger N, Veyradier A, et al. TTP: From empiricism for an enigmatic disease to targeted molecular therapies. *Br J Haematol*. 2022;197(2):156-70.
29. Wang W, He Y. [Research Advance in Treatment of Thrombotic Thrombocytopenic Purpura --Review]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi*. 2022;30(1):314-8.
30. Cataland SR, Coppo P, Scully M, Lämmle B. Thrombotic thrombocytopenic purpura: 100 years of research on Moschcowitz syndrome. *Blood*. 2024;144(11):1143-52.
31. Halkidis K, Meng C, Liu S, Mayne L, Siegel DL, Zheng XL. Mechanisms of inhibition of human monoclonal antibodies in immune thrombotic thrombocytopenic purpura. *Blood*. 2023;141(24):2993-3005.
32. Halkidis K, Lämmle B, Zheng XL. The history of thrombotic thrombocytopenic purpura research: a narrative review. *Ann Blood*. 2024;9.
33. Zheng XL, Vesely SK, Cataland SR, Coppo P, Geldziler B, Iorio A, et al. ISTH guidelines for the diagnosis of thrombotic thrombocytopenic purpura. *J Thromb Haemost*. 2020;18(10):2486-95.
34. Matsumoto M, Miyakawa Y, Kokame K, Ueda Y, Wada H, Higasa S, et al. Diagnostic and treatment guidelines for thrombotic thrombocytopenic purpura (TTP) in Japan 2023. *Int J Hematol*. 2023;118(5):529-46.

35. Kubo M, Matsumoto M. Frontiers in pathophysiology and management of thrombotic thrombocytopenic purpura. *Int J Hematol.* 2023;117(3):331-40.
36. Peyvandi F, Scully M, Kremer Hovinga JA, Cataland S, Knöbl P, Wu H, et al. Caplacizumab for Acquired Thrombotic Thrombocytopenic Purpura. *N Engl J Med.* 2016;374(6):511-22.
37. Scully M, Cataland SR, Peyvandi F, Coppo P, Knöbl P, Kremer Hovinga JA, et al. Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *N Engl J Med.* 2019;380(4):335-46.
38. Liu W, Xiao Y. [Advance in the diagnosis and treatment of hereditary thrombotic thrombocytopenic purpura]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2022;39(4):442-6.
39. Zhou W, Inada M, Lee TP, Benten D, Lyubsky S, Bouhassira EE, et al. ADAMTS13 is expressed in hepatic stellate cells. *Lab Invest.* 2005;85(6):780-8.
40. Zheng X, Chung D, Takayama TK, Majerus EM, Sadler JE, Fujikawa K. Structure of von Willebrand factor-cleaving protease (ADAMTS13), a metalloprotease involved in thrombotic thrombocytopenic purpura. *J Biol Chem.* 2001;276(44):41059-63.
41. Plaimauer B, Zimmermann K, Volkel D, Antoine G, Kerschbaumer R, Jenab P, et al. Cloning, expression, and functional characterization of the von Willebrand factor-cleaving protease (ADAMTS13). *Blood.* 2002;100(10):3626-32.
42. Zheng XL. Structure-function and regulation of ADAMTS-13 protease. *J Thromb Haemost.* 2013;11 Suppl 1(0 1):11-23.
43. DeYoung V, Singh K, Kretz CA. Mechanisms of ADAMTS13 regulation. *J Thromb Haemost.* 2022;20(12):2722-32.
44. Furlan M, Robles R, Lämmle B. Partial purification and characterization of a protease from human plasma cleaving von Willebrand factor to fragments produced by in vivo proteolysis. *Blood.* 1996;87(10):4223-34.
45. Tsai HM. Physiologic cleavage of von Willebrand factor by a plasma protease is dependent on its conformation and requires calcium ion. *Blood.* 1996;87(10):4235-44.
46. de Groot R, Bardhan A, Ramroop N, Lane DA, Crawley JT. Essential role of the disintegrin-like domain in ADAMTS13 function. *Blood.* 2009;113(22):5609-16.
47. Dong JF. Structural and functional correlation of ADAMTS13. *Curr Opin Hematol.* 2007;14(3):270-6.
48. South K, Luken BM, Crawley JT, Phillips R, Thomas M, Collins RF, et al. Conformational activation of ADAMTS13. *Proc Natl Acad Sci U S A.* 2014;111(52):18578-83.
49. Springer TA. Biology and physics of von Willebrand factor concatamers. *J Thromb Haemost.* 2011;9 Suppl 1(0 1):130-43.
50. Gogia S, Neelamegham S. Role of fluid shear stress in regulating VWF structure, function and related blood disorders. *Biorheology.* 2015;52(5-6):319-35.
51. Lynch CJ, Cawte AD, Millar CM, Rueda D, Lane DA. A common mechanism by which type 2A von Willebrand disease mutations enhance ADAMTS13 proteolysis revealed with a von Willebrand factor A2 domain FRET construct. *PLoS One.* 2017;12(11):e0188405.
52. Crawley JT, de Groot R, Xiang Y, Luken BM, Lane DA. Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor. *Blood.* 2011;118(12):3212-21.
53. Yu S, Liu W, Fang J, Shi X, Wu J, Fang Y, et al. AFM Imaging Reveals Multiple Conformational States of ADAMTS13. *J Biol Eng.* 2019;13:9.
54. Sakai K, Matsumoto M, De Waele L, Dekimpe C, Hamada E, Kubo M, et al. ADAMTS13 conformation and immunoprofiles in Japanese patients with immune-mediated thrombotic thrombocytopenic purpura. *Blood Adv.* 2023;7(1):131-40.
55. Joly BS, Roose E, Coppo P, Vanhoorelbeke K, Veyradier A. ADAMTS13 conformation is closed in non-immune acquired thrombotic thrombocytopenic purpura of unidentified pathophysiology. *Haematologica.* 2023;108(2):638-44.

56. Halkidis K, Zheng XL. ADAMTS13 conformations and mechanism of inhibition in immune thrombotic thrombocytopenic purpura. *J Thromb Haemost.* 2022;20(10):2197-203.
57. Lin J, Ding X, Yang P, Liu S, Li Q, Cruz MA, et al. Force-induced biphasic regulation of VWF cleavage by ADAMTS13. *Thromb Res.* 2023;229:99-106.
58. Shin Y, Akiyama M, Kokame K, Soejima K, Miyata T. Binding of von Willebrand factor cleaving protease ADAMTS13 to Lys-plasmin(ogen). *J Biochem.* 2012;152(3):251-8.
59. Pillai VG, Zheng XL. A novel mechanism underlying allosteric regulation of ADAMTS-13 revealed by hydrogen-deuterium exchange plus mass spectrometry. *Res Pract Thromb Haemost.* 2023;7(1):100012.
60. Kokame K, Nobe Y, Kokubo Y, Okayama A, Miyata T. FRETs-VWF73, a first fluorogenic substrate for ADAMTS13 assay. *Br J Haematol.* 2005;129(1):93-100.
61. Muia J, Gao W, Haberichter SL, Dolatshahi L, Zhu J, Westfield LA, et al. An optimized fluorogenic ADAMTS13 assay with increased sensitivity for the investigation of patients with thrombotic thrombocytopenic purpura. *J Thromb Haemost.* 2013;11(8):1511-8.
62. Graça NAG, Ercig B, Carolina Velásquez Pereira L, Kangro K, Kaijen P, Nicolaes GAF, et al. Modifying ADAMTS13 to modulate binding of pathogenic autoantibodies of patients with acquired thrombotic thrombocytopenic purpura. *Haematologica.* 2020;105(11):2619-30.
63. Moore GW, Meijer D, Griffiths M, Rushen L, Brown A, Budde U, et al. A multi-center evaluation of TECHNOSCREEN(®) ADAMTS-13 activity assay as a screening tool for detecting deficiency of ADAMTS-13. *J Thromb Haemost.* 2020;18(7):1686-94.
64. Stephenson J, Chapman K, Mohammed S, Zebeljan D, Ahuja M, Donikian D, et al. A multicenter evaluation of the Technoscreen ADAMTS13 activity semi-quantitative screening test for thrombotic thrombocytopenic purpura diagnosis and exclusion. *Int J Lab Hematol.* 2023;45(4):562-70.
65. Hubbard AR, Heath AB, Kremer Hovinga JA, Factor SovW. Establishment of the WHO 1st International Standard ADAMTS13, plasma (12/252): communication from the SSC of the ISTH. *J Thromb Haemost.* 2015;13(6):1151-3.
66. Padoan A, Sciacovelli L, Aita A, Antonelli G, Plebani M. Measurement uncertainty in laboratory reports: A tool for improving the interpretation of test results. *Clin Biochem.* 2018;57:41-7.
67. Plebani M, Sciacovelli L, Bernardi D, Aita A, Antonelli G, Padoan A. What information on measurement uncertainty should be communicated to clinicians, and how? *Clin Biochem.* 2018;57:18-22.
68. Gantzer ML, Miller WG. Harmonisation of Measurement Procedures: how do we get it done? *Clin Biochem Rev.* 2012;33(3):95-100.
69. Meyer SC, Sulzer I, Lämmle B, Kremer Hovinga JA. Hyperbilirubinemia interferes with ADAMTS-13 activity measurement by FRETs-VWF73 assay: diagnostic relevance in patients suffering from acute thrombotic microangiopathies. *J Thromb Haemost.* 2007;5(4):866-7.
70. Zhou Z, Han H, Cruz MA, López JA, Dong JF, Guchhait P. Haemoglobin blocks von Willebrand factor proteolysis by ADAMTS-13: a mechanism associated with sickle cell disease. *Thromb Haemost.* 2009;101(6):1070-7.
71. Mancini I, Valsecchi C, Lotta LA, Deforche L, Pontiggia S, Bajetta M, et al. FRETs-VWF73 rather than CBA assay reflects ADAMTS13 proteolytic activity in acquired thrombotic thrombocytopenic purpura patients. *Thromb Haemost.* 2014;112(2):297-303.
72. MacDonald S. Measurement uncertainty in the VWF-ADAMTS13 axis: the challenges of adding a calculated measurand in the haemostasis laboratory. *Journal of Laboratory and Precision Medicine.* 2022;8.
73. Kilpatrick ES, Sandberg S. An overview of EFLM harmonization activities in Europe. *Clinical Chemistry and Laboratory Medicine (CCLM).* 2018;56(10):1591-7.
74. Wild D. The immunoassay handbook: theory and applications of ligand binding, ELISA and related techniques: Newnes; 2013.

75. McPherson PAC. Approaches to nonlinear curve fitting in laboratory medicine. *Lab Med*. 2024;55(2):111-6.
76. Schneppenheim R, Kremer Hovinga JA, Becker T, Budde U, Karpman D, Brockhaus W, et al. A common origin of the 4143insA ADAMTS13 mutation. *Thromb Haemost*. 2006;96(1):3-6.
77. Camilleri RS, Scully M, Thomas M, Mackie IJ, Liesner R, Chen WJ, et al. A phenotype-genotype correlation of ADAMTS13 mutations in congenital thrombotic thrombocytopenic purpura patients treated in the United Kingdom. *J Thromb Haemost*. 2012;10(9):1792-801.
78. Joly BS, Coppo P, Veyradier A. Pediatric thrombotic thrombocytopenic purpura. *Eur J Haematol*. 2018;101(4):425-34.
79. Klaus C, Plaimauer B, Studt JD, Dorner F, Lammle B, Mannucci PM, et al. Epitope mapping of ADAMTS13 autoantibodies in acquired thrombotic thrombocytopenic purpura. *Blood*. 2004;103(12):4514-9.
80. Peyvandi F, Ferrari S, Lavoretano S, Canciani MT, Mannucci PM. von Willebrand factor cleaving protease (ADAMTS-13) and ADAMTS-13 neutralizing autoantibodies in 100 patients with thrombotic thrombocytopenic purpura. *Br J Haematol*. 2004;127(4):433-9.
81. Thomas MR, de Groot R, Scully MA, Crawley JT. Pathogenicity of Anti-ADAMTS13 Autoantibodies in Acquired Thrombotic Thrombocytopenic Purpura. *EBioMedicine*. 2015;2(8):942-52.
82. Falter T, Rossmann H, de Waele L, Dekimpe C, von Auer C, Müller-Calleja N, et al. A novel von Willebrand factor multimer ratio as marker of disease activity in thrombotic thrombocytopenic purpura. *Blood Adv*. 2023;7(17):5091-102.
83. De Waele L, Sakai K, Mancini I, Sinkovits G, Falter T, Inoue T, et al. Open ADAMTS-13 conformation index predicts earlier relapse in immune-mediated thrombotic thrombocytopenic purpura. *J Thromb Haemost*. 2024;22(2):493-502.
84. De Waele L, Curie A, Kangro K, Tellier E, Kaplanski G, Männik A, et al. Anti-cysteine/spacer antibodies that open ADAMTS13 are a common feature in iTTP. *Blood Adv*. 2021;5(21):4480-4.
85. Prasannan N, Dragunaite B, Subhan M, Thomas M, de Groot R, Singh D, et al. Peak ADAMTS13 activity to assess ADAMTS13 conformation and risk of relapse in immune-mediated thrombotic thrombocytopenic purpura. *Blood*. 2024;143(25):2644-53.
86. Horvath AR, Bell KJL, Ceriotti F, Jones GRD, Loh TP, Lord S, et al. Outcome-based analytical performance specifications: current status and future challenges. *Clin Chem Lab Med*. 2024;62(8):1474-82.
87. Narayan SK, Gudivada KK, Sivakoti S, Krishna B. Outcomes of patients with thrombotic thrombocytopenic purpura treated in an intensive care unit. *Natl Med J India*. 2023;36(5):295-300.
88. Du P, Cristarella T, Goyer C, Moride Y. A Systematic Review of the Epidemiology and Disease Burden of Congenital and Immune-Mediated Thrombotic Thrombocytopenic Purpura. *J Blood Med*. 2024;15:363-86.
89. Sukumar S, Mazepa MA, Chaturvedi S. Cardiovascular Disease and Stroke in Immune TTP-Challenges and Opportunities. *J Clin Med*. 2023;12(18).
90. Soares Ferreira Junior A, Pinheiro Maux Lessa M, Boyle SH, Sanborn K, Kuchibhatla M, Onwuemene OA. In patients with suspected immune TTP, admission source impacts hospital length of stay and time to therapeutic plasma exchange impacts clinical outcomes. *Thromb Res*. 2023;227:34-9.
91. Jaan A, Sarfraz Z, Khalid F, Anwar J. Impact of Concomitant Thrombotic Thrombocytopenic Purpura on COVID-19 Mortality and Morbidity: A Nationwide Inpatient Sample Analysis. *Clin Appl Thromb Hemost*. 2023;29:10760296231219252.
92. Rodeghiero F, Stasi R, Gernsheimer T, Michel M, Provan D, Arnold DM, et al. Standardization of terminology, definitions and outcome criteria in immune thrombocytopenic

- purpura of adults and children: report from an international working group. *Blood*. 2009;113(11):2386-93.
93. Cuker A, Cataland SR, Coppo P, de la Rubia J, Friedman KD, George JN, et al. Redefining outcomes in immune TTP: an international working group consensus report. *Blood*. 2021;137(14):1855-61.
 94. Sakai K, Fujimura Y, Miyata T, Isonishi A, Kokame K, Matsumoto M. Current prophylactic plasma infusion protocols do not adequately prevent long-term cumulative organ damage in the Japanese congenital thrombotic thrombocytopenic purpura cohort. *Br J Haematol*. 2021;194(2):444-52.
 95. Sakai K, Matsumoto M. Clinical Manifestations, Current and Future Therapy, and Long-Term Outcomes in Congenital Thrombotic Thrombocytopenic Purpura. *J Clin Med*. 2023;12(10).
 96. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med*. 2015;53(6):833-5.
 97. Braga F, Panteghini M. Performance specifications for measurement uncertainty of common biochemical measurands according to Milan models. *Clin Chem Lab Med*. 2021.
 98. Christensen H, Lauterlein JJ, Sørensen PD, Petersen ER, Madsen JS, Brandslund I. Home management of oral anticoagulation via telemedicine versus conventional hospital-based treatment. *Telemed J E Health*. 2011;17(3):169-76.
 99. Garcia-Alamino JM, Ward AM, Alonso-Coello P, Perera R, Bankhead C, Fitzmaurice D, et al. Self-monitoring and self-management of oral anticoagulation. *Cochrane Database Syst Rev*. 2010(4):CD003839.
 100. Çubukçu HC, Vanstapel F, Thelen M, van Schrojenstein Lantman M, Bernabeu-Andreu FA, Meško Brguljan P, et al. APS calculator: a data-driven tool for setting outcome-based analytical performance specifications for measurement uncertainty using specific clinical requirements and population data. *Clin Chem Lab Med*. 2024;62(4):597-607.
 101. Sandberg S, Zima T, Panteghini M. Analytical performance specifications - moving from models to practical recommendations. *Clin Chem Lab Med*. 2024;62(8):1451-4.
 102. Badrick T, Theodorsson E. Six Sigma - is it time to re-evaluate its value in laboratory medicine? *Clin Chem Lab Med*. 2024;62(12):2398-400.
 103. Panteghini M, Sandberg S. Total error vs. measurement uncertainty: the match continues. *Clin Chem Lab Med*. 2016;54(2):195-6.
 104. Sandberg S, Coskun A, Carobene A, Fernandez-Calle P, Diaz-Garzon J, Bartlett WA, et al. Analytical performance specifications based on biological variation data - considerations, strengths and limitations. *Clin Chem Lab Med*. 2024;62(8):1483-9.
 105. Kim HJ, Xu Y, Petri A, Vanhoorelbeke K, Crawley JTB, Emsley J. Crystal structure of ADAMTS13 CUB domains reveals their role in global latency. *Sci Adv*. 2021;7(16).
 106. Crawley JT, Lam JK, Rance JB, Mollica LR, O'Donnell JS, Lane DA. Proteolytic inactivation of ADAMTS13 by thrombin and plasmin. *Blood*. 2005;105(3):1085-93.
 107. Madarati H, DeYoung V, Singh K, Sparring T, Kwong AC, Fredenburgh JC, et al. Optimization of plasma-based BioID identifies plasminogen as a ligand of ADAMTS13. *Sci Rep*. 2024;14(1):9073.
 108. Kilercik M, Coskun A, Serteser M, Inan D, Unsal I. Biological variations of ADAMTS13 and von Willebrand factor in human adults. *Biochem Med (Zagreb)*. 2014;24(1):138-45.
 109. Aarsand AK, Roraas T, Fernandez-Calle P, Ricos C, Diaz-Garzon J, Jonker N, et al. The Biological Variation Data Critical Appraisal Checklist: A Standard for Evaluating Studies on Biological Variation. *Clin Chem*. 2018;64(3):501-14.
 110. Fraser CG. Biological variation: from principles to practice: Amer. Assoc. for Clinical Chemistry; 2001.

111. Gowans EM, Hyltoft Petersen P, Blaabjerg O, Hørder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest*. 1988;48(8):757-64.
112. de Maat MP, van Schie M, Kluft C, Leebeek FW, Meijer P. Biological Variation of Hemostasis Variables in Thrombosis and Bleeding: Consequences for Performance Specifications. *Clin Chem*. 2016;62(12):1639-46.
113. Fraser CG. Reference change values. *Clin Chem Lab Med*. 2011;50(5):807-12.
114. Braga F, Pasqualetti S, Borrillo F, Capoferri A, Chibireva M, Rovegno L, et al. Definition and application of performance specifications for measurement uncertainty of 23 common laboratory tests: linking theory to daily practice. *Clin Chem Lab Med*. 2023;61(2):213-23.
115. Borrillo F, Panteghini M. Current performance of C-reactive protein determination and derivation of quality specifications for its measurement uncertainty. *Clin Chem Lab Med*. 2023;61(9):1552-7.
116. Borrillo F, Panteghini M. State-of-the-art model for derivation of analytical performance specifications: how to define the highest level of analytical performance technically achievable. *Clin Chem Lab Med*. 2024;62(8):1490-6.
117. Pasqualetti S, Chibireva M, Borrillo F, Braga F, Panteghini M. Improving measurement uncertainty of plasma electrolytes: a complex but not impossible task. *Clin Chem Lab Med*. 2020.
118. Braga F, Panteghini M. Defining permissible limits for the combined uncertainty budget in the implementation of metrological traceability. *Clin Biochem*. 2018;57:7-11.
119. Rovegno L, Civera E, Infusino I, Panteghini M. State of the art of measurement uncertainty of serum ferritin. *Clin Chem Lab Med*. 2024;62(1):e6-e8.
120. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. *Clin Chem Lab Med*. 2017;55(2):189-94.
121. Panteghini M. A Comment about Analytical Performance Specifications for the Combined Measurement Uncertainty Budget in the Implementation of Metrological Traceability of Parathyroid Hormone. *Clin Chem*. 2024;70(2):456-7.
122. Panteghini M. Analytical performance specifications for combined uncertainty budget in the implementation of metrological traceability. *Clin Chem Lab Med*. 2024;62(8):1497-504.
123. Valsecchi C, Mirabet M, Mancini I, Biganzoli M, Schiavone L, Faraudo S, et al. Evaluation of a New, Rapid, Fully Automated Assay for the Measurement of ADAMTS13 Activity. *Thromb Haemost*. 2019;119(11):1767-72.
124. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255-68.
125. Martin RF. General deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clin Chem*. 2000;46(1):100-4.
126. Bilić-Zulle L. Comparison of methods: Passing and Bablok regression. *Biochem Med (Zagreb)*. 2011;21(1):49-52.
127. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society Series D: The Statistician*. 1983;32(3):307-17.
128. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141-51.
129. Clarke WL. The original Clarke Error Grid Analysis (EGA). *Diabetes Technol Ther*. 2005;7(5):776-9.
130. Clarke WL, Anderson S, Farhy L, Breton M, Gonder-Frederick L, Cox D, et al. Evaluating the clinical accuracy of two continuous glucose sensors using continuous glucose-error grid analysis. *Diabetes Care*. 2005;28(10):2412-7.

131. Clarke WL, Gonder-Frederick L, Cox D, Kovatchev B. A critical appraisal of the continuous glucose-error grid analysis: response to Wentholt et al. *Diabetes Care*. 2007;30(2):449-50; author reply 50-1.
132. Clarke WL, Anderson S, Kovatchev B. Evaluating clinical accuracy of continuous glucose monitoring systems: Continuous Glucose-Error Grid Analysis (CG-EGA). *Curr Diabetes Rev*. 2008;4(3):193-9.
133. Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care*. 2000;23(8):1143-8.
134. Barrett BJ, Fardy JM. Evaluation of Diagnostic Tests. *Methods Mol Biol*. 2021;2249:319-33.
135. Habibzadeh F, Habibzadeh P. The likelihood ratio and its graphical representation. *Biochem Med (Zagreb)*. 2019;29(2):020101.
136. Petersen PH, Sandberg S, Iglesias N, Sölétormos G, Aarsand AK, Brandslund I, et al. 'Likelihood-ratio' and 'odds' applied to monitoring of patients as a supplement to 'reference change value' (RCV). *Clin Chem Lab Med*. 2008;46(2):157-64.
137. Deane KD, Van Hoovels L, Joy VE, Olschowka N, Bossuyt X. From autoantibody test results to decision making: Incorporating likelihood ratios within medical practice. *Autoimmun Rev*. 2024;23(5):103537.
138. Yong J, MacDonald S, Downey C, Fretwell R, Lawrence C, Murphy P, et al. Practical Considerations for the Use of the Rapid AcuStar[®] ADAMTS13 Activity Assay in the Diagnosis of Acute Thrombotic Thrombocytopenic Purpura (TTP). *J Clin Med*. 2024;13(15).
139. Fierz W. Likelihood ratios of quantitative laboratory results in medical diagnosis: The application of Bézier curves in ROC analysis. *PLoS One*. 2018;13(2):e0192420.
140. De Bièvre P. The 2012 international vocabulary of metrology: "VIM". *Chemistry International--Newsmagazine for IUPAC*. 2012;34(3):26-7.
141. 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;57:56-61.
142. Rigo-Bonnin R, Alia P, Canalias F. Measurement uncertainty and metrological traceability of whole blood cyclosporin A mass concentration results obtained by UHPLC-MS/MS. *Clin Chem Lab Med*. 2018;56(9):1458-68.
143. Lim YK, Kweon OJ, Lee MK, Kim B, Kim HR. Top-down and bottom-up approaches for the estimation of measurement uncertainty in coagulation assays. *Clin Chem Lab Med*. 2020;58(9):1525-33.
144. Lim YK, Kweon OJ, Lee MK, Kim HR. Estimation of Measurement Uncertainty of Factor Assays Using the Monte Carlo Simulation. *Am J Clin Pathol*. 2021;156(5):717-21.
145. Braga F, Infusino I, Panteghini M. Role and Responsibilities of Laboratory Medicine Specialists in the Verification OF Metrological Traceability of. *J Med Biochem*. 2015;34(3):282-7.
146. Panteghini M, Braga F. Implementation of metrological traceability in laboratory medicine: where we are and what is missing. *Clin Chem Lab Med*. 2020;58(8):1200-4.
147. Panteghini M. Documenting and validating metrological traceability of serum alanine aminotransferase measurements: a priority for medical laboratory community for providing high quality service in hepatology. *Clin Chem Lab Med*. 2024;62(2):249-52.
148. Panteghini M. Redesigning the surveillance of in vitro diagnostic medical devices and of medical laboratory performance by quality control in the traceability era. *Clin Chem Lab Med*. 2023;61(5):759-68.
149. Panteghini M, Krintus M. Establishing, evaluating and monitoring analytical quality in the traceability era. *Crit Rev Clin Lab Sci*. 2025;62(3):148-81.
150. Lim YK, Kweon OJ, Lee MK, Kim HR. Estimation of Measurement Uncertainty of Factor Assays Using the Monte Carlo Simulation. *Am J Clin Pathol*. 2021.

151. Aloisio E, Carnevale A, Pasqualetti S, Birindelli S, Dolci A, Panteghini M. Random uncertainty of photometric determination of hemolysis index on the Abbott Architect c16000 platform. *Clin Biochem*. 2018;57:62-4.
152. Panteghini M. The simple reproducibility of a measurement result does not equal its overall measurement uncertainty. *Clin Chem Lab Med*. 2022;60(10):e221-e2.
153. Ishikawa K, Loftus JH. Introduction to quality control. 3rd ed. Tokyo: 3A Corporation; 1990.
154. Thelen MHM, van Schrojenstein Lantman M. When bias becomes part of imprecision: how to use analytical performance specifications to determine acceptability of lot-lot variation and other sources of possibly unacceptable bias. *Clin Chem Lab Med*. 2024;62(8):1505-11.
155. Farrance I, Frenkel R. Measurement uncertainty and the importance of correlation. *Clin Chem Lab Med*. 2020;59(1):7-9.
156. Braga F, Infusino I, Panteghini M. Performance criteria for combined uncertainty budget in the implementation of metrological traceability. *Clin Chem Lab Med*. 2015;53(6):905-12.
157. MacDonald S. Measurement Uncertainty in the Hemostasis Laboratory. *Semin Thromb Hemost*. 2025.
158. Joint Committee for Guides in Metrology Jre. International vocabulary of metrology | Basic and general concepts and associated terms (VIM). 2012; 3.
159. Farrance I, Frenkel R. Uncertainty of Measurement: A Review of the Rules for Calculating Uncertainty Components through Functional Relationships. *Clin Biochem Rev*. 2012;33(2):49-75.
160. (CLSI) CLSI. Expression of MU in Laboratory Medicine Approved Guideline. Clinical Laboratory Standards Institute (CLSI); 2012.
161. Farrance I, Frenkel R, Badrick T. ISO/TS 20914:2019 - a critical commentary. *Clin Chem Lab Med*. 2020;58(8):1182-90.
162. Chantarangkul V, van den Besselaar AM, Witteveen E, Tripodi A. International collaborative study for the calibration of a proposed international standard for thromboplastin, rabbit, plain. *J Thromb Haemost*. 2006;4(6):1339-45.
163. data-end="270" data-start="220">International Organization for Standardization td-c-ssd-ed-ssd-ed-std-c-sm. Medical laboratories — Requirements for quality and competence. <td data-col-size="sm" data-end="425" data-start="409"><strong data-end="424" data-start="411"><td data-col-size="md" data-end="432" data-start="425">ISO; 2022.
164. Dorgalaleh A, Favaloro EJ, Bahraini M, Rad F. Standardization of Prothrombin Time/International Normalized Ratio (PT/INR). *Int J Lab Hematol*. 2021;43(1):21-8.
165. Critchfield GC, Bennett ST. The influence of the reference mean prothrombin time on the international normalized ratio. *Am J Clin Pathol*. 1994;102(6):806-11.
166. Critchfield GC, Bennett ST. The international normalized ratio and uncertainty. Validation of a probabilistic model. *Am J Clin Pathol*. 1994;102(1):115-22.
167. Bowyer AE, Gosselin RC. Factor VIII and Factor IX Activity Measurements for Hemophilia Diagnosis and Related Treatments. *Semin Thromb Hemost*. 2023;49(6):609-20.
168. Mourey G, Tachon G, Pellequer JL, Zawadzki C, Trossaert M, Bertrand MA, et al. A Haut-Doubs FVII variant depending on species-derived-thromboplastin reagent (F7:p.Arg337His). *Haemophilia*. 2014;20(4):e347-9.
169. (ISO) IOFS. ISO/TS 20914:2019 - Medical Laboratories. Practical guidance for the estimation of measurement uncertainty. British Standards Institute 2019.
170. Farrance I, Frenkel R. Uncertainty in measurement: a review of monte carlo simulation using microsoft excel for the calculation of uncertainties through functional relationships, including uncertainties in empirically derived constants. *Clin Biochem Rev*. 2014;35(1):37-61.
171. Rigo-Bonnin R, Canalias F. Measurement uncertainty estimation for derived biological quantities. *Clin Chem Lab Med*. 2020;59(1):e1-e7.

172. BIPM I, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data | Guide to the expression of uncertainty

in measurement. Joint Committee for Guides in Metrology,.

https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6 . Joint Committee for Guides in Metrology, JCGM 2008.

173. Biedermann JS, van den Besselaar AM, de Maat MP, Leebeek FW, Kruip MJ. Monitoring of treatment with vitamin K antagonists: recombinant thromboplastins are more sensitive to factor VII than tissue-extract thromboplastins. *J Thromb Haemost*. 2017;15(3):500-6.

174. van den Besselaar AM, Witteveen E, van der Meer FJ. Uncertainty of international sensitivity index and international normalized ratio. *J Thromb Haemost*. 2013;11(8):1615-7.

175. van den Besselaar A, Stavelin A, Kitchen S, Bryant M, Tripodi A, Scalabrino E, et al. Defining a metrologically traceable and sustainable calibration hierarchy of international normalized ratio for monitoring of vitamin K antagonist treatment in accordance with International Organization for Standardization (ISO) 17511:2020 standard: communication from the International Federation of Clinical Chemistry and Laboratory Medicine-SSC/ISTH working group on prothrombin time/international normalized ratio standardization. *J Thromb Haemost*. 2024;22(4):1236-48.

176. Taberner DA, Poller L, Thomson JM, Darby KV. Effect of international sensitivity index (ISI) of thromboplastins on precision of international normalised ratios (INR). *J Clin Pathol*. 1989;42(1):92-6.

177. Kirkwood TB. Calibration of reference thromboplastins and standardisation of the prothrombin time ratio. *Thromb Haemost*. 1983;49(3):238-44.

178. Lim YK, Kweon OJ, Lee MK, Kim HR. Estimating the measurement uncertainties of the international sensitivity index of 12 thromboplastins through Monte Carlo simulation. *Thromb Res*. 2023;224:32-7.

179. Infusino I, Panteghini M. Measurement uncertainty: Friend or foe? *Clin Biochem*. 2018;57:3-6.

180. Aarsand AK, Roraas T, Bartlett WA, Coskun A, Carobene A, Fernandez-Calle P, et al. Harmonization initiatives in the generation, reporting and application of biological variation data. *Clin Chem Lab Med*. 2018;56(10):1629-36.

181. Panteghini M. Traceability, reference systems and result comparability. *Clin Biochem Rev*. 2007;28(3):97-104.

182. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. *Clin Biochem*. 2009;42(4-5):236-40.

183. Lee JH, Choi JH, 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;53(7):1025-32.

184. Favaloro E. Uncertainties in Measurement of Uncertainty Measurements: A Commentary

from the Perspective of a Specialist Haemostasis Laboratory.

<https://search.informit.org/doi/10.3316/informit.182099435319554>: Australian Journal of Medical Science; 2006. p. 72-83.

185. Panteghini M, Camara JE, Delatour V, Van Uytvanghe K, Vesper HW, Zhang T. Feasibility of Metrological Traceability Implementation Using the Joint Committee on Traceability in Laboratory Medicine Database Entries Including the Fulfillment of "Fit-for-Purpose" Maximum Allowable Measurement Uncertainty. *Clin Chem*. 2024;70(11):1321-33.

186. Zhang HP, Zhou WB, Li CB, Du ZL, Peng MT. [Evaluation and characterization of the certified reference materials for coagulation factor VIII and IX activity testing]. *Zhonghua Yi Xue Za Zhi*. 2016;96(20):1613-7.

187. Guy S, Sermon-Cadd AM, Shepherd FM, Kitchen S, Bowyer AE. A cost-effective approach to factor assay calibration using a truncated live calibration curve. *Int J Lab Hematol*. 2019;41(5):679-83.
188. Panteghini M. Not all biases are created equal: how to deal with bias on laboratory measurements. *Clin Chem Lab Med*. 2025;63(5):916-22.
189. Meijer P, de Maat MP, Kluft C, Haverkate F, van Houwelingen HC. Long-term analytical performance of hemostasis field methods as assessed by evaluation of the results of an external quality assessment program for antithrombin. *Clin Chem*. 2002;48(7):1011-5.
190. Qin Y, Zhou R, Wang W, Yin H, Yang Y, Yue Y, et al. Uncertainty evaluation in clinical chemistry, immunoassay, hematology and coagulation analytes using only external quality assessment data. *Clin Chem Lab Med*. 2018;56(9):1447-57.
191. St Ledger K, Feussner A, Kalina U, Horn C, Metzner HJ, Bensen-Kennedy D, et al. International comparative field study evaluating the assay performance of AFSTYLA in plasma samples at clinical hemostasis laboratories. *J Thromb Haemost*. 2018;16(3):555-64.
192. Peyvandi F, Oldenburg J, Friedman KD. A critical appraisal of one-stage and chromogenic assays of factor VIII activity. *J Thromb Haemost*. 2016;14(2):248-61.
193. Cui M, Xu L, Wang H, Ju S, Xu S, Jing R. Combining Nordtest method and bootstrap resampling for measurement uncertainty estimation of hematology analytes in a medical laboratory. *Clin Biochem*. 2017;50(18):1067-72.
194. Favaloro EJ, Pasalic L. Should multiple factor dilutions be performed for all patient coagulation factor assays? Let the debate begin! *Res Pract Thromb Haemost*. 2022;6(2):e12689.
195. Cui M, Xu L, Wang H, Ju S, Xu S, Jing R. Combining Nordtest method and bootstrap resampling for measurement uncertainty estimation of hematology analytes in a medical laboratory. *Clinical Biochemistry*. 2017;50(18):1067-72.
196. Veen AMHvd, Cox MG. Getting started with uncertainty evaluation using the Monte Carlo method in R.
197. BIPM I, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data | Supplement 1 to the \Guide to the expression of uncertainty in measurement" | Propagation of distributions using a Monte Carlo method.
https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c .2008.
198. Favaloro EJ, Hamdam S, McDonald J, McVicker W, Ule V. Time to think outside the box? Prothrombin time, international normalised ratio, international sensitivity index, mean normal prothrombin time and measurement of uncertainty: a novel approach to standardisation. *Pathology*. 2008;40(3):277-87.
199. Frenkel RB, Farrance I. Uncertainty in Measurement: Procedures for Determining Uncertainty With Application to Clinical Laboratory Calculations. *Adv Clin Chem*. 2018;85:149-211.
200. Baker P, Platton S, Gibson C, Gray E, Jennings I, Murphy P, et al. Guidelines on the laboratory aspects of assays used in haemostasis and thrombosis. *Br J Haematol*. 2020;191(3):347-62.
201. Balık AR, Gücel F. Evaluation of 20 clinical chemistry and 12 immunoassay analytes in terms of total analytical error and measurement uncertainty. *Scand J Clin Lab Invest*. 2021;81(7):517-22.
202. Farrance I, Badrick T, Sikaris KA. Uncertainty in measurement and total error - are they so incompatible? *Clin Chem Lab Med*. 2016;54(8):1309-11.
203. Bayat H, Westgard SA, Westgard JO. Clinical chemistry and laboratory medicine: The value of Sigma-metrics in laboratory medicine. 2024.

204. Ammer T, Schützenmeister A, Prokosch H-U, Rauh M, Rank CM, Zierk J. refineR: a novel algorithm for reference interval estimation from real-world data. *Scientific reports*. 2021;11(1):16023.
205. Sandberg S, Carobene A, Bartlett B, Coskun A, Fernandez-Calle P, Jonker N, et al. Biological variation: recent development and future challenges. *Clin Chem Lab Med*. 2023;61(5):741-50.
206. Scully M, Antun A, Cataland SR, Coppo P, Dossier C, Biebuyck N, et al. Recombinant ADAMTS13 in Congenital Thrombotic Thrombocytopenic Purpura. *N Engl J Med*. 2024;390(17):1584-96.
207. Scully M, Berger C, Biebuyck N, Boix SB, Dossier C, Hassenpflug W-A, et al. Recombinant ADAMTS13 Prophylaxis in Pediatric Patients with Congenital Thrombotic Thrombocytopenic Purpura: Results from the Phase 3 Randomized, Crossover Study and the Phase 3b Continuation Study. *Blood*. 2024;144:141.
208. Carobene A, Aarsand AK, Bartlett WA, Coskun A, Diaz-Garzon J, Fernandez-Calle P, et al. The European Biological Variation Study (EuBIVAS): a summary report. *Clin Chem Lab Med*. 2022;60(4):505-17.
209. Sandberg S, Carobene A, Aarsand AK. Biological variation - eight years after the 1st Strategic Conference of EFLM. *Clin Chem Lab Med*. 2022;60(4):465-8.
210. Avnon T, Rabinovich A, Pikovsky O, Pariente G, Erez O. Recombinant a disintegrin-like and metalloproteinase with thrombospondin type 1 motifs 13 (ADAMTS-13) for acute and prophylactic treatment of congenital thrombotic thrombocytopenic purpura in pregnancy. *J Thromb Haemost*. 2025.
211. Cai P, Luo H, Xu H, Zhu X, Xu W, Dai Y, et al. Recombinant ADAMTS 13 Attenuates Brain Injury After Intracerebral Hemorrhage. *Stroke*. 2015;46(9):2647-53.
212. Colliou É, Ribes A, Gaible C, Marlas M, Ribes D, Labadens I, et al. Prophylactic use of recombinant ADAMTS-13 during pregnancy for congenital thrombotic thrombocytopenic purpura. *Res Pract Thromb Haemost*. 2025;9(1):102687.
213. Cattaneo D, Panteghini M. Analytical performance specifications for measurement uncertainty in therapeutic monitoring of immunosuppressive drugs. *Clin Chem Lab Med*. 2024;62(4):e81-e3.
214. Fraser CG. Desirable standards of performance for therapeutic drug monitoring. *Clin Chem*. 1987;33(3):387-9.
215. Prasannan N, Thomas M, Stubbs M, Westwood JP, de Groot R, Singh D, et al. Delayed normalization of ADAMTS13 activity in acute thrombotic thrombocytopenic purpura in the caplacizumab era. *Blood*. 2023;141(18):2206-13.
216. Coşkun A, Sandberg S, Unsal I, Cavusoglu C, Serteser M, Kilercik M, et al. Personalized and Population-Based Reference Intervals for 48 Common Clinical Chemistry and Hematology Measurands: A Comparative Study. *Clin Chem*. 2023;69(9):1009-30.
217. Coşkun A, Sandberg S, Unsal I, Cavusoglu C, Serteser M, Kilercik M, et al. Personalized Reference Intervals in Laboratory Medicine: A New Model Based on Within-Subject Biological Variation. *Clin Chem*. 2021;67(2):374-84.
218. Coskun A, Sandberg S, Unsal I, Yavuz FG, Cavusoglu C, Serteser M, et al. Personalized reference intervals - statistical approaches and considerations. *Clin Chem Lab Med*. 2022;60(4):629-35.
219. Coskun A, Sandberg S, Unsal I, Serteser M, Aarsand AK. Personalized reference intervals: from theory to practice. *Crit Rev Clin Lab Sci*. 2022;59(7):501-16.
220. Bugdayci G, Oguzman H, Arattan HY, Sasmaz G. The use of reference change values in clinical laboratories. *Clin Lab*. 2015;61(3-4):251-7.
221. Haeckel R, Carobene A, Wosniok W. Problems with estimating reference change values (critical differences). *Clin Chim Acta*. 2021;523:437-40.

222. Lund F, Petersen PH, Fraser CG, Sölétormos G. Calculation of limits for significant unidirectional changes in two or more serial results of a biomarker based on a computer simulation model. *Ann Clin Biochem.* 2015;52(Pt 2):237-44.
223. Lund F, Petersen PH, Fraser CG, Sölétormos G. Different percentages of false-positive results obtained using five methods for the calculation of reference change values based on simulated normal and ln-normal distributions of data. *Ann Clin Biochem.* 2016;53(6):692-8.
224. Kwak H, Kim MS, Kim S, Kim J, Aoki Y, Chung SJ, et al. Kinetic modeling of the plasma pharmacokinetic profiles of ADAMTS13 fragment and its Fc-fusion counterpart in mice. *Front Pharmacol.* 2024;15:1352842.
225. Aarsand AK, Díaz-Garzón J, Fernandez-Calle P, Guerra E, Locatelli M, Bartlett WA, et al. The EuBIVAS: Within- and Between-Subject Biological Variation Data for Electrolytes, Lipids, Urea, Uric Acid, Total Protein, Total Bilirubin, Direct Bilirubin, and Glucose. *Clin Chem.* 2018;64(9):1380-93.
226. Çubukçu HC. Computer simulation approaches to evaluate the interaction between analytical performance characteristics and clinical (mis)classification: a complementary tool for setting indirect outcome-based analytical performance specifications. *Clin Chem Lab Med.* 2025;63(7):1292-300.
227. Bendapudi PK, Li A, Hamdan A, Uhl L, Kaufman R, Stowell C, et al. Impact of severe ADAMTS13 deficiency on clinical presentation and outcomes in patients with thrombotic microangiopathies: the experience of the Harvard TMA Research Collaborative. *Br J Haematol.* 2015;171(5):836-44.
228. Li A, Khalighi PR, Wu Q, Garcia DA. External validation of the PLASMIC score: a clinical prediction tool for thrombotic thrombocytopenic purpura diagnosis and treatment. *J Thromb Haemost.* 2018;16(1):164-9.
229. Oosterhuis WP, Coskun A, Sandberg S, Theodorsson E. Performance specifications for sodium should not be based on biological variation. *Clin Chim Acta.* 2023;540:117221.
230. Fokkema MR, Herrmann Z, Muskiet FA, Moecks J. Reference change values for brain natriuretic peptides revisited. *Clin Chem.* 52. England2006. p. 1602-3.
231. Sölétormos G, Hyltoft Petersen P, Dombernowsky P. Progression criteria for cancer antigen 15.3 and carcinoembryonic antigen in metastatic breast cancer compared by computer simulation of marker data. *Clin Chem.* 2000;46(7):939-49.
232. Røraas T, Støve B, Petersen PH, Sandberg S. Biological Variation: The Effect of Different Distributions on Estimated Within-Person Variation and Reference Change Values. *Clin Chem.* 2016;62(5):725-36.
233. Petersen PH, Fraser CG, Sandberg S, Goldschmidt H. The index of individuality is often a misinterpreted quantity characteristic. *Clin Chem Lab Med.* 1999;37(6):655-61.
234. Petersen PH, Sandberg S, Fraser CG, Goldschmidt H. Influence of index of individuality on false positives in repeated sampling from healthy individuals. *Clin Chem Lab Med.* 2001;39(2):160-5.
235. Walton RM. Subject-based reference values: biological variation, individuality, and reference change values. *Vet Clin Pathol.* 2012;41(2):175-81.
236. Coşkun A, Sandberg S, Unsal I, Topcu DI, Aarsand AK. Reference Intervals Revisited: A Novel Model for Population-Based Reference Intervals, Using a Small Sample Size and Biological Variation Data. *Clin Chem.* 2024;70(10):1279-90.
237. Siest G, Henny J, Gräsbeck R, Wilding P, Petitclerc C, Queraltó JM, et al. The theory of reference values: an unfinished symphony. *Clin Chem Lab Med.* 2013;51(1):47-64.
238. Dimopoulos K, Philips M, Goetze JP. Diagnosis of Thrombotic Thrombocytopenic Purpura by ADAMTS13 Activity Quantification. *J Appl Lab Med.* 2022;7(3):637-49.
239. Pascual C, Nieto JM, Fidalgo T, Seguí IG, Díaz-Ricart M, Docampo MF, et al. Multicentric evaluation of the new HemosIL Acustar. *Int J Lab Hematol.* 2021;43(3):485-93.

240. Beranger N, Benghezal S, Joly BS, Capdenat S, Delton A, Stepanian A, et al. Diagnosis and follow-up of thrombotic thrombocytopenic purpura with an automated chemiluminescent ADAMTS13 activity immunoassay. *Res Pract Thromb Haemost*. 2021;5(1):81-93.
241. Deshpande SR, Tarawneh H, Deitelzweig CC, Tong J, Zhou T, Chen Y, et al. Rapid ADAMTS13 activity assays for thrombotic thrombocytopenic purpura: a systematic review and meta-analysis. *Blood*. 2025.
242. Westgard JO. Use and interpretation of common statistical tests in method comparison studies. *Clinical chemistry*. 2008;54(3):612-.
243. Dunn G, Roberts C. Modelling method comparison data. *Stat Methods Med Res*. 1999;8(2):161-79.
244. Payne RB. Method comparison: evaluation of least squares, Deming and Passing/Bablok regression procedures using computer simulation. *Ann Clin Biochem*. 1997;34 (Pt 3):319-20.
245. Stöckl D, Dewitte K, Thienpont LM. Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data? *Clin Chem*. 1998;44(11):2340-6.
246. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. *Clinical chemistry*. 1979;25(3):432-8.
247. Payne RB. Deming's regression analysis in method comparison studies. *Ann Clin Biochem*. 1985;22 (Pt 4):430.
248. Deming WE. Statistical adjustment of data. 1943.
249. Konings H. Use of Deming regression in method-comparison studies. *Surv Immunol Res*. 1982;1(4):371-4.
250. MacDonald S, Murphy P, Fretwell R, Downey C, Baker P, Foxton E, et al. Abstract. *Research and Practice in Thrombosis and Haemostasis*. 2022;6.
251. Favaloro EJ, Mohammed S, Chapman K, Swanepoel P, Zebeljan D, Sefhore O, et al. A multicenter laboratory assessment of a new automated chemiluminescent assay for ADAMTS13 activity. *J Thromb Haemost*. 2021;19(2):417-28.
252. Irsara C, Anliker M, Egger AE, Harasser L, Lhotta K, Feistritz C, et al. Evaluation of two fully automated ADAMTS13 activity assays in comparison to manual FRET assay. *Int J Lab Hematol*. 2023;45(5):758-65.
253. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories. *Clinical Chemistry and Laboratory Medicine (CCLM)*. 2020;58(9):1407-13.
254. Shumway RH, Stoffer DS. Time series analysis and its applications: Springer; 2000.
255. Montgomery DC, Jennings CL, Kulahci M. Introduction to time series analysis and forecasting: John Wiley & Sons; 2015.
256. Chatfield C, Xing H. The analysis of time series: an introduction with R: Chapman and hall/CRC; 2019.
257. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*. 2018;13(3):e0194889.
258. Plebani M, Lippi G. Standardization and harmonization in laboratory medicine: not only for clinical chemistry measurands. *Clin Chem Lab Med*. 61. Germany2023. p. 185-7.
259. Ceriotti F, Barhanovic NG, Kostovska I, Kotaska K, Perich Alsina MC. Harmonisation of the laboratory testing process: need for a coordinated approach. *Clin Chem Lab Med*. 2016;54(12):e361-e3.
260. Bowyer A, Gray E, Lowe A, Murphy P, Platton S, Riddell A, et al. Laboratory coagulation tests and recombinant porcine factor VIII: A United Kingdom Haemophilia Centre Doctors' Organisation guideline. *Haemophilia*. 2022;28(3):515-9.
261. Gray E, Kitchen S, Bowyer A, Chowdary P, Jenkins PV, Murphy P, et al. Laboratory measurement of factor replacement therapies in the treatment of congenital haemophilia: A United Kingdom Haemophilia Centre Doctors' Organisation guideline. *Haemophilia*. 2020;26(1):6-16.

262. Jenkins PV, Bowyer A, Burgess C, Gray E, Kitchen S, Murphy P, et al. Laboratory coagulation tests and emicizumab treatment A United Kingdom Haemophilia Centre Doctors' Organisation guideline. *Haemophilia*. 2020;26(1):151-5.
263. Ceriotti F. Harmonization Initiatives in Europe. *EJIFCC*. 2016;27(1):23-9.
264. Banfi G, Del Fabbro M. Biological variation in tests of hemostasis. *Semin Thromb Hemost*. 2008;34(7):635-41.
265. Kopic A, Benamara K, Piskernik C, Plaimauer B, Horling F, Höbarth G, et al. Preclinical assessment of a new recombinant ADAMTS-13 drug product (BAX930) for the treatment of thrombotic thrombocytopenic purpura. *J Thromb Haemost*. 2016;14(7):1410-9.
266. Sarode R. Recombinant ADAMTS-13: goodbye, allergic reactions! *Blood*. 2017;130(19):2045-6.
267. Kwak H, Choi G, Kim S, Park JM, Kwon Y, Lee Y, et al. GC1126A, a novel ADAMTS13 mutein, evades autoantibodies in immune-mediated thrombotic thrombocytopenic purpura. *Sci Rep*. 2025;15(1):1613.
268. Cornes M, Simundic AM, Cadamuro J, Costelloe SJ, Baird G, Kristensen GBB, et al. The CRESS checklist for reporting stability studies: on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group for the Preanalytical Phase (WG-PRE). *Clin Chem Lab Med*. 2020;59(1):59-69.
269. Cornes M, Vermeersch P, Šimundić AM, Von Meyer A, Šálek T, Meyer B, et al. The final part of the CRESS trilogy - how to evaluate the quality of stability studies. *Clin Chem Lab Med*. 2024;62(11):2128-39.
270. Langley K, Fretwell R, Kitchen S, MacDonald S, Dutt T, Baker P, et al. Multiple centre evaluation study of ADAMTS13 activity and inhibitor assays. *Int J Lab Hematol*. 2018;40(1):21-5.
271. Ludbrook J. Comparing methods of measurements. *Clin Exp Pharmacol Physiol*. 1997;24(2):193-203.
272. Singh D, Subhan MO, de Groot R, Vanhoorelbeke K, Zadvydaite A, Dragūnaitė B, et al. ADAMTS13 activity testing: evaluation of commercial platforms for diagnosis and monitoring of thrombotic thrombocytopenic purpura. *Res Pract Thromb Haemost*. 2023;7(2):100108.
273. Pipe S, Sadeghi-Khomami A, Konkole BA, Kitchen S, Negrier C, Liu M, et al. A global comparative field study to evaluate the factor VIII activity of efanesoctocog alfa by one-stage clotting and chromogenic substrate assays at clinical haemostasis laboratories. *Haemophilia*. 2024;30(1):214-23.
274. Hollestelle MJ, Ruinemans-Koerts J, Idema RN, Meijer P, de Maat MPM. Determination of sigma score based on biological variation for haemostasis assays: fit-for-purpose for daily practice? *Clin Chem Lab Med*. 2019;57(8):1235-41.
275. Arkin CF, Greco FA. Likelihood ratio and positive predictive value. *Arch Pathol Lab Med*. 1987;111(2):111.
276. Caraguel CGB, Colling A. Diagnostic likelihood ratio - the next-generation of diagnostic test accuracy measurement. *Rev Sci Tech*. 2021;40(1):299-309.
277. Simel DL, Samsa GP, Matchar DB. Likelihood ratios for continuous test results--making the clinicians' job easier or harder? *J Clin Epidemiol*. 1993;46(1):85-93.
278. Simel DL, Easter J, Tomlinson G. Likelihood ratios, sensitivity, and specificity values can be back-calculated when the odds ratios are known. *J Clin Epidemiol*. 2013;66(4):458-60.
279. Carey M, Brennan J, Ciconelli R, Dayan P, Eskin B, Fabris L, et al. Likelihood ratios when the results of diagnostic tests are continuous variables. *Am J Emerg Med*. 16. United States 1998. p. 722-3.
280. Bossuyt X, Van Hoovels L. Reporting Likelihood Ratio in Autoimmune Serology: Just a Theory or Feasible in Practice? *Clin Chem*. 2024;70(3):482-5.
281. Fierz W, Bossuyt X. Likelihood Ratio Approach and Clinical Interpretation of Laboratory Tests. *Front Immunol*. 2021;12:655262.
282. Cebul RD. The likelihood ratio and diagnostic tests. *Ann Intern Med*. 1986;104(6):888-9.

283. Pascual C, Nieto JM, Fidalgo T, Segui IG, Diaz-Ricart M, Docampo MF, et al. Multicentric evaluation of the new HemosIL Acustar((R)) chemiluminescence ADAMTS13 activity assay. *Int J Lab Hematol*. 2021;43(3):485-93.
284. Stratmann J, Ward JN, Miesbach W. Evaluation of a rapid turn-over, fully-automated ADAMTS13 activity assay: a method comparison study. *J Thromb Thrombolysis*. 2020;50(3):628-31.

12 Appendices

12.1 Appendix A - Comparison of measurand definitions of

ADAMTS13 activity assays

This appendix presents a summary comparison of measurand definitions used in different commercial ADAMTS13 activity assays. It highlights key inconsistencies in how each method defines and reports activity, based on information provided in manufacturers' Instructions for Use (IFUs). These discrepancies have implications for inter-method comparability and interpretation, as discussed in the main thesis.

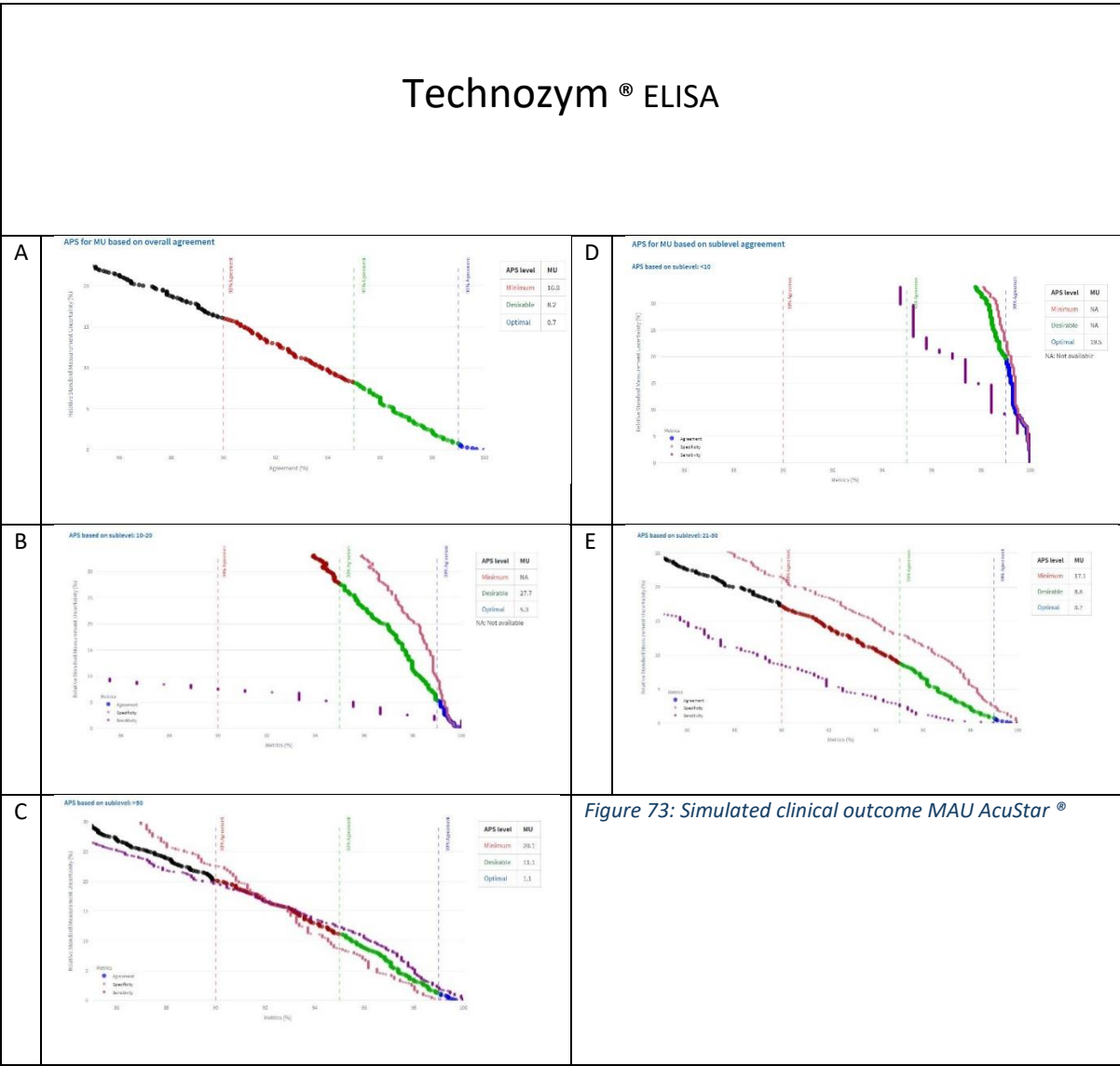
Table 48: Measurand definition comparison for ADAMTS13 activity quantification.

	Technozym ELISA	AcuStar	Ceveron	Technoscreen
Component/Analyte	a disintegrin and metalloproteinase with a thrombospondin type-1 motif, member 13			
Common name	ADAMTS13			
Biological system	Human plasma			
Sample type	Sodium Citrate (0.109mol sodium citrate, 1 part to 9 parts)			
Calibrator	1 st WHO international standard 12/252 (Hubbard et al, 2015)			
Calibrator Traceability	In house calibration to WHO 1 st International standard (12/252) NA			
Unit	IU/mL, IU/dL			
Unit traceability	1 IU is equivalent to the ADAMTS13 activity in pooled normal plasma (WHO/BS/2014.2246)			
Measurement method	Cleavage of GST-vWF73 substrate - End point absorbance changes proportional to activity of ADAMTS13 at 450nm	Cleavage of GST-vWF73 substrate - End point chemiluminescence generation proportional to activity of ADAMTS13 at 450nm	Cleavage of GST-vWF73 substrate - End point fluorescence resonance energy transfer (FRET) proportional to activity of ADAMTS13 at 450nm	Semi-Quantitative flow through assay

	Technozym ELISA	AcuStar	Ceveron	Technoscreen
Measurement Procedure	Technozym ® ADAMTS13 activity ELISA	HemosIL AcuStar ® ADAMTS13 activity	Technofluor ® ADAMTS13 activity	Technoscreen ® Rapid ADAMTS13 activity
Intended purpose	Quantification of ADAMTS13 activity for the diagnosis and monitoring of TTP_and other Microangiopathic Haemolytic Anaemias (MAHA)			
Measurand	The quantification of ADAMTS13 activity by ELISA in citrated human plasma			

12.2 Appendix B – Clinical outcome simulations for MAU

This appendix presents clinical outcome simulations used to derive maximum allowable measurement uncertainty (MAU) for clinical outcome based analytical performance specifications (APS, Model 1b). Simulations were performed using a Monte Carlo modelling approach implemented via the web-based APS Calculator (100). These outputs provide supporting evidence for the derivation of MAU limits, demonstrate the application of simulation-based APS methods, and present full results not included in the main text for clarity.



HemoSIL AcuStar®

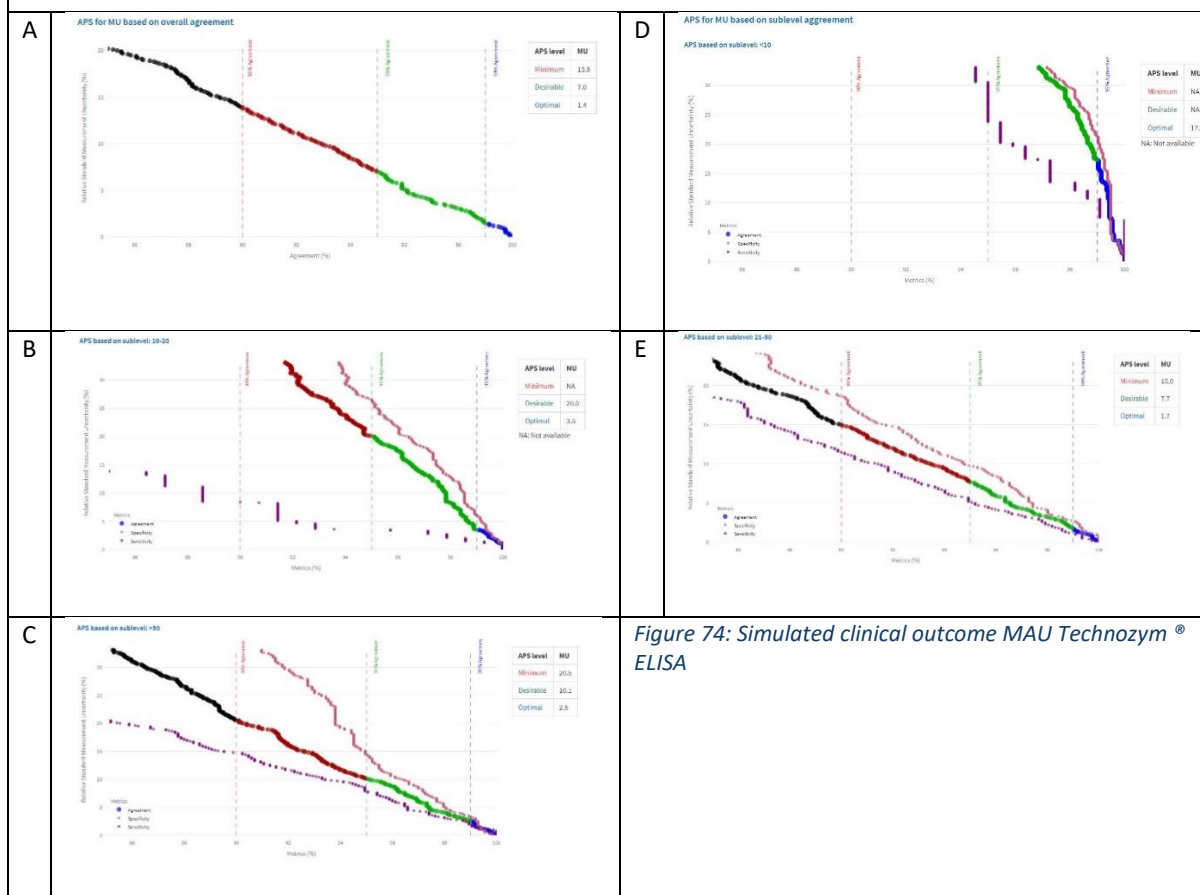


Figure 74: Simulated clinical outcome MAU Technozym® ELISA

12.3 Appendix C – Reference Change Value proportions

This appendix presents a comparative analysis of reference change values (RCVs) for ADAMTS13 activity using four published calculation models: Fraser (1991), Fraser updated (2011), Solé-Morata et al., and EFLM-based specifications (Fokkema et al.). The table summarises the frequency with which each model flagged significant changes across a dataset of ADAMTS13 results, highlighting the impact of RCV method choice on result interpretation.

Table 49: RCV of ADAMTS13 activity based on four different models.

Significant change in results across RCV calculation methods. Proportion of sample results where a significant RCV was detected by four statistical measures of result differences.

Patient	Common RCV	Solétormos	Fraser	Fokkema
A	38%	38%	38%	38%
B	8%	8%	8%	8%
C	33%	25%	17%	25%
D	50%	50%	21%	50%
E	47%	53%	42%	47%
F	43%	43%	36%	43%
G	20%	20%	27%	13%
H	19%	13%	6%	13%
I	21%	21%	5%	21%
J	22%	24%	19%	22%
K	0%	35%	16%	27%
L	0%	44%	33%	44%
M	9%	27%	27%	27%
N	0%	11%	11%	11%
O	0%	50%	33%	39%
P	0%	15%	25%	10%
Q	0%	25%	13%	25%
R	0%	35%	29%	35%
S	0%	21%	21%	14%
T	0%	34%	29%	34%
U	5%	18%	18%	18%
V	15%	12%	12%	10%
W	100%	100%	100%	100%
X	25%	25%	33%	17%
Y	35%	35%	35%	29%
Average	14%	29%	23%	27%

12.4 Appendix D – ANOVA variance component analysis

This appendix presents a variance component analysis of internal quality control data from the Technozym® ADAMTS13 ELISA assay. The analysis was conducted using ANOVA to quantify within-run and between-run sources of variability. Descriptive statistics are provided for both normal and low-level QC material. The resulting variance components support later calculations of measurement uncertainty for this method.

Table 50: Variance component analysis of Technozym® internal quality control.

Descriptive statistics of the normal and low quality control (QC) data used for the variance component analysis from the Technozym® ELISA method. SD – standard Deviation. Min – minimum result, max – maximum result, N – number of QC points

Row Labels	N of QC points	Mean Result	SD	Min Result	Max result
Normal	88	72.1	8.9	56.3	100.0
Control					
RU5450701	2	69.8	5.9	65.6	74.0
RU83B00	24	72.8	8.1	62.3	100.0
RU85B00	10	68.7	7.0	60.0	85.1
RU88B00	48	71.8	9.8	56.3	93.0
RU89B00	4	80.2	2.5	76.7	82.5
Low Control	88	22.7	2.7	14.7	29.2
RU5450701	2	23.7	1.6	22.5	24.8
RU83B00	24	22.1	3.0	14.7	27.7
RU85B00	10	20.6	2.8	17.5	26.4
RU88B00	48	23.4	2.4	19.0	29.2
RU89B00	4	23.5	1.5	22.1	25.3

Table 51: Variance components analysis ANOVA output for Technozym® ELISA High QC.

Example data output from Anova calculation of variance components for a single lot number of Normal QC (RU83B00). SS – Sum of squares, MS = Mean Square and represents the data used for the calculations shown below, df – degrees of freedom, F – obtained F statistic (ratio), F-crit – critical F statistics at $p < 0.05$

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1117.644	11	101.604	3.029794	0.034645	2.717331
Within Groups	402.4195	12	33.53495			
Total	1520.063	23				

Table 52: Variance components analysis ANOVA output for Technozym® ELISA Low QC

Example data output from Anova calculation of variance components for a single lot number of Low QC. SS – Sum of squares, MS = Mean Square and represents the data used for the calculations shown below, df – degrees of freedom, F – obtained F statistic (ratio), F-crit – critical F statistics at $p < 0.05$

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	181.2653	11	16.47867	8.599791	0.000408	2.717331
Within Groups	22.99405	12	1.916171			
Total	204.2594	23				

Calculations (Technozym® Normal QC used as an example, using data from Table 51)

Using the CLSI EP29-A methodology:

Within-run standard deviation:

$$S_{wth} = \sqrt{MS_{within}}$$

$$S_{wth} = \sqrt{33.53495}$$

$$S_{wth} = 5.79$$

Between-run standard deviation (assuming $n_0=2$ replicates per run):

$$s_{btw} = \sqrt{MAX\left(0, \frac{MS_{btw} - MS_{wth}}{n_0}\right)}$$

$$s_{btw} = \sqrt{MAX\left(0, \frac{101.604 - 33.535}{2}\right)}$$

$$s_{btw} = \sqrt{34.0345}$$

$$s_{btw} = 5.83$$

Combined standard uncertainty:

$$u_c = \sqrt{s_{wth}^2 + s_{btw}^2}$$

$$u_c = \sqrt{5.79^2 + 5.83^2}$$

$$u_c = \sqrt{33.54 + 34.0}$$

$$u_c = 8.22$$

12.5 Appendix E - ANOVA variance component analysis for

Preliminary MU

This appendix presents a variance component analysis of internal quality control data from the AcuStar ADAMTS13 activity assay. ANOVA was used to estimate within-run and between-run sources of imprecision across both high and low QC levels. These results provide a basis for preliminary measurement uncertainty estimation and allow comparison of assay performance across control levels.

Table 53: Imprecision and ANOVA output data for high and low ADAMTS13 QC - AcuStar.

df – degrees of freedom, VWF – von Willebrand Factor, SEM – Standard Error of the Mean, SD – Standard Deviation

Parameter	VWF Low	VWF Normal
Between group df	4	4
Within group df	20	20
Number of observations	25	25
Mean	32.81	74.77
SEM	0.32	0.81
Average of average of series	32.81	74.77
Average of series' SD	0.94	2.12
Within series variance	0.89	4.48
Between series variance	1.95	14.54
Intra-laboratory variance	2.84	19.02
Within series imprecision (sX)	0.94	2.12
Between series imprecision (sX)	1.4	3.81
Intra-laboratory imprecision (SD)	1.68	4.36
Within series relative imprecision (%CV)	2.9	2.8
Between series relative imprecision (%CV)	4.3	5.1
Intra-laboratory imprecision (%CV)	5.1	5.8

12.6 Appendix F – Top-Down uncertainty of the Technozym® ELISA

This appendix presents the final measurement uncertainty estimate for the Technozym® ADAMTS13 ELISA, calculated using the top-down approach based on internal quality control data. The analysis demonstrates the application of this method to local performance data, providing an empirical uncertainty value used to support method validation and result interpretation.

Table 54: Calculation of MU of the Technozym ELISA method

Measurement unit	IU/dL	
Measurement method	The amount of substance concentration of a disintegrin and metalloproteinase with a thrombospondin type-1 motif, member 13 (ADAMTS13) activity by cleavage of vWF73 using ELISA in citrated human plasma	
Measurement procedure	ELISA based method using Technozym® ELISA	
Highest order traceability	1 st WHO International standard for ADAMTS13 function	
Method of International standard value assignment	Collaborative study, value assignment as mean of mean values provided	
In use Calibrator	Manufacturer provided calibrator with assigned value (in IU/dL) assigned using manufacturer in house procedures (Data available on website and IFU)	
u_{ref} - Uncertainty of Reference standard, method	Standard Error of mean value assignment from collaborative study Uncertainty of vial concentration based on coefficient of variation of the volume within each vial	
u_{cal} - Calibrator uncertainty *	Assessed separately at the concentration nearest to each QC level	
Bias	EQA scheme participation over the previous 12 months – 3 surveys using a commutable material with target value set by consensus median. Average bias not medically significant. No bias noted on any individual survey	
u_{bias}	Not required – bias acceptable	
IQC	Plasma IQC Level 1 – Technozym ADAMTS13 activity High QC	Plasma IQC Level 1 – Technozym ADAMTS13 activity Low QC

Lot Numbers	RU01C0	RU01C0
Level monitored	Normal (80-100 IU/dL)	Low Abnormal (20-40 IU/dL)
Assay reference range clinical decision limit		<10 IU/dL
Clinical application of level monitored	Quantification of ADAMTS13 activity (function) for monitoring of congenital and immune mediated Thrombotic Thrombocytopenic Purpura	Quantification of ADAMTS13 activity (function) for diagnosis of congenital and immune mediated Thrombotic Thrombocytopenic Purpura
Data collection period	01/01/2022 – 31/07/2022	01/01/2022 – 31/07/2022
Number of QC points assessed	44	44
Manufacturer Target value (IU/dL)	56-88	12-32
Local Target Value (IU/dL)	72	22
Mean value (IU/dL)	72.6	22.9
Locally assigned SD	5.00	2.20
$u_{RW}(\text{ADAMTS13})$ (IU/dL)	4.64	1.71
$u(\text{ADAMTS13}) = \sqrt{(u_{RW}^2 + u_{cal}^2 + u_{ref}^2)}$ IU/dL	$\sqrt{(4.64^2 + 1.28^2 + 0.4^2)} = 4.82$	$\sqrt{(1.71^2 + 3.0^2 + 0.4^2)} = 3.48$
% $U_{rel}(\text{ADAMTS13})$	6.6%	6.7%
Expanded MU, (k=2), 95% confidence	13.2%	13.4%
Maximum Allowable MU assignment	Local agreement with clinical lead, evidenced using clinical outcome modelling (Model 1b, simulation studies)	Local agreement with clinical lead, evidenced using clinical outcome modelling (Model 1b, simulation studies)
Maximum Allowable MU	20%	20%
MU acceptable	Yes	Yes

12.7 Appendix G – Pooled SD Method for combined uncertainty of

HemosIL AcuStar®

This appendix presents a pooled standard deviation (SD_{pooled}) calculation across high and low-level ADAMTS13 quality control material measured using the HemosIL AcuStar® assay. This pooled estimate provides a simplified representation of method imprecision and forms the basis for an approximate calculation of measurement uncertainty using the top-down approach.

Table 55: The pooled standard deviation (SD_{pooled}) across ADAMTS13 quality control material.

Q-VWFL (Low QC)	n	SD
3738	2	0.02
4189	14	0.03
4284	29	0.06
4538	35	0.04
4879	37	0.02
SD_{pooled}		0.03
Q-VWFN (Normal QC)		
3738	2	0.08
4189	12	0.08
4284	24	0.12
4538	31	0.07
4879	37	0.04
SD_{pooled}		0.08

12.8 Appendix H - Top-Down uncertainty of the HemoSIL AcuStar®

Table 56: Calculation of MU for the HemoSIL AcuStar® method

This appendix presents the final measurement uncertainty estimate for the HemoSIL AcuStar® ADAMTS13 activity assay using the top-down approach. The calculation incorporates internal QC data and includes the manufacturer-supplied uncertainty of the calibrator-assigned value (u_{cal}). This analysis demonstrates the application of the top-down method and supports later comparison of uncertainty estimates across different assay platforms.

Measurement unit	%	
Measurement method	The amount of substance concentration of a disintegrin and metalloproteinase with a thrombospondin type-1 motif, member 13 (ADAMTS13) activity by cleavage of vWF73 using Chemiluminescence by HemoSIL AcuStar® in citrated human plasma	
Measurement procedure	Automated coagulometer system	
Highest order traceability	1 st WHO International standard for ADAMTS13 function	
Method of International standard value assignment	Collaborative study, value assignment as mean of mean values provided	
In use Calibrator	Manufacturer provided calibrator with assigned value (in %) assigned using manufacturer in house procedures (Data available on website and IFU)	
u_{ref} - Uncertainty of Reference standard, method	Standard Error of mean value assignment from collaborative study Uncertainty of vial concentration based on coefficient of variation of the volume within each vial	
u_{cal} - Calibrator uncertainty *	Assigned as 10% as per manufacturer recommendations through personal communication	
Bias	EQA scheme participation over the previous 12 months – 3 surveys using a commutable material with target value set by consensus median. Average bias not medically significant. No bias noted on any individual survey	
u_{bias}	Not required – bias acceptable	
IQC	VWF High (HemoSIL, Werfen)	VWF Low (HemoSIL, Werfen)

Lot Numbers	Analysed separately as detailed above – 5 lot numbers over the 6 month period	Analysed separately as detailed above – 5 lot numbers over the 6 month period
Level monitored	Normal (80-100 IU/dL)	Low Abnormal (20-40 IU/dL)
Assay reference range clinical decision limit		<10 IU/dL
Clinical application of level monitored	Quantification of ADAMTS13 activity (function) for monitoring of congenital and immune mediated Thrombotic Thrombocytopenic Purpura	Quantification of ADAMTS13 activity (function) for diagnosis of congenital and immune mediated Thrombotic Thrombocytopenic Purpura
	VWF High	VWF low
Number of QC points assessed	117	117
Manufacturer Target value (IU/dL)	Varied by lot number	Varied by lot number
$u_{RW}(\text{ADAMTS13 } \%)$	8.00	3.00
$u(\text{ADAMTS13}) = \sqrt{(u_{RW}^2 + u_{cal}^2 + u_{ref}^2)} \text{ IU/dL}$	$\sqrt{(8^2 + 10^2 + 0.4^2)} = 12.8$	$\sqrt{(3^2 + 10^2 + 0.4^2)} = 10.4$
$U(\text{ADAMTS13}); k = 2; \text{ mmol/l}$	25.6	20.8
Applied to patients' results (at approx. 95 % confidence)	$\pm 16.0\%$	$\pm 29.8\%$
Maximum Allowable MU assignment	Local agreement with clinical lead	Local agreement with clinical lead
Maximum Allowable MU	20%	20%
MU acceptable	Yes	Yes – after discussion regarding influence of inappropriate information provided with ucal, uRW alone accounts for Uc of 10%

12.9 Appendix I – Published method comparison studies

The below table shows a review of published literature up to 2023, highlighting issues with reporting of statistical methods in method comparison studies of rapid assays (HemoSIL AcuStar[®] and Technofluor[®] FRET) and standard methodologies (Technozym[®] ELISA and FRET-VWF73).

Table 57: Summary of published method comparison studies for AcuStar.

PB – Passing-Bablok, ELISA – Enzyme Linked Immunosorbent assay, FRET, Fluorescent Resonance Energy Transfer, vWF – von Willebrand Factor, aHUS atypical Haemolytic Uraemic Syndrome, TMA – Thrombotic Microangiopathy. LOA – Limits of Agreement, BA – Bland-Altman analysis.

Study	Assays Compared	Sample Selection	Correlation (r)	Regression Type & Slope	Regression Intercept	Bland-Altman	Findings and Limitations
(240)	HemosIL AcuStar vs FRET-VWF73	539 plasma samples, TTP and other TMAs	$r^2 = 0.83$ (P < .0001)	Linear Regression, Slope = 0.9698 (CI not provided)	Intercept = 9.608 (CI not provided)	Mean bias = 8.328 IU/dL, LoA = -24.32 to +40.97 IU/dL	Good correlation with reference method ($r^2 = 0.83$), but sample selection may limit generalisability. Bland-Altman shows a bias indicating systematic differences between methods.
(238)	HemosIL AcuStar vs FRET	706 plasma samples over 14 years	$r = 0.919$ (full dataset), $r = 0.912$ (diagnostic samples)	Passing-Bablok, Slope = 0.962 (CI: 0.937 to 1.00)	Intercept = -0.0096 (CI: -0.01 to -0.0006)	Mean bias = -0.03 kIU/L (CI: -0.04 to -0.02)	High correlation ($r = 0.919$) and strong diagnostic performance reported. Use of a very large sample size improves reliability. Passing-Bablok regression strengthens results but Bland-Altman bias suggests minor overestimation.
(251)	HemosIL AcuStar vs Technozym ELISA	700+ plasma samples	$r =$ not explicitly stated	Linear regression, specifics not detailed	Not provided	Negative bias for HemosIL compared to Technozym	Large multicentre evaluation improves external validity. However, negative bias observed in comparison to reference ELISA assay. Uncertainty

Study	Assays Compared	Sample Selection	Correlation (r)	Regression Type & Slope	Regression Intercept	Bland-Altman	Findings and Limitations
							about sample selection criteria weakens diagnostic conclusions.
(252)	HemosIL vs AcuStar vs Technofluor vs FRET	Acute TTP, congenital TTP, controls, follow-ups	HemosIL vs Technofluor r = 0.98	Passing-Bablok, Slope = 1.1 (CI: 1.0 to 1.2)	Intercept = 2.3	Bias not explicitly stated	High correlation (r = 0.98), but unclear if all clinical subgroups were equally represented. Sample sizes in some categories are not well defined, potentially affecting agreement conclusions.
(283)	HemosIL vs AcuStar vs FRET-VWF73, Technozym, DG-EIA	241 frozen plasma samples	r = not explicitly stated	Passing-Bablok, Slope varies per comparison	Intercept varies per comparison	Mean bias = 8.40%, LoA = -21.60 to 38.50%	Underestimation bias noted. High specificity (100%) but moderate agreement with gold standard assays. Use of frozen plasma may impact reproducibility.
(272)	HemosIL vs AcuStar vs Technozym, Ceveron, FRET-VWF73	94 patient samples, mix of prospective and retrospective	AcuStar vs FRET-VWF73 r = 0.84	Passing-Bablok, specifics not detailed	Not provided	AcuStar results significantly lower than other methods	Significant differences between methods noted (AcuStar lower values). Small sample size may limit generalisability. Inclusion of retrospective samples could introduce selection bias.
(284)	HemosIL vs AcuStar® vs Technozym®	24 paired test samples, including TTP patients	r = 0.93	Passing-Bablok, Slope = 1.04 (CI: 0.84 to 1.17)	Intercept = -2.56 (CI: -5.07 to -0.86)	Mean bias = 2.54%, LoA not explicitly stated	Strong correlation (r = 0.93) but small sample size raises concerns about statistical power. Regression intercept (-2.56) suggests systematic differences.

Study	Assays Compared	Sample Selection	Correlation (r)	Regression Type & Slope	Regression Intercept	Bland-Altman	Findings and Limitations
(123)	HemosIL AcuStar vs FRETs- VWF73, Technozym	176 plasma samples including TTP, aHUS, and controls	HemosIL vs FRETs- VWF73 r = 0.96, vs Technozy m r = 0.96	Passing- Bablok, Slope = 1.05 (FRETs), 1.02 (Technozym)	Intercept = -0.41 (FRETs), 0.05 (Technozy m)	Mean bias = -15.9% (FRETs), 2.9% (Technozy m)	High agreement with gold standard (kappa 0.97), good correlation with reference methods. Some evidence of underestimation bias at high activity levels.

12.10 Appendix J – Stability raw data

This appendix contains the raw dataset used to assess the stability of ADAMTS13 activity over time using the Technozym® ELISA method. Data are presented for individual participants at each measured time point and include calculated statistics such as standard deviation (SD), coefficient of variation (CV), 95% confidence intervals (CI), standard error of the mean (SEM), and percent difference (PD). These data support the analysis of temporal stability under defined storage conditions.

Table 58: Stability of ADAMTS13 activity.

Raw data for individual participants at each time point demonstrating the stability of ADAMTS13 activity using the Technozym® ELISA method. SD – Standard Deviation, CV – Coefficient of Variation, CI – 95% confidence interval, SEM – Standard Error of the Means, PD - Percent Difference, h - hours

ADAMTS13 IU/dL		Baseline	24h	48h	72h
Patient 1	Replicate 1	36.5	33.5	38.5	29.7
Patient 1	Replicate 2	40.6	33.9	36.2	31.2
Patient 1	Mean	38.6	33.1	34.7	31.5
Patient 1	SD	2.9	0.3	1.6	1.1
Patient 1	CV%	8%	1%	5%	3%
Patient 1	Difference from baseline (%)		-14%	-10%	-18%
Patient 2	Replicate 1	42.9	38.6	36.9	34.9
Patient 2	Replicate 2	46.9	37.6	35.6	33.3
Patient 2	Mean	44.9	38.1	36.3	34.1
Patient 2	SD	2.8	0.7	0.9	1.1

ADAMTS13		Baseline	24h	48h	72h
IU/dL					
Patient 2	CV%	6%	2%	3%	3%
Patient 2	Difference from baseline (%)		-15%	-19%	-24%
Patient 3	Replicate 1	79.9	83.1	75.2	74.1
Patient 3	Replicate 2	83.2	81.2	77.1	75.9
Patient 3	Mean	81.6	82.2	76.2	75.0
Patient 3	SD	2.3	1.3	1.3	1.3
Patient 3	CV%	3%	2%	2%	2%
Patient 3	Difference from baseline (%)		1%	-7%	-8%
Patient 4	Replicate 1	56.2	50.1	48.9	48.5
Patient 4	Replicate 2	52.6	53.2	51.2	44.9
Patient 4	Mean	54.4	51.7	50.1	46.7
Patient 4	SD	2.5	2.2	1.6	2.5
Patient 4	CV%	5%	4%	3%	5%
Patient 4	Difference from baseline (%)		-5%	-8%	-14%
Patient 5	Replicate 1	72.6	74.2	67.8	65.2
Patient 5	Replicate 2	76.1	69.1	67.4	64.2
Patient 5	Mean	74.4	71.7	67.6	64.7

ADAMTS13 IU/dL		Baseline	24h	48h	72h
Patient 5	SD	2.5	3.6	0.3	0.7
Patient 5	CV%	3%	5%	0%	1%
Patient 5	Difference from baseline (%)		-4%	-9%	-13%
Patient 6	Replicate 1	8.9	9.5	8.8	8.2
Patient 6	Replicate 2	11.2	9.1	8.9	8.4
Patient 6	Mean	10.1	9.3	8.9	8.3
Patient 6	SD	1.6	0.3	0.1	0.1
Patient 6	CV%	16%	3%	1%	2%
Patient 6	Difference from baseline (%)		-7%	-12%	-17%
Patient 7	Replicate 1	23.8	21.4	19.2	18.7
Patient 7	Replicate 2	20.2	22.3	19.5	18.6
Patient 7	Mean	22.0	21.9	19.4	18.7
Patient 7	SD	2.5	0.6	0.2	0.1
Patient 7	CV%	12%	3%	1%	0%
Patient 7	Difference from baseline (%)		-1%	-12%	-15%
Patient 8	Replicate 1	45.6	42.3	43.9	42
Patient 8	Replicate 2	44.8	44.1	43.5	44.2

ADAMTS13		Baseline	24h	48h	72h
IU/dL					
Patient 8	Mean	45.2	43.2	43.7	43.1
Patient 8	SD	0.6	1.3	0.3	1.6
Patient 8	CV%	1%	3%	1%	4%
Patient 8	Difference from baseline (%)		-4%	-3%	-5%
Patient 9	Replicate 1	70.3	67.5	63.2	61.2
Patient 9	Replicate 2	70.8	69.2	66.2	62.8
Patient 9	Mean	70.6	68.4	64.7	62.0
Patient 9	SD	0.4	1.2	2.1	1.1
Patient 9	CV%	1%	2%	3%	2%
Patient 9	Difference from baseline (%)		-3%	-8%	-12%
Patient 10	Replicate 1	52.1	50	53.1	50.5
Patient 10	Replicate 2	57.1	54.9	52.3	47.6
Patient 10	Mean	54.6	52.5	52.7	49.1
Patient 10	SD	3.5	3.5	0.6	2.1
Patient 10	CV%	6%	7%	1%	4%
Patient 10	Difference from baseline (%)		-4%	-3%	-10%
Global	Average (PD (%))		-6%	-9%	-14%

ADAMTS13		Baseline	24h	48h	72h
IU/dL					
Global	SD of means		5%	5%	6%
Global	SEM		2%	1%	2%
Global	95% CI Low		-9%	-12%	-17%
Global	95% CI High		-2%	-6%	-10%