Quality assurance for multiplex quantitative clinical chemistry proteomics in large clinical trials

Reijnders, E.; Romijn, F.P.H.T.M.; Arslan, F.; Georges, J.J.J.; Pieterse, M.M.; Schipper, E.R.; ... ; Ruhaak, L.R.

ARTICLE

Quality Assurance for Multiplex Quantitative Clinical Chemistry Proteomics in Large Clinical Trials

Esther Reijnders [a,*] Fred P.H.T.M. Romijn,[a] Figen Arslan,[a] Julien J.J. Georges,[a] Mervin M. Pieterse,[a] Edwin R. Schipper,[a] Sonja Didden-Buitendijk,[a] Machteld C. Martherus-Bultman,[a] Nico P.M. Smit,[a] Nina M. Diederiks,[a] Maxim M. Treep,[a] J. Wouter Jukema [b,c] Christa M. Cobbaert,[a] and L. Renee Ruhaak [a]

Background: To evaluate the clinical performance and effectiveness of a multiplex apolipoprotein panel in the context of cardiovascular precision diagnostics, clinical samples of patients with recent acute coronary syndrome in the ODYSSEY OUTCOMES trial were measured by quantitative clinical chemistry proteomics (qCCP). The ISO15189-accredited laboratory setting, including the total testing process (TTP), served as a foundation for this study. Consequently, tailored quality assurance measures needed to be designed and implemented to suit the demands of a multiplex LC-MS/MS test.

Methods: Nine serum apolipoproteins were measured in 23 376 samples with a laboratory-developed multiplex apolipoprotein test on 4 Agilent 6495 LC-MS/MS systems. A fit-for-purpose process was designed with tailored additions enhancing the accredited laboratory infrastructure and the TTP. Quality assurance was organized in 3 steps: system suitability testing (SST), internal quality control (IQC) evaluation with adjusted Westgard rules to fit a multiplex test, and interpeptide agreement analysis. Data was semi-automatically evaluated with a custom R script.

Results: LC-MS/MS analyses were performed with the following between-run CVs: for apolipoprotein (Apo) (a) 6.2%, Apo A-I 2.3%, Apo A-II 2.1%, Apo A-IV 2.9%, Apo B 1.9%, Apo C-I 3.3%, Apo C-II 3.3%, Apo C-III 2.7%, and for Apo E 3.3% and an average interpeptide agreement Pearson *r* of 0.981.

Conclusions: This is the first study of its kind in which qCCP was performed at this scale. This research successfully demonstrates the feasibility of high-throughput LC-MS/MS applications in large clinical trials.

ClinicalTrials.gov Registration Number: NCT01663402

[a]Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Leiden, the Netherlands; [b]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands; [c]Netherlands Heart Institute, Utrecht, the Netherlands.
*Address correspondence to this author at: Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Albinusdreef 2, Leiden, 2333 ZA, the Netherlands. Tel +31-71-52-64630; e-mail e.reijnders@lumc.nl.
Received March 01, 2024; accepted June 20, 2024.
https://doi.org/10.1093/jalm/jfae092

IMPACT STATEMENT

Quantitative clinical chemistry proteomics allows for the in-house development of multiplexed laboratory-developed tests for protein measurands. Quality assurance measures of these multiplex tests are challenging. We established a laboratory-developed apolipoprotein test to address residual cardiovascular risk. To determine the clinical performance and clinical effectiveness of this test in a large randomized controlled clinical trial, we developed and evaluated a custom quality assurance strategy. The strategy ensured generation of high-quality data. This study shows the feasibility of LC-MS/MS applications in clinical trials, a step towards cardiovascular precision diagnostics.

INTRODUCTION

There is an unmet clinical need to address residual cardiovascular risk beyond optimal lipid-lowering therapy (1). Even after low-density lipoprotein cholesterol (LDL-C) targets are met, a substantial risk of 70% of major adverse cardiovascular events (MACE) remains present (2). To address this residual risk, a molecularly defined status of cardiovascular health and disease (CVD) is needed. The current lipid panel, including LDL-C, high-density lipoprotein cholesterol, total cholesterol, and triglycerides, is lacking this molecular definition and is not fit for purpose anymore (3, 4). For example, LDL-C cannot be accurately measured (direct or indirect) at its low clinical target levels (5, 6). Apolipoproteins, the functional proteins of lipid metabolism, are considered potential candidates to fulfill this unmet clinical need as they are molecularly defined and can be measured directly with mass spectrometry (3, 7).

In 2016, we developed a quantitative clinical chemistry proteomics (qCCP) multiplex apolipoprotein panel test (8). The test now comprises quantitation of apolipoprotein (Apo) (a), A-I, A-II, A-IV, B, C-I, C-II, C-III, and E, as well as phenotyping of Apo E, and has been analytically validated with stable performance over a longer period of time (8–10). According to the test evaluation framework from the European Federation of Clinical Chemistry and Laboratory Medicine Working Group Test Evaluation (11), implementation of a new medical test requires assessment of the clinical performance and clinical effectiveness of the apolipoprotein panel for diagnosing and monitoring patients with CVD.

For prognostic biomarkers such as our apolipoprotein panel, clinical performance and effectiveness should ideally be proven in prognostic studies, in which a large number of individuals are followed over a longer period of time, until sufficient events have occurred to evaluate the prognostic and predictive value of a new test (12). However, such a dedicated study is expensive, time-consuming, and impractical (11, 13). As an alternative, we leveraged an existing study, the ODYSSEY OUTCOMES trial (14), which is a randomized controlled trial initially designed to evaluate the clinical efficacy of a proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitor, alirocumab, in patients with recent acute coronary syndrome. Measurement of our apolipoprotein panel in 23 376 samples of individuals from the ODYSSEY OUTCOMES trial allows evaluation of the clinical effectiveness of this panel. In addition, 12 other biomarkers in the context of CVD were measured on Cobas and Diazyme analyzers.

The analysis of large numbers of samples using tests and technology that is not yet implemented in routine patient practice is challenging. In our

general clinical chemistry laboratory setting, a total testing process (TTP) is in place comprising an ISO15189-accredited quality management system (QMS), standard operating procedures (SOPs), a laboratory information management system (LIMS), and track-and-trace registration. This ensures accurate test results and fulfillment of predefined quality performance indicators, including total turnaround times, analytical precision, accuracy, and sensitivity (15). However, the apolipoprotein multiplex test is a research-based test originally not covered in the TTP. Therefore, concepts of the TTP were adapted and adopted for this test. Here we describe the design and implementation of the study process within our laboratory, with special attention for the evaluation of the quality assurance of LC-MS/MS analyses. Quality assurance in multiplex proteomics is challenging, prompting numerous published efforts (16–20) and the development of various tools (21–24). However, at the start of the current study, publications addressing this, in particular in the context of classic clinical chemistry applications, were scarce. Even in the most recent clinical chemistry guidelines there is no clear consensus on quality assurance rules for multiplex tests (25). Consequently, we designed our own quality assurance procedure tailored to our multiplex test. The presented study process may serve as an example for the implementation of high-throughput LC-MS/MS applications in large clinical trials.

MATERIALS AND METHODS

Apolipoprotein Panel Analyses

In total, 23 376 serum samples were received on dry ice, thawed, mixed, centrifuged, mixed, and divided into aliquots. Samples for LC-MS/MS analysis were stored at −80°C and thawed prior to analysis (see online Supplemental Information and Supplemental Fig. 1 for more details). The trial (NCT01663402) was approved by the institutional review board of each site, and all patients provided informed consent.

Serum apolipoprotein levels for Apo (a), Apo B, Apo A-I, Apo A-II, Apo A-IV, Apo C-I, Apo C-II, Apo C-III, and Apo E including Apo E phenotypes were determined as published earlier for 7 apolipoproteins (8, 26). In brief, serum samples were 20× diluted in 96-well plates and stable isotope-labelled (SIL) peptides were added as internal standard (IS). Serum proteins underwent denaturation, reduction, alkylation, and tryptic digestion. Subsequently, the reaction was quenched, and peptides were concentrated through solid-phase extraction and measured on an 6495 QQQ-MS (6495A or 6495C) (Agilent). Sample preparation was performed semi-automated on a 96-channel BRAVO automated liquid handling platform (Agilent).

A single lot of sequencing grade trypsin (V5111) (Promega) was used. Additionally, one lot of 5 native serum calibrators, traceable to a WHO–IFCC reference material, and one lot of IS mix containing SIL peptides were utilized. Bilevel, internal quality control (IQC) native serum samples were measured in triplicate per batch. IQC target concentrations (online Supplemental Table 1) were determined following CLSI protocols (27). Two lots of IQC samples were used for this study.

A system suitability sample (SSS) containing synthetic peptides reflecting both endogenous (endo) and SIL peptides was prepared. Five replicate measurements of an SSS, followed by one blank, were run before and after sample measurement. Overall, 2 lots of SSS were used for this study.

Data Validation and Data set Creation

LC-MS/MS data were processed using Agilent MassHunter Quantitative Analysis. Concentrations, peak areas, ion ratios, relative responses, retention times, full width half maximum values, and peak symmetry results were exported as one CSV file per batch for further data evaluation using a custom-

made R-script (R version 4.2.2) (28) with RMarkdown and Knitr (version 1.33) (29) in RStudio (version 2023.6.1.524) (30). Data was reported weekly semi-automated to the local principal investigator. Patient results were evaluated on 3 levels: the first line was the technical evaluation by a laboratory technician in Excel templates including monitoring of system suitability test (SST) results and IQC levels on a single-batch level; second-line evaluation was performed, using R-script, by the project leader who evaluated the data over multiple batches: and third-line evaluation and authorization was performed by the local principal investigator who evaluated the data from a clinical chemistry perspective on its clinical soundness. Results were transmitted to LIMS upon approval by all 3 lines. After completion of the full data set, data validation comprised of consistency evaluation of retention times, ion ratios, and IS areas. The LC-MS/MS data set was set up using R.

RESULTS

This section describes the process, including the applied quality measures for multiplex qCCP testing, followed by an analysis of the results obtained from these measures.

Total Testing Process to Support the ODYSSEY OUTCOMES Trial

The laboratory infrastructure with the TTP, operating under ISO15189 standards, including QMS, LIMS, and a team of proficient laboratory technicians, was extensively utilized with a carefully planned approach from sample receipt to data management (Supplemental Fig. 1) (15). This infrastructure was essential for the LC-MS/MS analyses.

Quality Assurance. To ensure high-quality data, quality measures were applied. Because LC-MS/MS analysis is rather complex with semi-automated sample preparation and independent analytical instrumentation, special care was given to the quality assurance process. Specifically, a flow chart for data evaluation was constructed in collaboration with laboratory technicians involving: (*a*) SST, (*b*) IQC evaluation, and (*c*) interpeptide agreement (Fig. 1). In addition, a responsibility–accountability–support–consultation–informed (RASCI) matrix was constructed to delineate the different roles throughout the process (www.rascimethode.nl). Data folders were organized following findable, accessible, interoperable, and reusable (FAIR) principles to allow track-and-trace (31).

SST. SST evaluated the performance of the LC-MS/MS instruments (Fig. 2) (19, 32). Analytical sensitivity was assessed through a minimum threshold for both the IS and endo area of 20 000 counts. The variability of ion ratios and relative responses, which ensures analytical specificity and precision, respectively, was limited to 10% within 5 replicates and 15% between the SSS set before and after the run. Additionally, the maximum deviation of retention times from the average within-run retention time was 12 s to ensure that the full peak was being detected. The maximum allowable carryover for both endo and IS peptides is 1% and is determined by dividing the peptide area of the blank by the peptide area of the last SSS. If any of these criteria were not met for the first SSS set, the instrument was checked before running the clinical samples. Depending on the nature of the issue, either the HPLC or the MS underwent maintenance procedures, which might involve tasks such as cleaning the ion source or replacing the inlet filter of the HPLC (19).

IQC. Bilevel native serum IQC samples were measured in triplicate per batch. Concentrations of all 22 peptides were evaluated with Levey–Jennings plots (33, 34). Adjustments to the Westgard rules were needed to fit a multiplex test (35, 36). The LC-MS/MS test includes 22 peptides yielding 132 IQC data points per batch for which the application of Westgard rules is too restrictive. For instance, for one analyte the $1 \times 3SD$

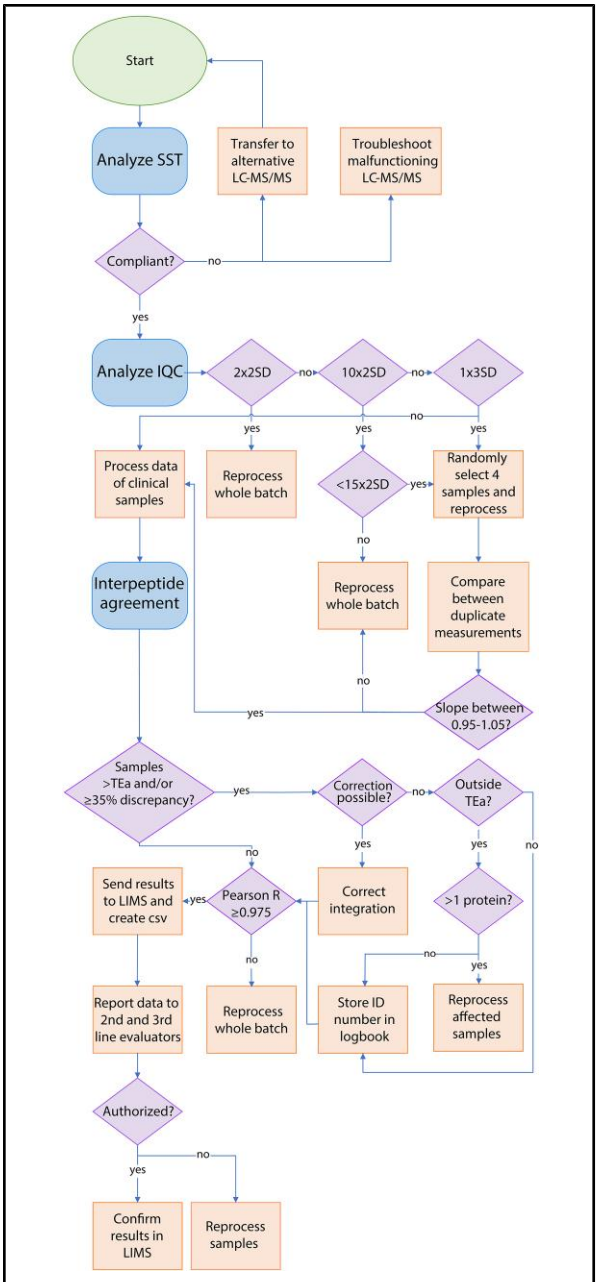
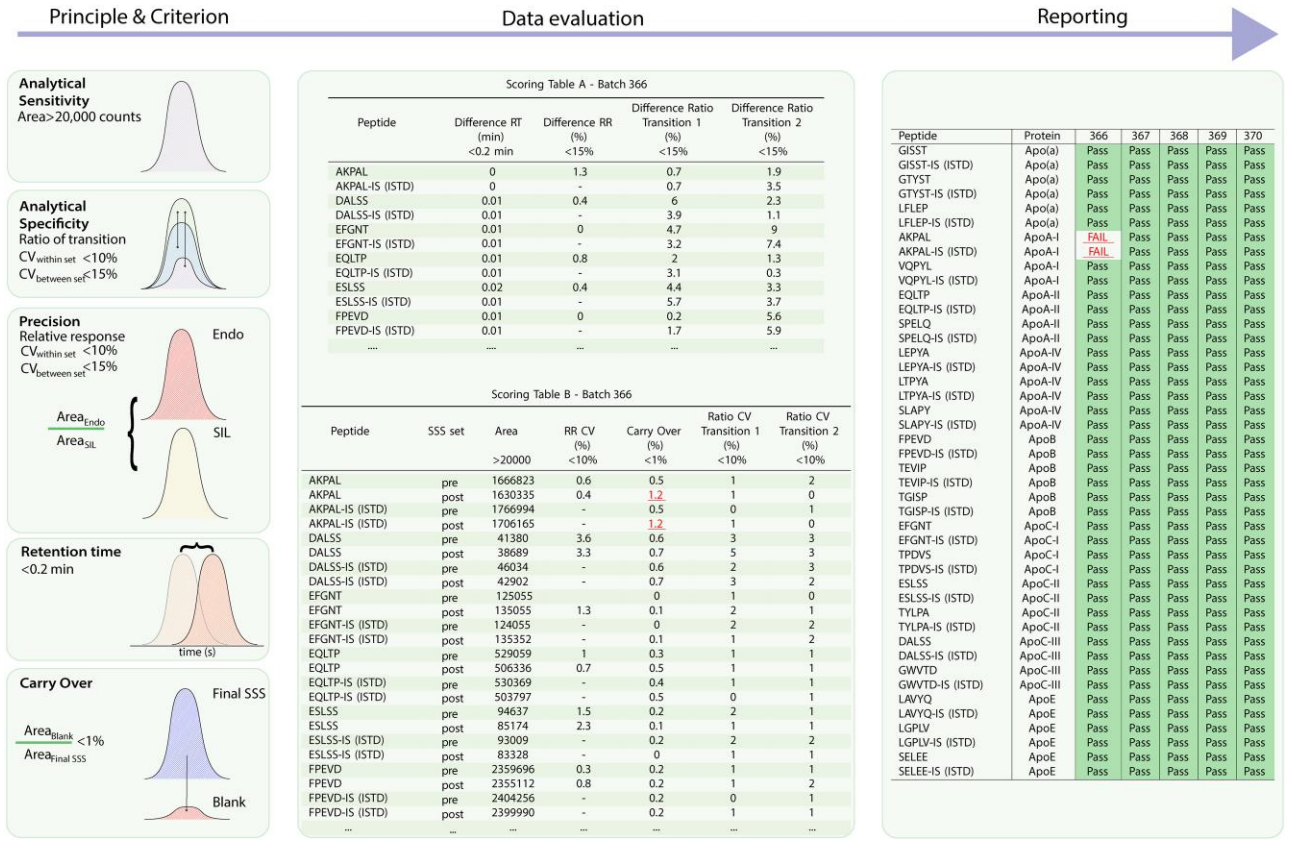


Fig. 1. Quality assurance process of LC-MS/MS measurements. In rounded rectangles, the different validation steps; in rectangles, the actions; and in diamond shapes, the decisions or criteria. Abbreviations: SST, system suitability testing; IQC, internal quality control; TE_a, total allowable error; LIMS, laboratory information management system; csv, comma separated value.

rule provides a 0.27% chance of a false rejection. Translated to our multiplex test, the probability of a batch being falsely rejected based on a 3SD alarm within 132 data points is approximately 30%, which is too stringent. Therefore, 3 rules were applied to our multiplex test: (*a*) 2 × 2SD within the same IQC lot for the same peptide, (*b*) 10 × 2SD for the whole batch, and (*c*) 1 × 3SD for the quantifying peptides only. All 3 rules were established based on our experience with this multiplex test and were applied per batch. The theoretic false rejection rate of the 3 rules was calculated in an independent manner: the 2 × 2SD rule would result in a false rejection rate of 23%, while the 10 × 2SD rule and the 1 × 3SD rule both result in a 13% false rejection rate. The 2 × 2SD rule identifies issues with specific peptides, potentially impacting reported concentrations of clinical samples as these concentrations cannot be confirmed. The 10 × 2SD rule was implemented to monitor overall data dispersion within the batch. The 1 × 3SD rule adheres to a classic Westgard rule but is solely applied to quantifying peptides to make it more suitable for a multiplex test. The 2 × 2SD violation resulted in an immediate rejection of the batch and all samples were reprocessed. In the case of a 10 × 2SD violation there were 2 options: (*a*) in the case of violation of ≥15 × 2SD within the batch, all samples of this batch were reprocessed, or (*b*) in the case where it was <15 × 2SD within the batch, results were confirmed by reprocessing 4 samples of this batch that were randomly selected as representative of the affected batch. Concentrations of all 4 samples of all peptides from the second measurement were plotted against the concentrations of the original measurement and compared with Deming regression (online Supplemental Fig. 2). A Deming slope between 0.95 and 1.05 was acceptable; otherwise, the entire batch was rejected, and all samples were reprocessed. In the case where the final rule, 1 × 3SD, was violated, 4 samples of that batch were randomly selected for reprocessing following the same procedure as described above.



Scoring Table A - Batch 366

Peptide	Difference RT (min) <0.2 min	Difference RR (%) <15%	Difference Ratio Transition 1 (%) <15%	Difference Ratio Transition 2 (%) <15%
AKPAL	0	1.3	0.7	1.9
AKPAL-IS (ISTD)	0	-	0.7	3.5
DALSS	0.01	0.4	6	2.3
DALSS-IS (ISTD)	0.01	-	3.9	1.1
EFGNT	0.01	0	4.7	9
EFGNT-IS (ISTD)	0.01	-	3.2	7.4
EQLTP	0.01	0.8	2	1.3
EQLTP-IS (ISTD)	0.01	-	3.1	0.3
ESLSS	0.02	0.4	4.4	3.3
ESLSS-IS (ISTD)	0.01	-	5.7	3.7
FPEVD	0.01	0	0.2	5.6
FPEVD-IS (ISTD)	0.01	-	1.7	5.9
....	....	...	...	...

Scoring Table B - Batch 366

Peptide	SSS set	Area >20000	RR CV (%) <10%	Carry Over (%) <1%	Ratio CV Transition 1 (%) <10%	Ratio CV Transition 2 (%) <10%
AKPAL	pre	1666823	0.6	0.5	1	2
AKPAL	post	1630335	0.4	1.2	1	0
AKPAL-IS (ISTD)	pre	1766994	-	0.5	0	1
AKPAL-IS (ISTD)	post	1706165	-	1.2	1	0
DALSS	pre	41380	3.6	0.6	3	3
DALSS	post	38689	3.3	0.7	5	3
DALSS-IS (ISTD)	pre	46034	-	0.6	2	3
DALSS-IS (ISTD)	post	42902	-	0.7	3	2
EFGNT	pre	125055		0	1	0
EFGNT	post	135055	1.3	0.1	2	1
EFGNT-IS (ISTD)	pre	124055	-	0	2	2
EFGNT-IS (ISTD)	post	135352	-	0.1	1	2
EQLTP	pre	529059	1	0.3	1	1
EQLTP	post	506336	0.7	0.5	1	1
EQLTP-IS (ISTD)	pre	530369	-	0.4	1	1
EQLTP-IS (ISTD)	post	503797	-	0.5	0	1
ESLSS	pre	94637	1.5	0.2	2	1
ESLSS	post	85174	2.3	0.1	1	1
ESLSS-IS (ISTD)	pre	93009	-	0.2	2	2
ESLSS-IS (ISTD)	post	83328	-	0	1	1
FPEVD	pre	2359696	0.3	0.2	1	1
FPEVD	post	2355112	0.8	0.2	1	2
FPEVD-IS (ISTD)	pre	2404256	-	0.2	0	1
FPEVD-IS (ISTD)	post	2399990	-	0.2	1	1
...	...	...	...	...	...	...

Peptide	Protein	366	367	368	369	370
GISST	Apo(a)	Pass	Pass	Pass	Pass	Pass
GISST-IS (ISTD)	Apo(a)	Pass	Pass	Pass	Pass	Pass
GTYST	Apo(a)	Pass	Pass	Pass	Pass	Pass
GTYST-IS (ISTD)	Apo(a)	Pass	Pass	Pass	Pass	Pass
LFLEP	Apo(a)	Pass	Pass	Pass	Pass	Pass
LFLEP-IS (ISTD)	Apo(a)	Pass	Pass	Pass	Pass	Pass
AKPAL	ApoA-I	FAIL	Pass	Pass	Pass	Pass
AKPAL-IS (ISTD)	ApoA-I	FAIL	Pass	Pass	Pass	Pass
VQPYL	ApoA-I	Pass	Pass	Pass	Pass	Pass
VQPYL-IS (ISTD)	ApoA-I	Pass	Pass	Pass	Pass	Pass
EQLTP	ApoA-II	Pass	Pass	Pass	Pass	Pass
EQLTP-IS (ISTD)	ApoA-II	Pass	Pass	Pass	Pass	Pass
SPELQ	ApoA-II	Pass	Pass	Pass	Pass	Pass
SPELQ-IS (ISTD)	ApoA-II	Pass	Pass	Pass	Pass	Pass
LEPYA	ApoA-IV	Pass	Pass	Pass	Pass	Pass
LEPYA-IS (ISTD)	ApoA-IV	Pass	Pass	Pass	Pass	Pass
LTPYA	ApoA-IV	Pass	Pass	Pass	Pass	Pass
LTPYA-IS (ISTD)	ApoA-IV	Pass	Pass	Pass	Pass	Pass
SLAPY	ApoA-IV	Pass	Pass	Pass	Pass	Pass
SLAPY-IS (ISTD)	ApoA-IV	Pass	Pass	Pass	Pass	Pass
FPEVD	ApoB	Pass	Pass	Pass	Pass	Pass
FPEVD-IS (ISTD)	ApoB	Pass	Pass	Pass	Pass	Pass
TEVIP	ApoB	Pass	Pass	Pass	Pass	Pass
TEVIP-IS (ISTD)	ApoB	Pass	Pass	Pass	Pass	Pass
TGISP	ApoB	Pass	Pass	Pass	Pass	Pass
TGISP-IS (ISTD)	ApoB	Pass	Pass	Pass	Pass	Pass
EFGNT	ApoC-I	Pass	Pass	Pass	Pass	Pass
EFGNT-IS (ISTD)	ApoC-I	Pass	Pass	Pass	Pass	Pass
TPDVS	ApoC-I	Pass	Pass	Pass	Pass	Pass
TPDVS-IS (ISTD)	ApoC-I	Pass	Pass	Pass	Pass	Pass
ESLSS	ApoC-II	Pass	Pass	Pass	Pass	Pass
ESLSS-IS (ISTD)	ApoC-II	Pass	Pass	Pass	Pass	Pass
TYLPA	ApoC-II	Pass	Pass	Pass	Pass	Pass
TYLPA-IS (ISTD)	ApoC-II	Pass	Pass	Pass	Pass	Pass
DALSS	ApoC-III	Pass	Pass	Pass	Pass	Pass
DALSS-IS (ISTD)	ApoC-III	Pass	Pass	Pass	Pass	Pass
GWVTD	ApoC-III	Pass	Pass	Pass	Pass	Pass
GWVTD-IS (ISTD)	ApoC-III	Pass	Pass	Pass	Pass	Pass
LAVYQ	ApoE	Pass	Pass	Pass	Pass	Pass
LAVYQ-IS (ISTD)	ApoE	Pass	Pass	Pass	Pass	Pass
LGPLV	ApoE	Pass	Pass	Pass	Pass	Pass
LGPLV-IS (ISTD)	ApoE	Pass	Pass	Pass	Pass	Pass
SELEE	ApoE	Pass	Pass	Pass	Pass	Pass
SELEE-IS (ISTD)	ApoE	Pass	Pass	Pass	Pass	Pass

Fig. 2. System suitability testing (SST) for the peptides of all apolipoproteins for 5 batches generated with R-script that was run weekly. In cases where SST criteria are met, a "pass" is generated, and in cases where criteria are not met, a "fail" is shown. In this case, batch 366 was rejected and remeasured on another LC-MS/MS system.

Interpeptide agreement. The final step of quality assurance involves the interpeptide agreement assessment between the concentrations of quantifying and qualifying peptides. Disagreement could indicate sample preparation errors, such as inconsistent digestion kinetics, resulting in unreliable results. However, biological variations due to mutations on the measured peptide may also result in discordant concentrations between quantifying and qualifying peptides. Individual samples were reprocessed based on 2 criteria: (*a*) a concentration difference between quantifying and qualifying peptide exceeding the total allowable error (TE_a) for more than one protein and/or (*b*) a concentration difference of >35% between quantifying and qualifying peptide to identify possible mutations that affect protein quantification. The 35% rule was empirically established based on typical discrepancies observed for heterozygous and homozygous mutations, leading to biases of approximately 50% and approximately 100%, respectively. TE_a was defined based on biological variation for Apo (a), Apo A-I, and Apo B resulting in 24.1%, 9.1% and 11.6%, respectively. For the other apolipoproteins, a TE_a of 20% was selected based on the state of the art (19). To assess batch performance, a Pearson's *r* of ≥0.975 was applied as the criterion. A coefficient <0.975 could suggest a potential batch effect. In such cases, all samples in that batch were reprocessed.

Data Verification and Validation.

Daily/weekly data verification. First-line evaluators processed the measurement data in Agilent MassHunter Quantitative Analysis, conducted the evaluation of SST and IQC results with Excel templates, and exported the raw data to a designated folder. A tailor-made R-script was run weekly on the CSV files. This R-script was developed for semi-automated quality evaluation, generating reports including SST, IQC evaluation, and interpeptide agreement plots, as well as the concentrations of the apolipoproteins for weekly review and discussion by second- and third-line evaluators (15). Conclusions were automatically generated based on the predefined criteria. Inconsistencies in individual samples (e.g., caused by improper peak integration) were identified based on visual inspection and corrected if possible after which the sample results underwent the same process and review as other samples. Upon approval by second- and third-line evaluators, results were transmitted to LIMS.

Final data validation. While quality assessment was performed daily, final data evaluation was performed after the study was completed. This included retention time checks between endo and IS peptides. If differences exceeded 0.1 min, data was visually inspected for integration errors. In addition, ion ratios of both endo and IS peptides, as well as IS areas, were visually checked for outliers. If outliers were detected, integration was manually checked and corrected if needed.

Process Performance

The entire process, spanning 74 weeks from sample reception to data set completion, was monitored weekly using a dashboard constructed in RStudio (online Supplemental Fig. 3). Although LC-MS/MS throughput started relatively low, the expansion from 2 LC-MS/MS systems to 4 and from one BRAVO to 3 led to enhanced throughput. This resulted in handling 16 batches per week, translating to a total of 1280 clinical samples per week.

SST. After measurements were conducted, the R-script generated an overview of the SST results of all peptides per batch. An example for 5 batches is presented in Fig. 2 to illustrate the weekly R-output. Thirty-one (9.6%) of the 322 batches were remeasured due to failed SST.

IQC. IQC evaluation was semi-automated and results were reported weekly with the R-script. An example of the 4 batches left, after one batch failed the SST in the previous section, is depicted in Fig. 3 (left panel).

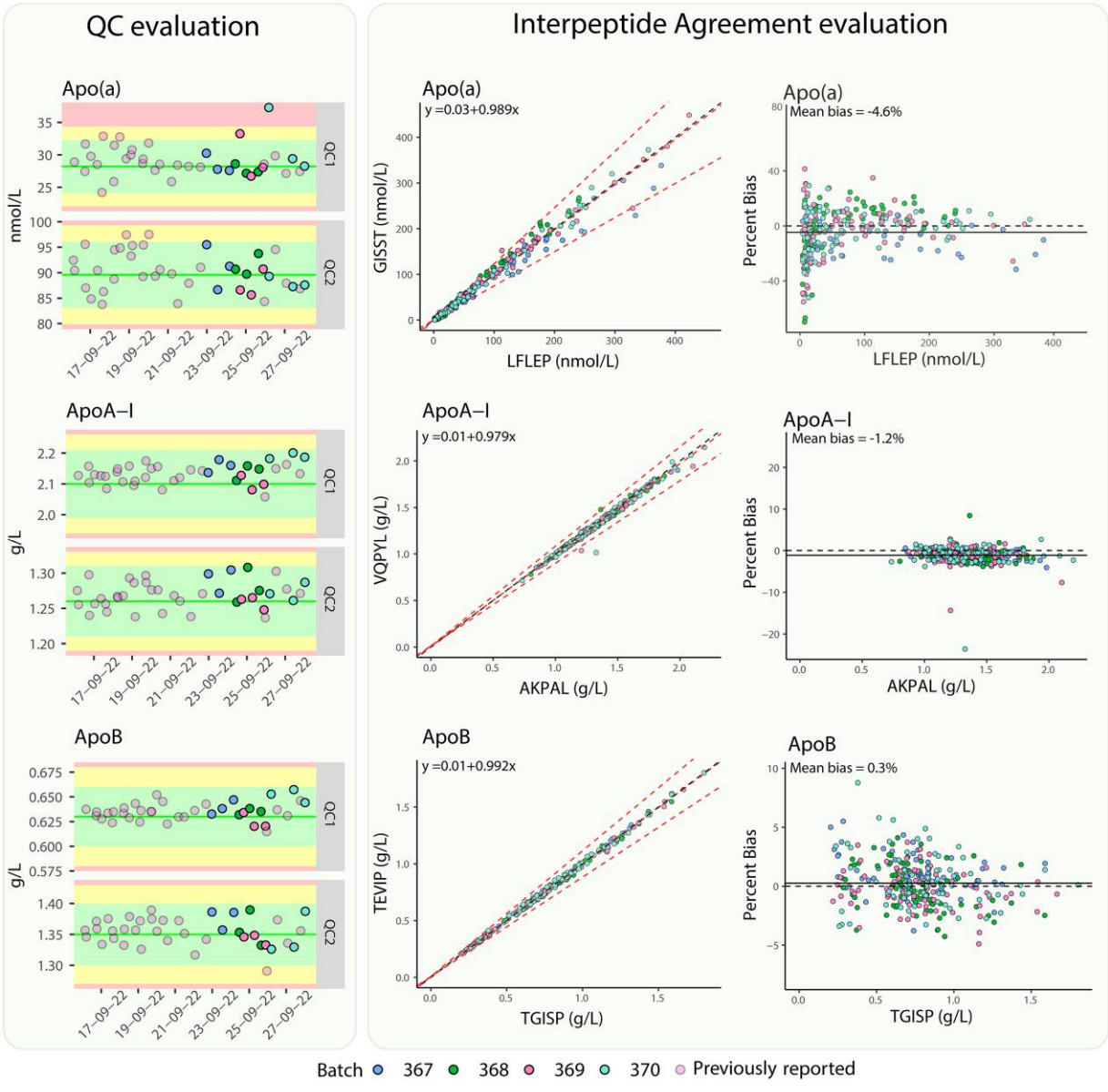


Fig. 3. **Example of QC evaluation and interpeptide agreement evaluation for 4 batches generated with R-script that was run weekly. Left panel: Bilevel IQC monitoring for Apo (a) (quantifying peptide LFLEP), Apo A-I (quantifying peptide AKPAL), and Apo B (quantifying peptide TGISP) with 2SD (inner borders in light-green), 3SD (middle borders in yellow), and outside 3SD (outer borders in red) borders and the target value is depicted as the center line. Right panel: Interpeptide agreement evaluation of the same 4 batches with Deming regression plots (left) with the quantifying peptide on the *x* axis and the qualifying peptide on the *y* axis. On the right: Bland–Altman percentage plots: on the *x* axis the concentration of quantifying peptide, and on the *y* axis the percent bias calculated as the difference between quantifying and qualifying concentrations divided by the quantifying concentration times 100. A bias of 0% is depicted as a solid line and mean bias is depicted as a dashed line. For mean bias calculations of Apo (a), concentrations below the limit of quantitation (3.8 nmol/L) were excluded from the calculation.**

Table 1. Average between-run CV (%) per protein, per IQC lot (1A, 1B, 2A, and 2B) and per LC-MS/MS type (Agilent QQQ-MS 6495A or Agilent QQQ-MS 6495C) for quantifying peptides.[a]

	CV, %								
	IQC level 1A		IQC level 1B		IQC level 2A		IQC level 2B		
Protein	6495A	6495C	6495A	6495C	6495A	6495C	6495A	6495C	Overall (%)
Apo (a)	14.3	13.3	5.4	4.3	7.1	6.6	3.2	3.0	6.2
Apo A-I	3.5	2.0	3.9	2.0	3.8	1.5	3.5	1.6	2.3
Apo A-II	3.2	2.0	2.9	2.1	2.6	1.7	2.4	1.6	2.1
Apo A-IV	4.6	2.3	5.2	2.2	4.5	2.1	4.2	1.9	2.9
Apo B	2.8	2.1	2.0	2.2	2.2	1.5	2.5	1.5	1.9
Apo C-I	5.2	2.8	5.8	3.1	4.0	2.4	4.3	2.7	3.3
Apo C-II	4.5	3.0	4.7	2.7	4.5	2.7	4.0	2.9	3.3
Apo C-III	3.1	2.6	4.0	2.6	3.3	2.0	4.0	2.3	2.7
Apo E	4.4	3.0	4.3	3.2	3.9	3.0	3.0	3.1	3.3

[a]Abbreviations: IQC, internal quality control; Apo, apolipoprotein; 6495A, Agilent QQQ-MS 6495A; 6495C, Agilent QQQ-MS 6495C.

IQC evaluation based on multiplex-adjusted rules for LC-MS/MS. Violation of 2 × 2SD within the same IQC lot for the same peptide was true for 25 batches (322 batches total study [7.7%]). Most of these batches were immediately reprocessed except 5, where upon visual inspection, it was decided that the affected peptide LFLEP concentration of this lot was so low (7 nmol/L) compared with the clinical decision limit (90 nmol/L) that it was clinically of limited importance. To prevent clinically less relevant alarms, IQC levels at more suitable concentrations were implemented. To establish whether the 2 × 2SD rule was effective, duplicate measurements for the full batches were compared retrospectively. For 16 of the 20 batches (5.0% of total study), concentrations between duplicate measurements were not comparable for at least one peptide as the Deming slope was not between 0.95 and 1.05 and, therefore, reprocessing of the samples was a correct decision.

Additionally, 4 (1.2% of total study) batches violated the 10 × 2SD rule. Four random samples were reprocessed for each of the batches. Comparison between results of all peptides from the 4 samples for the first and second preparation indicated equivalence (Deming slope 0.95 to 1.05) and showed no clinically relevant differences in concentrations between both measurements. Therefore, all 4 batches retrospectively passed IQC evaluation.

Violation of the 1 × 3SD rule for the quantifying peptides occurred in 7 batches (2.2%). Four randomly selected samples were reprocessed following the same procedure as described for the 10 × 2SD rule. All 7 batches retrospectively passed IQC evaluation demonstrating equivalence through comparison with a Deming slope ranging between 0.95 and 1.05 and showed no clinically relevant differences in concentrations between both measurements.

Overall IQC results. IQC concentrations were monitored for all 22 peptides with Levey–Jennings plots over a period of 14 months (online Supplemental Fig. 4). After study data set completion, the average between-run CV was calculated (Table 1). CVs vary among different levels of IQC, proteins, and instruments. The average CVs per apolipoprotein were as follows: Apo (a) 6.2%, Apo A-I 2.3%, Apo A-II 2.1%, Apo A-IV 2.9%, Apo B 1.9%, Apo C-I 3.3%,

Apo C-II 3.3%, Apo C-III 2.7%, and for Apo E 3.3%, resulting in an overall between-run CV of 3.1%.

Interpeptide Agreement. An example of the 4 batches left, after all example batches passed IQC evaluation in the previous section, is depicted in the right panel of Fig. 3.

Samples that were annotated to be remeasured were gathered during the study. When 160 samples were reached, an interim analysis was performed by reprocessing these samples in 2 batches. Concentrations from the original and second measurement were compared with Deming regression. Comparison of concentrations per peptide and per sample led to the conclusion that second measurements of 152 samples were similar to the original measurements, confirming the observed discrepancies. For 8 samples, no discrepancy was observed in a second measurement, indicating a false result in the first measurement. The most affected protein was Apo A-I in 6 of the cases. Consequently, the rule was adjusted from discrepancies in 2 or more to 3 or more proteins, except when the discrepancy is observed in Apo A-I as Apo A-I was overrepresented in the 8 samples that were rejected correctly. An overview of the rejected plates per evaluation step is depicted in Fig. 5.

Overall interpeptide agreement. Interpeptide agreement plots for the complete study were created and the results for Apo (a), Apo A-I, and Apo B are depicted in Fig. 4 (the other apolipoproteins are in online Supplemental Table 2 and Supplemental Fig. 5). Discrepancies of ±50% or ±100%, as observed for Apo B, could indicate the presence of a mutation on the measured peptide, either heterozygous or homozygous, respectively. There were no indications of batch effects for any of the batches during the weekly evaluation.

Final Data Validation. After completion of the study data set, final data validation was performed. Thirty samples had a retention time difference between the endo and IS peak of ≥0.1 min, which could be resolved by signal reintegration.

DISCUSSION

For quality assurance of routine laboratory tests Westgard rules are applied (36). However, these rules cannot be transferred to a multiplex LC-MS/MS test in which a multitude of IQC data points is generated. Here, adjusted Westgard rules were developed and tested. The application of adjusted Westgard rules in a multiplex test proved to be challenging. In the total study, 11.1% of the batches were rejected based on the adjusted rules (Fig. 5). Of the 3 applied IQC rules, only the 2 × 2SD rule within the same peptide of the same IQC lot rejected batches correctly. The other 2 rules had no additional benefit as duplicate measurements yielded similar results.

An additional criterion for the Deming regression comparison in future applications could be that the mean bias of the 4 reprocessed samples must fall within twice the average CV of the IQCs. Upon retrospective evaluation, we determined that all batches we compared met this criterion, except one. In future measurements, it could be necessary to reprocess batches that do not meet this criterion. An important note is that the applicability of quality assurance rules in multiplex tests depend on multiple parameters, for example, the number of IQC samples measured in a batch, the number of proteins, and therefore the number of quantifying peptides measured in a batch, as well as the total number of peptides measured. These rules should be fit for purpose for a specific multiplex test and cannot be universally applied to any multiplex test.

An interim analysis as part of the interpeptide agreement evaluation revealed that rejecting samples with discrepancies between peptides of at least 2 proteins was too strict. The rule was modified

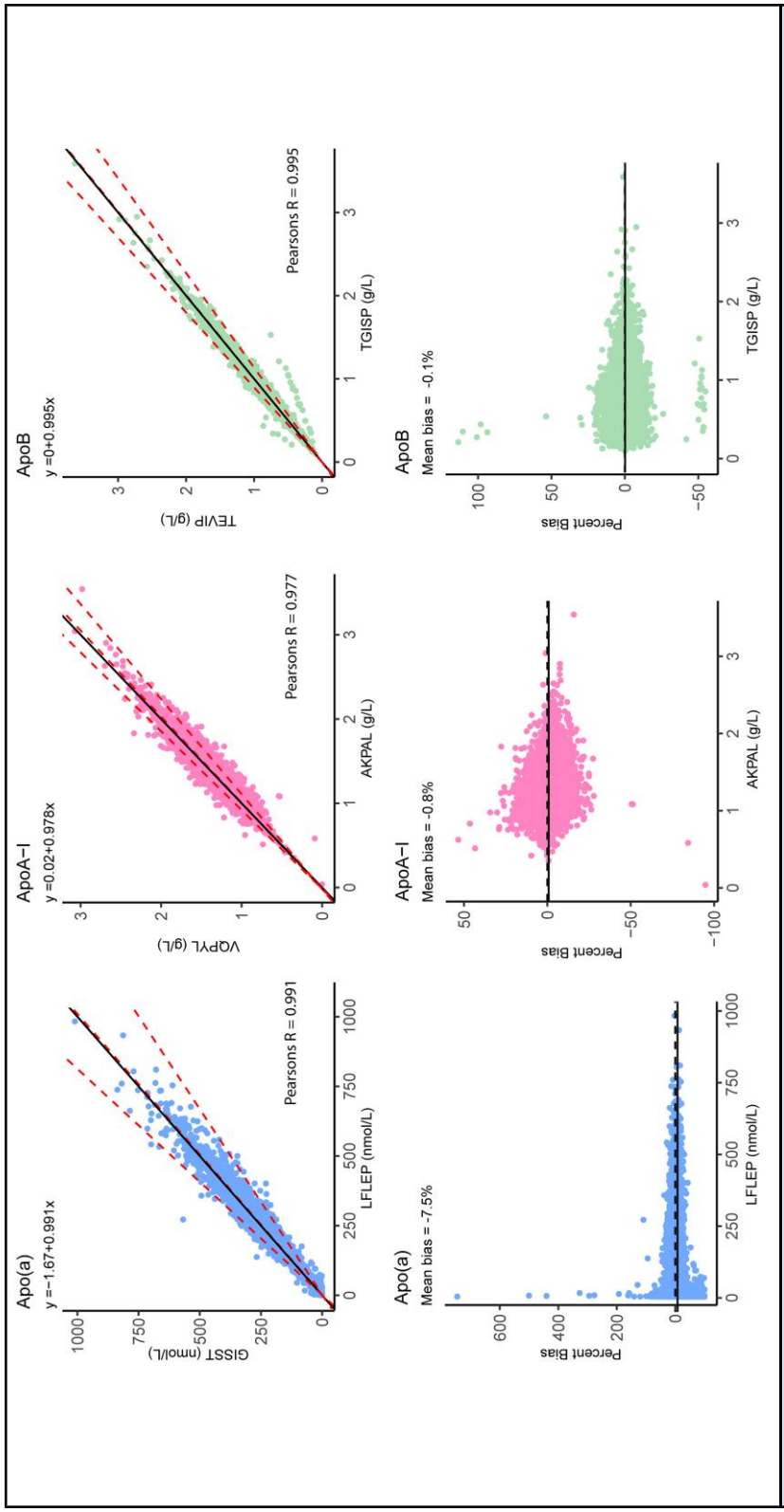


Fig. 4. Interpeptide agreement plots of Apo (a), Apo A-I, and Apo B. Top: Deming regression plots in which the *x* axis represent quantifying peptides and *y* axis the qualifying peptide of the corresponding protein, including the Pearson correlation coefficient (*r*). The black line represents the line of identity, while the red dashed line represents the Deming regression and the total allowable error from the Deming regression (n = 23 376). Bottom: Bland–Altman percentage plots: on the *x* axis, the concentration of quantifying peptide and on the *y* axis the percent bias calculated as the difference between quantifying and qualifying concentrations divided by the quantifying concentration times 100. A bias of 0% is depicted as a black line and mean bias is depicted as a dashed black line. Discordances in interpeptide agreement between peptides could indicate the presence of a mutation in one of the peptides which can be observed for Apo B.

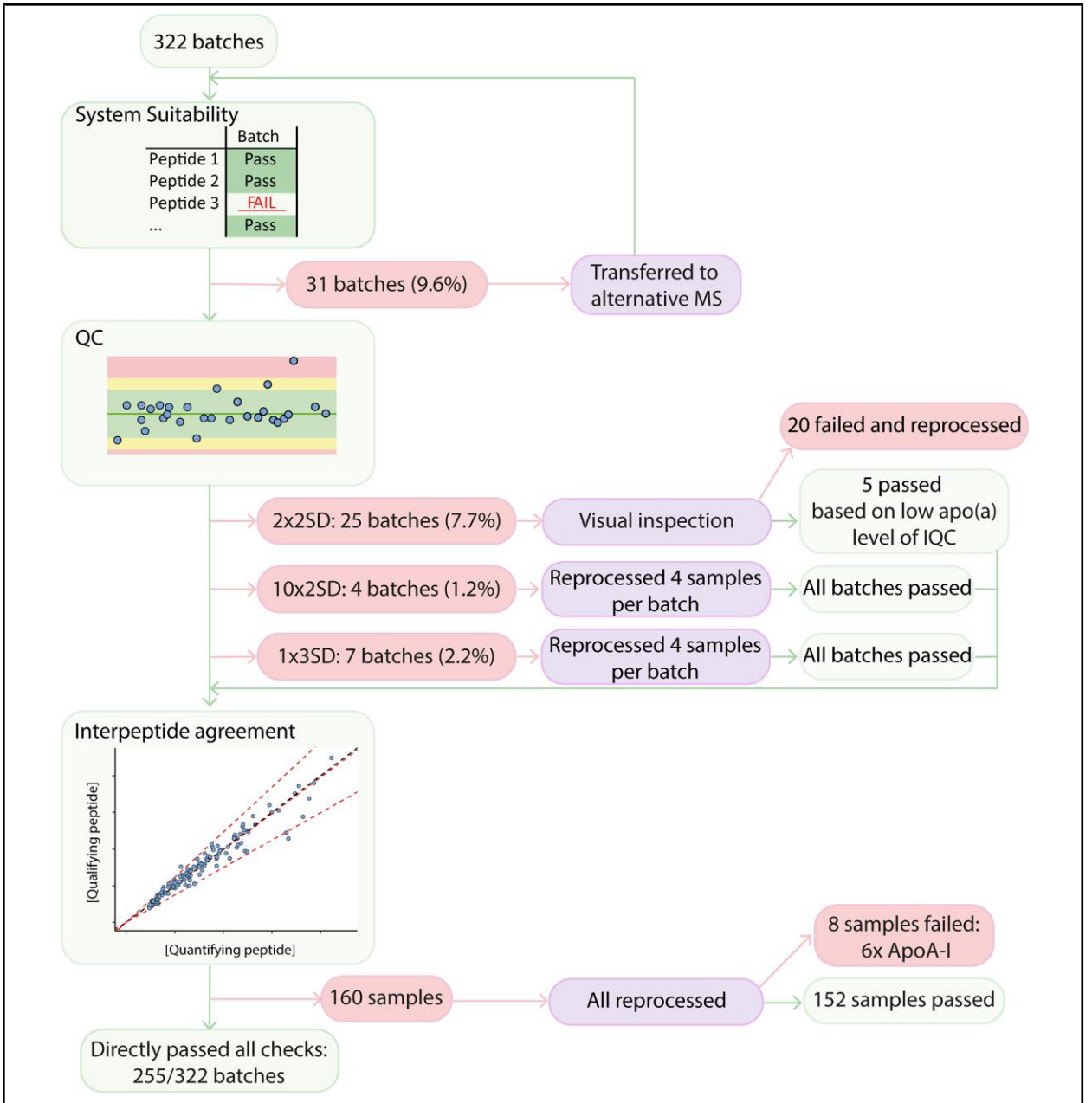


Fig. 5. Overview of results of quality measures per evaluation step. A total of 322 batches were measured of which 31 failed the SST step. Batches were transferred to an alternative LC-MS/MS and remeasured. A total of 36 batches failed QC initially, of which 20 were confirmed rejections upon further evaluation. A total of 160 samples failed the interpeptide agreement evaluation, of which 8 were confirmed rejections upon further evaluation.

to evaluate discrepancies within the sample for more than 2 proteins instead of more than one, except for Apo A-I. Overall, the interpeptide agreement provides an additional layer of quality control at individual sample level in quantitative proteomics procedures.

The presence of a routine general clinical chemistry laboratory as well as research laboratory facilities within the same department was essential for the successful execution of the study. The procedure of the LC-MS/MS measurements required attention to ensure high-quality data. Although the flowchart (Fig. 1) seems complex, it was easy to adhere to and gave clear guidance to all stakeholders.

CONCLUSION

This is the first qCCP trial at this scale performed in a diagnostic clinical chemistry laboratory, which meets both the test process requirements as well as the predefined analytical performance criteria that make medical tests fit for clinical purpose. Anchoring tailored additions to the TTP ensured the generation of high-quality data. These achievements would not have been possible without the dedication of skilled laboratory technicians, whose contributions were crucial to successful execution. In conclusion, we successfully demonstrated the feasibility of employing high-throughput clinical chemistry LC-MS/MS analyses in a large clinical trial setting.

SUPPLEMENTAL MATERIAL

Supplemental material is available at *The Journal of Applied Laboratory Medicine* online.

Nonstandard Abbreviations: qCCP, quantitative clinical chemistry proteomics; TTP, total testing process; SST, system suitability testing; IQC, internal quality control; Apo, apolipoprotein; LDL-C, low-density lipoprotein cholesterol; LIMS, laboratory information management system; IS, internal standard; SSS, system suitability sample; endo, endogenous.

Author Contributions: *The corresponding author takes full responsibility that all authors on this publication have met the following required criteria of eligibility for authorship: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved. Nobody who qualifies for authorship has been omitted from the list.*

Esther Reijnders (Conceptualization-Equal, Data curation-Lead, Formal analysis-Lead, Investigation-Lead, Software-Lead, Validation-Lead, Visualization-Lead, Writing—original draft-Lead, Writing—review & editing-Lead), Fred Romijn (Data curation-Equal, Formal analysis-Equal, Validation-Equal, Writing—review & editing-Supporting), Figen Arslan (Data curation-Supporting, Validation-Supporting), Julien J.J. Georges (Data curation-Supporting, Validation-Supporting), Mervin Pieterse (Data curation-Supporting, Validation-Supporting), Edwin R. Schipper (Data curation-Supporting, Validation-Supporting), Sonja Didden-Buitendijk (Data curation-Supporting, Validation-Supporting), Machteld C. Martherus-Bultman (Data curation-Supporting, Validation-Supporting), Nico Smit (Data curation-Supporting, Validation-Supporting, Writing—review & editing-Supporting), Nina M. Diederiks (Data curation-Supporting, Validation-Supporting), Maxim Treep (Data curation-Supporting, Validation-Supporting), J. Wouter Jukema (Funding acquisition-Equal, Supervision-Equal), Christa Cobbaert (Conceptualization-Lead, Funding acquisition-Lead, Supervision-Lead, Writing—review & editing-Equal), and L. Renee Ruhaak (Conceptualization-Lead, Data curation-Equal, Investigation-Equal, Methodology-Equal, Supervision-Lead, Writing—original draft-Equal, Writing—review & editing-Equal)

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form.*

Research Funding: The ODYSSEY OUTCOMES trial was funded by Sanofi and Regeneron Pharmaceuticals. Cobas biomarker tests were provided by Roche Diagnostics Int. Ltd., Switzerland, whereas the multiplex quantitative clinical chemistry proteomics test was a laboratory-developed test from KCL/LUMC, Leiden, the Netherlands. This study was part of a research collaboration between LUMC and Agilent (Santa Clara, CA, United States).

Disclosures: J.W. Jukema and his department have received research grants from and/or was speaker on a.o. (CME accredited) meetings sponsored or supported by Abbott, Amarin, Amgen, Athera, Biotronik, Boston Scientific, Dalcor, Daiichi Sankyo, Edwards Lifesciences, GE Healthcare, Johnson and Johnson, Lilly, Medtronic, Merck-Schering-Plough, Novartis, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, the Netherlands Heart Foundation, CardioVascular Research the Netherlands (CVON), the Netherlands Heart Institute, and the European Community Framework KP7 Programme. C.M. Cobbaert is chair of the IFCC Scientific Division and chair of the EFLM Task Force Regulatory Affairs. L.R Ruhaak is chair of the IFCC working group for standardization of apolipoproteins by mass spectrometry.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: The authors express their gratitude towards the ODYSSEY OUTCOMES trial steering committee, especially the cardiology professors G.G. Schwartz, P.G. Steg; the biostatistician M. Szarek; and from Sanofi, G. Garon. In addition, the authors would like to thank J.A.M. Aalbers-Geerdink, H. Ahmed, R. Bahi, W. van der Bent, M. Bourik, E. Dantuma, P.A.M. Does, W. Endlich, E. Galjaard, A.P.M. Giliams, S.M.D. van Gog, A. Grummels, D. Kolle, J.L.M. van der Kroft, A. van der Laarse, M.L Luiten, R.N. Malhoe Mishre-Lalai, J.A.L. Mambi, F.A. van der Meer, S. Meivis, R.T.J. Niesthoven, C. Oolbekkink, G.J. Pieters, M.J.A. Rijks, I. Sabbar, G.C.J. van Steen, R. Swillens, D.I. Tromp, C.J. Trompert, T.J. Trompert, and R.C. van Wissen for their contribution.

REFERENCES

1. Reijnders E, van der Laarse A, Jukema JW, Cobbaert CM. High residual cardiovascular risk after lipid-lowering: prime time for predictive, preventive, personalized, participatory, and psycho-cognitive medicine. Front Cardiovasc Med 2023;10:1264319.
2. Libby P. The forgotten majority: unfinished business in cardiovascular risk reduction. J Am Coll Cardiol 2005;46: 1225–8.
3. Renee Ruhaak L, van der Laarse A, Cobbaert CM. Apolipoprotein profiling as a personalized approach to the diagnosis and treatment of dyslipidaemia. Ann Clin Biochem 2019;56:338–56.
4. Contois JH, Langlois MR, Cobbaert C, Sniderman AD. Standardization of apolipoprotein b, ldl-cholesterol, and non-hdl-cholesterol. J Am Heart Assoc 2023;12:e030405.
5. Langlois MR, Nordestgaard BG, Langsted A, Chapman MJ, Aakre KM, Baum H, et al. Quantifying atherogenic lipoproteins for lipid-lowering strategies: consensus-based recommendations from EAS and EFLM. Clin Chem Lab Med 2020;58:496–517.
6. Langlois MR, Chapman MJ, Cobbaert C, Mora S, Remaley AT, Ros E, et al. Quantifying atherogenic lipoproteins: current and future challenges in the era of personalized medicine and very low concentrations of LDL cholesterol. A consensus statement from EAS and EFLM. Clin Chem 2018;64:1006–33.
7. Reijnders E, van der Laarse A, Ruhaak LR, Cobbaert CM. Closing the gaps in patient management of dyslipidemia: stepping into cardiovascular precision diagnostics with apolipoprotein profiling. Clin Proteomics 2024;21:19.
8. van den Broek I, Romijn FP, Nouta J, van der Laarse A, Drijfhout JW, Smit NP, et al. Automated multiplex lc-ms/ms assay for quantifying serum apolipoproteins A-I, B, C-I, C-II, C-III, and E with qualitative apolipoprotein e phenotyping. Clin Chem 2016;62:188–97.
9. Ruhaak LR, Smit NPM, Suchiman HED, Pieterse MM, Romijn FPHTM, Beekman M, Cobbaert CM. MS-based proteomics: a metrological sound and robust alternative for apolipoprotein e phenotyping in a multiplexed test. Clin Chem Lab Med 2019;57:e102–4.
10. Ruhaak LR, Smit NPM, Romijn F, Pieterse MM, van der Laarse A, van der Burgt YEM, et al. Robust and accurate 2-year performance of a quantitative mass spectrometry-based apolipoprotein test in a clinical chemistry laboratory. Clin Chem 2018;64:747–9.
11. Horvath AR, Lord SJ, StJohn A, Sandberg S, Cobbaert CM, Lorenz S, et al. From biomarkers to medical tests: the changing landscape of test evaluation. Clin Chim Acta 2014;427:49–57.
12. Mathes T, Pieper D. An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews. Syst Rev 2019;8:226.
13. Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem 2012;58:1636–43.
14. Schwartz GG, Steg PG, Szarek M, Bhatt DL, Bittner VA, Diaz R, et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. N Engl J Med 2018;379: 2097–107.
15. Cobbaert C, Albersen A, Zwiers I, Schippers P, Gillis J. Designing a diagnostic total testing process as a base for supporting diagnostic stewardship. Clin Chem Lab Med 2020;59:473–89.
16. Tsantilas KA, Merrihew GE, Robbins JE, Johnson RS, Park J, Plubell DL, et al. A framework for quality control in quantitative proteomics. Preprint at https://doi.org/10.1101/2024.04.12.589318 (2024).
17. Bittremieux W, Meysman P, Martens L, Valkenborg D, Laukens K. Unsupervised quality assessment of mass spectrometry proteomics experiments by multivariate quality control metrics. J Proteome Res 2016; 15:1300–7.
18. Parker CE, Borchers CH. Mass spectrometry based biomarker discovery, verification, and validation—quality assurance and control of protein biomarker assays. Mol Oncol 2014;8:840–58.
19. Smit NPM, Ruhaak LR, Romijn F, Pieterse MM, van der Burgt YEM, Cobbaert CM. The time has come for quantitative protein mass spectrometry tests that target unmet clinical needs. J Am Soc Mass Spectrom 2021;32: 636–47.
20. Smit NPM, van den Broek I, Romijn FPHTM, Haex M, Deelder AM, van der Burgt YEM, et al. Quality requirements for quantitative clinical chemistry proteomics. Transl Proteom 2014;2:1–13.
21. Pichler P, Mazanek M, Dusberger F, Weilnböck L, Huber CG, Stingl C, et al. Simpatiqco: a server-based software suite which facilitates monitoring the time course of LC–MS performance metrics on Orbitrap instruments. J Proteome Res 2012;11:5540–7.
22. Dogu E, Mohammad-Taheri S, Abbatiello SE, Bereman MS, MacLean B, Schilling B, Vitek O. Msstatsqc: longitudinal system suitability monitoring and quality control for

targeted proteomic experiments. Mol Cell Proteomics 2017;16:1335–47.
23. Stanfill BA, Nakayasu ES, Bramer LM, Thompson AM, Ansong CK, Clauss TR, et al. Quality control analysis in real-time (QC-ART): a tool for real-time quality control assessment of mass spectrometry-based proteomics data. Mol Cell Proteomics 2018;17:1824–36.
24. Taylor RM, Dance J, Taylor RJ, Prince JT. Metriculator: quality assessment for mass spectrometry-based proteomics. Bioinformatics 2013;29:2948–9.
25. Clarke W, Molinaro RJ, Bachmann LM, Cook J. CLSI C62-ED2:2022 liquid chromatography-mass spectrometry methods. 2nd Ed. Wayne (PA): Clinical and Laboratory Standards Institute; 2022.
26. Diederiks NM, Ruhaak LR, Romijn F, Pieterse MM, Smit NPM, Cobbaert CM. An LC-MS-based designated comparison method with similar performance to the LP(a) reference measurement procedure to guide molar LP(a) standardization. Clin Proteomics 2024; 21:5.
27. CLSI, Parvin CA, Person NB, Baumann N, Duan L, Durham AP, et al. In: Parvin CA, editors. Statistical quality control for quantitative measurement procedures: principles and definitions. 4th Ed. Wayne (PA): Clinical and Laboratory Standards Institute; 2016. 74 p.
28. R Core Team. R: A language and environment for statistical computing. 4.2.2 Ed. Vienna, Austria: R Foundation for Statistical Computing; 2014.
29. Xie Y. Knitr: A general-purpose package for dynamic report generation in R. Version 1.33. Boca Raton (FL): Chapman and Hall/CRC; 2021.
30. Rstudio Team. Rstudio: integrated development for R. 2023.6.1.524 Ed. Boston (MA): RStudio, PBC; 2020.
31. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The fair guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018.
32. Briscoe CJ, Stiles MR, Hage DS. System suitability in bioanalytical LC/MS/MS. J Pharm Biomed Anal 2007;44: 484–91.
33. Levey S, Jennings ER. The use of control charts in the clinical laboratory. Am J Clin Pathol 1950;20:1059–66.
34. Anand U. The Levey–Jennings plot. Clin Chem 2013;59: 865–6.
35. Westgard JO. Internal quality control: planning and implementation strategies. Ann Clin Biochem 2003;40: 593–611.
36. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule shewhart chart for quality control in clinical chemistry. Clin Chem 1981;27:493–501.