First steps in the standardization of immunoglobulin IgG myeloperoxidase-anti-neutrophil cytoplasmic antibody measurements

D. P. Hutu,* E. Tuddenham,[†]
E. Monogioudi,* P. Meroni,[‡]
H. Schimmel,* J. Sheldon,[†] I. Zegers*
and the collaborating laboratories

Institute for Reference Materials and Measurements (IRMM), Joint Research Centre, European Commission, Geel, Belgium, [†]Protein Reference Unit and Immunopathology Department, St Georges' Hospital, London, UK, and [‡]Department of Clinical Sciences and Community Health and Istituto Auxologico Italiano, Milan, Italy

Summary

The standardization of immunoassays for immunoglobulin (Ig)G myeloperoxidase-anti-neutrophil cytoplasmic antibodies (MPO-ANCA) could contribute to a more accurate diagnosis and follow-up of small vessels-associated vasculitis, a systemic autoimmune disorder that leads to necrosis of blood vessel walls. Despite significant efforts by different groups, the level of comparability of results from commercially available immunoassays used for IgG MPO-ANCA detection is still poor. Therefore, the potential for improvement using reference materials was assessed. The evaluation of a set of 30 patient samples with 11 assays showed that differences between assays result in different interpretations for individual patients. Only 10 of 30 patient samples had the same clinical interpretation among 11 assays applying the cut-off values provided by each respective manufacturer. The correlation between results from 13 different assays was assessed in a pairwise manner. The correlation between results from patient samples was systematically very good for combinations of seven of those assays. The correlation of results ranged from reasonable to good for combinations with four other assays, therefore it should be possible to improve the comparability of results using a commutable reference material for calibration. Feasibility studies were conducted in order to find a reference material format most suitable for a calibrator. Two sets of candidate reference materials were produced from different raw materials, and assessed according to their suitability. A final format was selected, and a candidate reference material was produced.

Keywords: autoimmune disease, immunoassay, MPO-ANCA, reference material, standardization

Introduction

The detection of anti-neutrophil cytoplasmic antibodies (ANCAs) is crucial for the diagnosis of small vessel-associated vasculitis [1] such as microscopic polyangiitis (MPA), eosinophilic granulomatosis with polyangiitis (EGPA, formerly known as Churg–Straus syndrome [2]) and granulomatosis with polyangiitis (formerly Wegener's granulomatosis, GPA) [3]. In MPA and EGPA, ANCAs are directed mainly against the neutrophil enzyme myeloperoxidase (MPO). In GPA the ANCAs mainly target the neutrophil enzyme proteinase 3 (PR3). Routine laboratory investigation includes a screening test by indirect immunofluorescence (IIF) followed by confirmation with specific enzyme-linked immunosorbent assays (ELISA) for immunoglobulin (Ig)G antibodies to MPO and PR3 [4]. In recent years, several automated assays for the quantification of IgG MPO-ANCA (and IgG PR3-ANCA) have been developed, including multiplex-based assays [5], fluorescence immunoassay [6] and chemiluminescent assays [7].

Quantification of IgG MPO-ANCA (and IgG MPR3-ANCA) presents a significant problem, because there is a wide variation in the values generated for the same sample between different commercially available immunoassays as shown in external quality assurance schemes. This is the case even when values are reported in the same units. In the absence of any reference material (RM), each manufacturing company applies their arbitrary units to express the IgG MPO-ANCA concentration in patient serum, making it impossible to compare the concentrations measured with different assays.

Since 1998, the European Union Directive on In-Vitro Diagnostic Medical Devices (IVD-MD) (Directive 98/79/EC) requires the traceability of quantity values of calibrants and control materials to reference measurement procedures and/or reference materials of higher order. According to this legislation, all assay manufacturers must ensure that the *in-vitro* diagnostic devices they produce are calibrated properly against certified reference materials and reference measurement procedures, provided that such common references are available and suitable. A number of manufacturers produce kits that provide results traceable to the quantity value of the World Health Organization (WHO) international serum (MPO-ANCA, Human Reference Serum 15, code IS2720) available from the Center for Disease Control and Prevention (Atlanta, GA, USA). This material has not been tested for commutability, i.e. for its suitability, and has arbitrary values assigned to it (100 International Units).

The standardization of the measurement of complex biomarkers is a challenging task, but it has been shown to be feasible, for example, for the measurement of serum proteins such as IgG and IgM [8]. The first step in the process is to evaluate the degree of correlation between results from different assays; this should establish whether the various assays are measuring the same parameter or closely related parameters. Poor correlation between assays will make it impossible to use a common scale for the measurements [9].

The next step is the development of a reference system that can provide and maintain comparability of measurement results over a long time. This should deliver an anchor point through which the measurement results can be traced back, ideally to the International System of Units (SI). Inherent to this is a clear definition of the measurand being considered. In the case of complex biomarkers it may not be possible currently to give a structural (molecular) definition of the measured parameter; instead, a parameter may be defined by following a prescribed procedure (involving, for example, a chromatographic procedure, or a particular ELISA).

In December 2009, the International Federation for Clinical Chemistry and Laboratory Medicine (IFCC) formed a working group to collaborate with the Institute for Reference Materials and Measurements on the harmonization of autoimmune antibody testing. The first objective of this group was to establish whether it is possible to develop reference systems, including certified reference materials (CRMs) that have property values traceable metrologically to the International System of Units (SI) for selected autoantibody tests. The ultimate aim is to prepare CRMs that would be widely available for standardization of serology testing of several autoantibodies, among them IgG MPO-ANCA and anti-apolipoprotein H (formerly beta 2-glycoprotein) [10].

IgG MPO-ANCA was selected as the first autoantibody to be considered and the processes used were assessed critically to form the model system for the development of CRMs for further autoantibodies. An important requirement for a CRM is commutability, i.e. the CRM has to resemble patient samples in order to minimize calibration bias [11,12]. Therefore, two commutability studies were performed, comparing the results for patient samples and candidate reference materials across various assays. In the first study the comparability of results from a selected number of assays was assessed, and different raw materials and CRM formats were evaluated. This study was performed in a single laboratory, at St Georges' Hospital. A second study was performed in order to validate further the most promising candidate CRMs, using a larger number of assays. The measurements for the second study were performed by the manufacturers of the assays using identical samples (split sample study setup).

This paper presents the results of correlation studies for IgG MPO-ANCA assays, and investigates the results in analytical terms (e.g. repeatability, calibration, scales). Results from feasibility studies for the development of a reference system for anti-MPO ANCA serology testing are presented.

Materials and methods

Serum samples

Patient samples were selected purely on the basis of the fact that IgG MPO-ANCA measurements had been requested from the laboratory, and that sufficient sample volume was available. A set of 29 serum samples was used in the first commutability study, whereas 30 samples were used in the second study. Serum samples were routine samples from patients undergoing IgG MPO-ANCA testing at the Protein Reference Unit and Immunopathology Department, St Georges' Hospital. Each sample was anonymized in accordance with national ethical laws.

Candidate reference materials processing

The raw materials used were three plasmapheresis materials, collected from patients with high concentrations of IgG MPO-ANCA. The materials were collected at the Staten Serum Institute, Denmark and coded as SSIB, SSIE and SSIF. Table 1 summarizes the different candidate reference materials that were produced on the basis of these three materials.

Plasma conversion into serum. Aliquots (500 ml) of the three plasmapheresis materials were thawed and warmed to 37°C. Each was treated with a 1/100 volume of 5 mg/ml protamine sulphate solution (Sigma-Aldrich, Seelze, Germany). The material was incubated for 1 h at 37°C followed by a second incubation at 4°C for 40 h in order to allow the formation of fibrin. The fibrin was removed by

Table 1. Candidate reference material formats

1st study, all materials both stored lyophilised and liquid frozen
- SSI B plasma
- SSI B plasma converted to serum
- SSI E plasma converted to serum
- SSI F plasma converted to serum
- SSI serum B with NaN₃
- SSI serum B processed (delipidated, with additives)
- IgG MPO-ANCA purified from SSI B, spiked into processed serum
- IgG MPO-ANCA purified, spiked into filtered unprocessed serum
- IgG MPO-ANCA purified, spiked into filtered unprocessed serum with additives

2nd study
- SSIB serum (MPO-4)
- SSIB processed serum (delipidated, with additives, MPO-6)
- SSIB processed serum diluted 1 : 6 in serum from healthy donors (MPO-6D1)
- SSIB processed serum diluted 1 : 6 in an albumin solution (MPO-6D2)
- IgG MPO-ANCA purified, spiked into filtered unprocessed serum (MPO-7)

Ig = immunoglobulin; MPO-ANCA = myeloperoxidase-anti-neutrophil cytoplasmic antibodies.

centrifugation during 10 min at 12 000 g and the supernatant was further filter-sterilized using a 0·22 μm PES filter (Corning Incorporated, Midland, NC, USA).

Serum processing. The procedure for serum processing was based on that used for the CRM ERM-DA470k/IFCC [13], with some modifications. NaCl (final concentration 50 g/l, Sigma-Aldrich) was added to the defibrinated serum and the pH was adjusted to 8·5 with a saturated tris(hydroxymethyl)aminomethane (Tris) solution (Sigma-Aldrich). The lipids present in serum were removed by incubation with synthetic amorphous silica. Protein precipitates and silica particles were removed by centrifugation for 30 min at 10 000 g and the clear supernatant was dialysed further in isotonic NaCl solution (0·9% w/v) (Sigma-Aldrich). The dialysis was completed after four changes of the dialysis solution over 24 h. Subsequently, the pH was adjusted to 7·2 with a 100 mM 4-(2-hydroxyethyl)-1-piperazineethane-sulphonic acid (HEPES) solution (Sigma-Aldrich) and the following preservatives were added: sodium azide (final concentration of 0·95 g/l), aprotinin (final concentration of 61·5 TIU/ml) and benzamidine hydrochloride monohydrate (final concentration of 1 mM). All three preservatives were purchased from Sigma-Aldrich. Furthermore, a sterile filtration with a 0·22 μm PES filter (Corning Incorporated, Glendale, AZ, USA) was performed.

In order to produce the candidate reference materials for the second commutability study, the processed serum was diluted further 1 : 6 (v/v) with either Zenalb® 4·5 (Bio Products Laboratory, Herts, UK) or pooled serum collected from healthy donors.

Purification of IgG MPO-ANCA. IgG MPO-ANCA was purified from SSIB, the plasmapheresis material with the highest titre of IgG MPO-ANCA of the three donations. Two chromatographic techniques were combined: immunoaffinity chromatography and size exclusion chromatography (SEC). The column used for immunoaffinity was prepared in-house using commercially available purified human myeloperoxidase (Scipac, Sittingbourne, UK) coupled covalently to CNBr activated Sepharose 4B (GE Healthcare, Uppsala, Sweden), according to the protocol provided by the manufacturer. SSIB plasma diluted with Tris-buffered saline (TBS) solution pH 8 (Sigma-Aldrich) in a 1 : 1 (v/v) ratio was passed over the affinity column at a flow rate of 0·5 ml/min in order to allow antigen–antibody complex formation. After washing the column the bound antibodies were eluted by a brief exposure to a 100 mM glycine solution pH 2·7 (Sigma-Aldrich). In order to neutralize the pH of the eluted fractions, a 1/10 volume of 1·5 M Tris-HCl pH 8·8 solution (Bio-Rad Laboratories, Hercules, CA, USA) was added to each collected fraction. In order to enhance the purity, all fractions were pooled, concentrated to a final volume of 500 μl using Amicon Ultra centricons with a 30 kDa cut-off (Millipore, Billerica, CA, USA) and applied subsequently on a Superdex 200 10/300 GL column (GE Healthcare). The purification efficiency was evaluated by analysing purified IgG MPO-ANCA present in each fraction on 10% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE; Bio-Rad Laboratories) and fluorescent staining with Syproruby (Invitrogen, Waltham, MA, USA).

Filling and freeze-drying. One ml serum was filled into uncoloured threaded and siliconized glass vials. For all types of materials vials were stored both with materials liquid-frozen (−80°C) and freeze-dried and frozen at −80°C. In the latter case, freeze-drying was started within 1 h after filling.

IgG MPO-ANCA assays used

Nine commonly used immunoassays were selected for the first commutability study and 11 immunoassays were used in the second commutability study. Six assays were identical in both studies. All assays are listed in Table 2.

Performance of the measurements

Due to the large number of candidate reference material formats and clinical samples, the IgG MPO-ANCA measurements in the first commutability study were performed as follows: the candidate reference materials that had been stored liquid-frozen and 29 clinical samples were measured in duplicate on one ELISA plate, while the candidate reference materials that had been lyophilized and reconstituted were measured in duplicate alongside the clinical samples on a separate plate. Reconstitution was performed by the procedure specified in the certificate of ERM-DA470k/IFCC.

Table 2. Immunoassays used in the commutability studies.

1st study			
Manufacturer	Assay	Principle	Name
Eurodiagnostica	DIASTAT™ Anti-MPO	ELISA	Diastat
Eurodiagnostica	Wieslab® Capture MPO-ANCA	ELISA	Wieslab C
Eurodiagnostica	Wieslab® MPO-ANCA	ELISA	Wieslab D
Euroimmun AG	Anti-MPO ELISA	ELISA	Euroimmune
Inova	QUANTA Lite® MPO SC	ELISA	Quanta Lite
Menarini Diagnostics	Myeloperoxidase antibody	ELISA	Menarini
Orgentec	ORG 519 Anti-MPO	ELISA	Orgentec
Phadia GmbH	EliA MPO^S	FEI	Phadia EliA
Phadia GmbH	Varelisa™ MPO ANCA	ELISA	Varelisa

2nd study			
Manufacturer	Assay	Principle	Name
Aesku Diagnostics	AESKULISA® MPO	ELISA	Aesku
Bio-Rad Laboratories	Anti-MPO EIA	ELISA	Bio-Rad EIA
Bio-Rad Laboratories	BioPlex 2200	MFI	BioPlex 2200
Eurodiagnostica	Wieslab® Capture MPO-ANCA	ELISA	Wieslab C
Euroimmun AG	Anti-MPO ELISA	ELISA	Euroimmun
IMMCO Diagnostics	Anti-MPO ELISA	ELISA	IMMCO
INOVA Diagnostics	QUANTA Lite® MPO IgG	ELISA	Quanta Lite
INOVA Diagnostics	MPO BIOFLASH	CI	Bioflash
Orgentec GmbH	ORG 519 Anti-MPO (pANCA)	ELISA	Orgentec
Phadia GmbH	EliA MPO^S	FEI	Phadia EliA
Phadia GmbH	Varelisa™ MPO ANCA	ELISA	Varelisa

MPO = myeloperoxidase; ELISA = enzyme linked immunosorbent assay; FEI = fluorescence enzyme immunoassay; MFI = multiplex flow immunoassay; CI = chemiluminescent immunoassay.

Additionally, all formats of candidate reference materials and 20 clinical samples were measured in duplicate on a third ELISA plate. For measurements performed on the automated platform (Phadia EliA), all candidate reference materials and clinical samples were measured in triplicate. All reconstitution and dilution volumes were controlled gravimetrically, and the dilution levels were calculated from the masses and the densities rather than from the intended volumes. All measurements were performed at the Protein Reference Unit and Immunopathology Department, St Georges' Hospital, UK.

In the second commutability study 30 clinical samples and the five candidate reference materials were measured in duplicate each on three different plates when measured by ELISA, and in triplicate when measured on automated platforms. The measurements were performed by the manufacturers' laboratories according to their validated procedures.

### Immunoassay response upon dilution of candidate reference materials

The linearity of the assay responses for dilutions of the candidate reference materials was evaluated in the second study. Four dilutions (10, 25, 50 and 75%) (w/w) were measured, together with the undiluted materials, in duplicate. Separately a 33·3% (w/w) dilution (using the kit diluent) of each clinical sample was measured in duplicate alongside the undiluted clinical samples.

Data analysis

In both studies, the measurement results that were outside the measuring interval of each individual assay were excluded. Data analysis was performed using Analyse-it software (Analyse-it software, Leeds, UK).

Commutability was assessed by comparing the results for candidate reference materials with the results obtained for clinical samples [11]. For each sample the average values per plate were calculated and used for the analysis. First, Pearson's correlation coefficient was calculated for the results for patient samples for all pairs of assays. For the pairs with a correlation coefficient higher than 0·75 a linear regression was applied and the 95% prediction interval calculated. Samples outside the prediction interval were excluded, and a new linear regression performed. A candidate reference material or dilution thereof was considered commutable when its results were within the 95% prediction interval of this regression performed on results from patient samples.
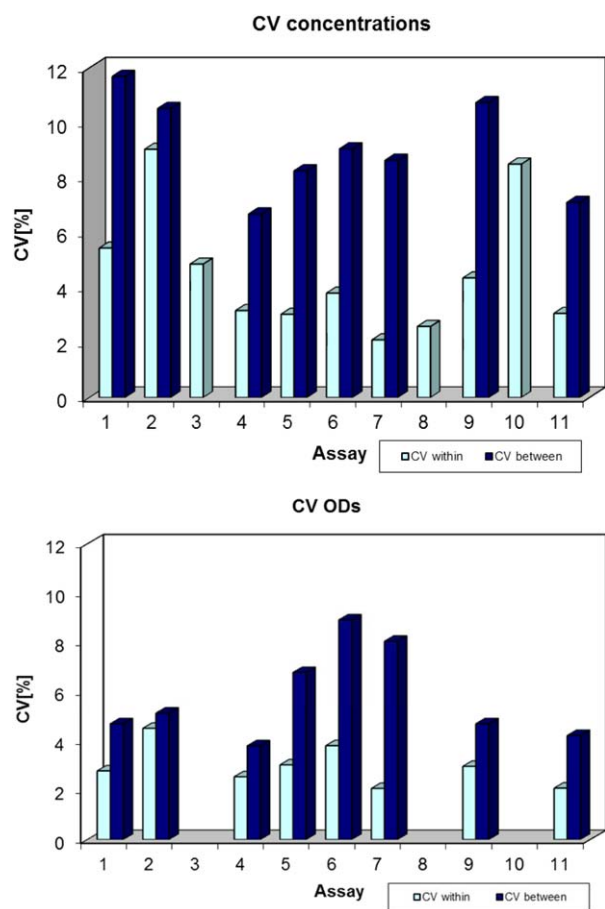
Fig. 1. Coefficients of variation (CVs) within plate or run and between plate or run for the immunoassays used in the second commutability study. Presented are the coefficients of variation within plates (shown as light bars) and within plates (shown as dark bars). Results for ODS are only shown for enzyme-linked immunosorbent assay (ELISA)-based assays.

Results

Assay performance

Evaluation of the assay performance was not the main objective of this work. However, the repeatability, intermediate precision, sensitivity for interferences and particularly the assay selectivity have an impact upon the degree of equivalence that can be achieved upon the development and use of a reference system.

The repeatability and intermediate precision of each assay were assessed by calculating the coefficient of variation (CV) of averages within and between plates for all patient samples. For the measurements performed by ELISA CVs were calculated both for the absorbance signal produced by the ELISA and the concentrations calculated from the absorbances. Results from the second commutability study are shown in Fig. 1. The results from absorbances showed a low intraplate variation for all ELISA assays. The interplate variation was higher, but well below 10% for all assays. When the absorbances were con-

verted into concentrations higher CVs were observed. However, all within-assay, between-plate CVs were below 12% for all assays. Similar results were observed for the measurements performed in the first commutability study (data not shown).

The higher CVs for concentration values seem to be due, to a considerable extent, to the fact that assays have been optimized for detection of IgG MPO-ANCA. In order not to miss weakly positive samples the assays are designed in such a manner that the measurement interval includes the upper part of the response curve, where a small change in signal corresponds to a large change in concentration. For samples with high concentrations of IgG MPO-ANCA, a small analytical variation has a large impact upon the concentration value determined.

A second reason for higher variations in between-plate concentration values than in absorbances is the manner in which the calibration is performed. Most of the assays used a six-point calibration curve, but others used calibration curves consisting of either three or five points. Moreover, it was observed that few calibration points are present in the upper range of the measuring interval, leading to less accurate concentration estimation in that part.

Immunoassay response for dilutions of clinical samples

Therefore, the average recovery was evaluated for clinical samples for dilutions with relative concentrations of 0·33. Results were obtained for nine assays, as summarized in Table 3. The correlation of the results for the dilutions *versus* those for the undiluted materials was good for all assays (R^2 from 0·94 to 0·99). Three samples were found to be outliers in most assays. This means that a majority of the clinical samples have the same dilution behaviour within each assay, at least for this dilution. The recovery was sometimes significantly higher than 100, which indicates a non-linear response curve and calibration problems.

Correlation between results from different assays

The degree of correlation of results from different assays was evaluated by comparing results from all assays two × two. In the first commutability study the Pearson product–moment correlation coefficient (r) was calculated for 28 assay pairs. Results from one assay were not evaluated due to the small number of values within the measuring interval (Menarini). The interpretation of correlation coefficients was based on clinical biostatistics rules [14], according to which r-values ranged from 0.75–1 points to a very good to excellent correlation. As can be seen in Table 4a, the correlation coefficients for all assay comparisons in the first commutability study ranged from 0·56 to 0·95. For 18 assay pairs the value of r was above 0·75, indicating a very good correlation for combinations among five assays. Moreover, seven of these comparisons had an r-value above 0·9, indicating an excellent correlation of the measurement

Table 3. Dilution behaviour of clinical samples.

Assay	Outliers removed	Slope of values for diluted *versus* undiluted samples	R^2	Average recovery for dilutions (%)
Aesku	2	0·32	0·94	88
Bio-Rad EIA	2	0·27	0·97	85
Wieslab C	2	0·25	0·93	115
Euroimmun	1	0·58	0·95	122
IMMCO	1	0·38	0·98	116
Quanta Lite	2	0·46	0·97	159
Bioflash	1	0·37	0·97	136
Orgentec	2	0·35	0·95	115
Varelisa	3	0·38	0·99	115

results. However, it was also obvious that there are individual patient samples that behaved differently from the majority of samples and were outliers in many of the comparisons.

In the second study the degree of correlation was assessed by calculating *r* for 55 assay pairs. The correlation coefficients ranged from 0·29 to 0·98 (Table 4b). For 34 of 55 assay pairs the *r*-value was above 0·75, and for 10 assay pairs it was above 0·9.

The correlation coefficient describes the strength of the linear relationship between results from two assays, but does not detect it if there is a constant or proportional difference between two assays. Visual inspection of the plots was per-

formed in order to verify the form of the relationship between assays in each pair. The relationship between results from different assays is, in many cases, not linear. For example, in Fig. 2c the relationship for Quanta Lite *versus* Orgentec for the candidate reference material is curved, and so is the relationship between the results for clinical samples. This is in agreement with the finding that the values obtained for dilutions of the candidate reference materials is curved for Quanta Lite and linear for Orgentec (Fig. 3). However, no attempt was made to find the best curve fit in this study, as the number of patient samples was not appropriate.

Classification and interpretation of measurement results

The measurement scales used by the assays are very different. In some cases this is also reflected in the use of different units (U, IU, CU, RU, AI). In the second study numerical differences up to factor 5 were found between assay averages for patient samples (as evidenced from the slopes of regression lines for results from different assays) even when using the same units. What is more problematic is that these differences in values and differences in decision limits result in different interpretations of the results for individual patients. Results of the second study were classified as being negative, equivocal or positive for IgG MPO-ANCA by applying the cut-offs provided by the different manufacturers. Only 10 of 30 samples had the same classification in all assays. Nine samples were classified as being positive by all assays, while only one sample was found to be negative in all assays.

Table 4. Correlation between results for clinical samples.

(a) 1st study							
	Wieslab C	Phadia EliA	Euroimmun	Varelisa	Orgentec	Quanta Lite	Diastat
Wieslab D	0·78	0·91	0·76	0·65	0·91	0·56	0·66
Wieslab C		0·72	0·72	0·77	0·73	0·74	0·63
Phadia EliA			0·72	0·76	0·98	0·77	0·58
Euroimmun				0·95	0·87	0·85	0·90
Varelisa					0·76	0·94	0·85
Orgentec						0·86	0·75
Quanta Lite							0·83

(b) 2nd study									
	Wieslab C	Phadia EliA	Euroimmun	Varelisa	Orgentec	Quanta Lite	IMMCO	Bio-Rad EIA	Bioflash	Aesku
BioPlex 2200	0·29	0·71	0·58	0·60	0·55	0·79	0·69	0·65	0·71	0·39
Wieslab C		0·69	0·74	0·84	0·74	0·77	0·55	0·65	0·59	0·79
Phadia EliA			0·83	0·88	0·90	0·94	0·79	0·95	0·90	0·84
Euroimmun				0·79	0·90	0·84	0·66	0·79	0·65	0·80
Varelisa					0·83	0·95	0·79	0·90	0·85	0·87
Orgentec						0·88	0·74	0·89	0·79	0·95
Quanta Lite							0·85	0·93	0·87	0·89
IMMCO								0·88	0·81	0·72
Bio-Rad EIA									0·92	0·84
Bioflash										0·74
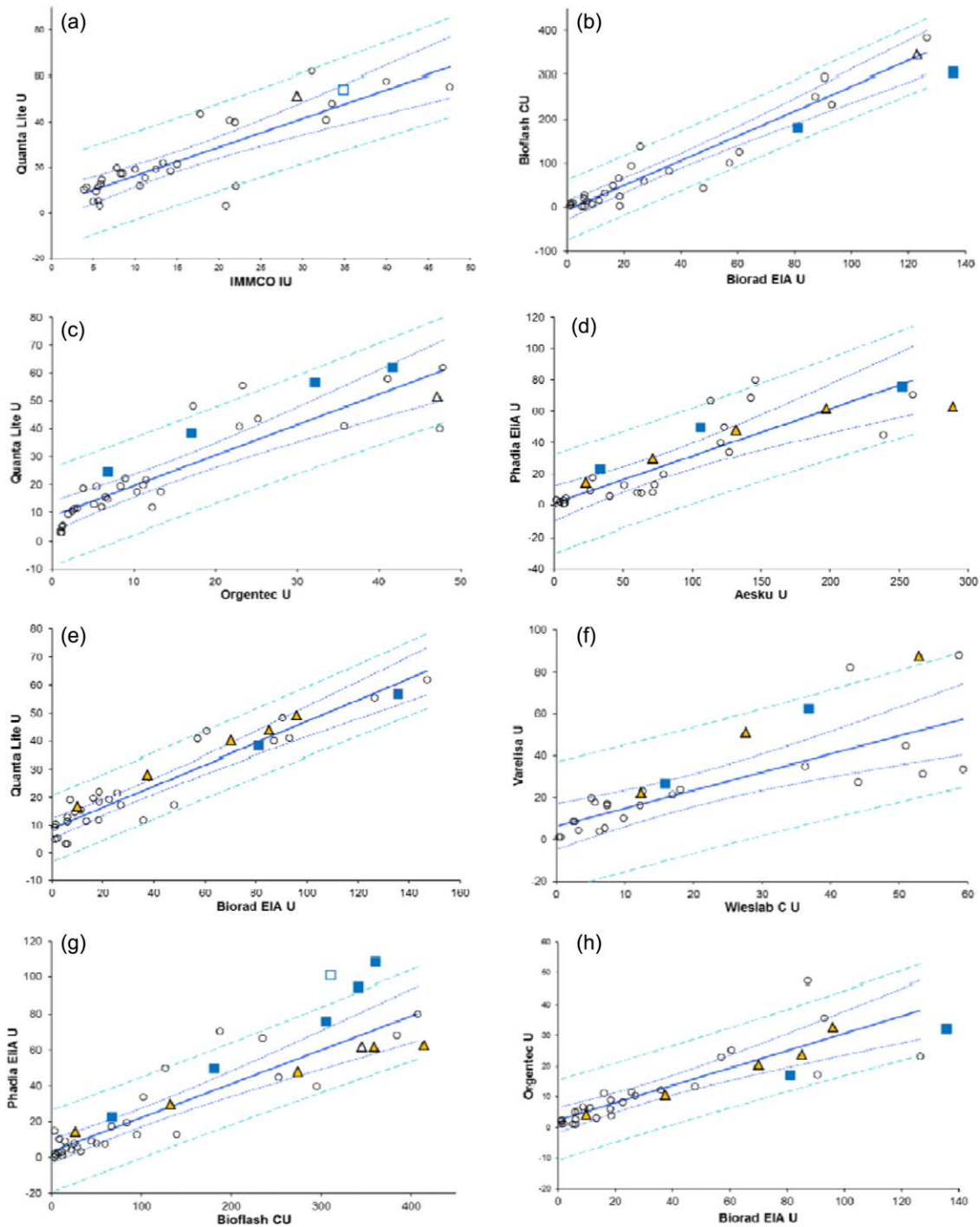
Fig. 2. Examples of results for the commutability study. Results for clinical samples are shown as circles, results for undiluted myeloperoxidase MPO-7 as an empty triangle, results for undiluted MPO-6D2 as an empty square, results for dilutions of MPO-7 as filled triangles and results for dilutions of MPO-6D2 as filled squares. Example (a) shows a case where both undiluted RMs are commutable. (b,c) The situation where the undiluted MPO-7 is commutable, the undiluted MPO-6D2 is out of range, but dilutions of MPO-6D2 are commutable. (e,d,h) All dilutions of MPO-7 and MPO-6D2 in the measurement range are commutable. In (f), one dilution of MPO-7 is not commutable, in (g), undiluted MPO-6D2 is not commutable.

Potential for harmonization of IgG MPO-ANCA measurements

The degree of comparability that could potentially be achieved by using single-factor recalibrations of the assays was evaluated on the basis of the results from the second study. Correction factors were calculated for all assays, and applied to the results for the clinical samples. The correction factors were based on the slope and bias obtained for
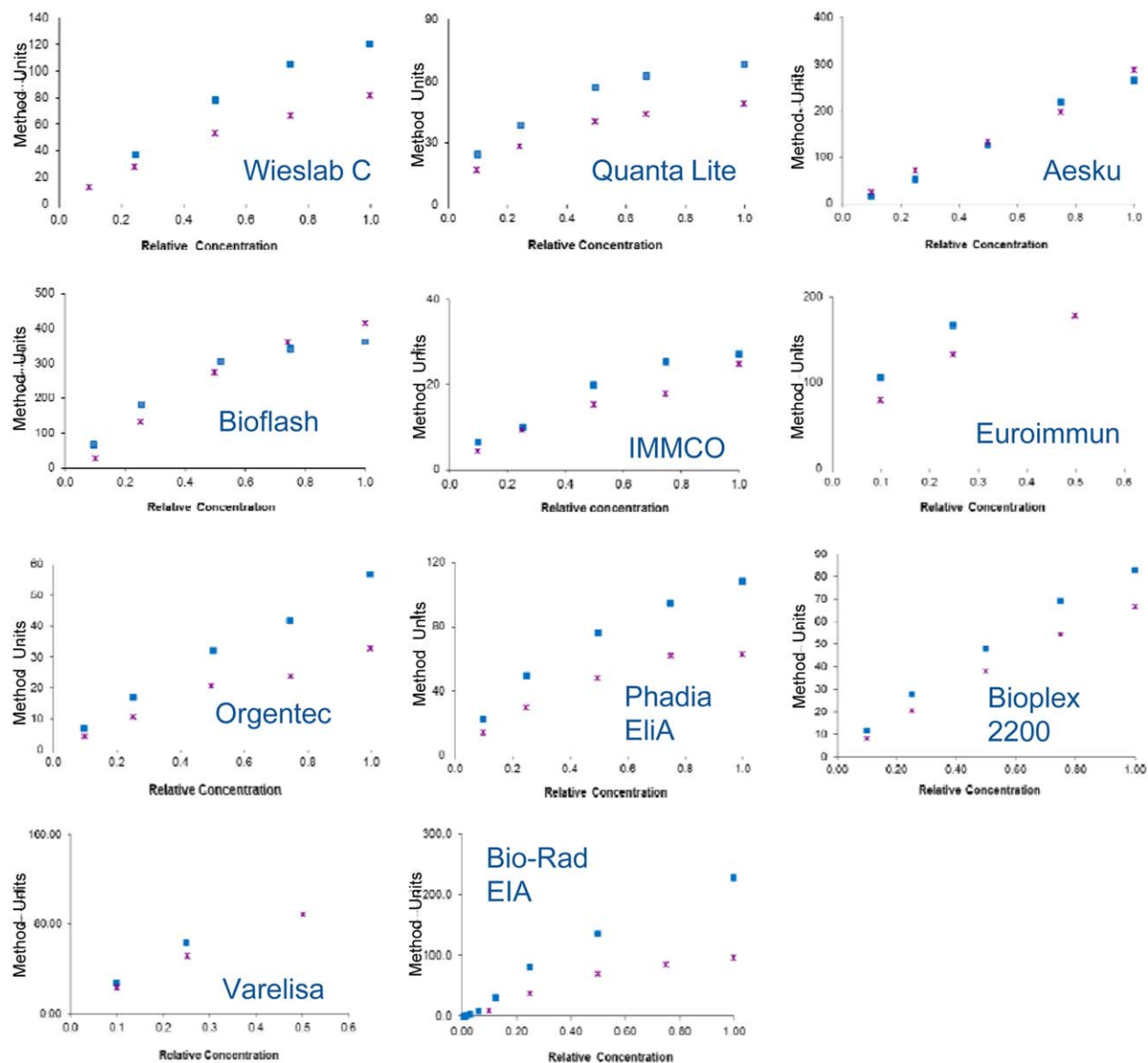
Fig. 3. Results from the dilution studies for candidate reference materials myeloperoxidase (MPO)−6D2 (■) and MPO-7 (✗).

the Passing–Bablok regression of the results of individual assays with respect to the assay averages. Figure 4 shows the data before and after applying the correction factors for the seven assays for which correlation factors are higher than 0·75 (Phadia EliA, Varelisa, Orgentec, Quanta Lite, Bio-Rad EIA, Bioflash, Aesku). Before mathematical recalibration the average between-assay CV for the clinical samples is 97% for all assays and 87% for the seven best correlating assays. After applying the correction factors the average CV is reduced to 47% for all assays and 32% for the seven best-correlating assays. It should be taken into account that this approach is very basic; it does not take into account non-linearity in the relation between results from different assays, and samples with large variability are not excluded, possibly causing bias. It should be possible to improve this

further on the basis of a more careful recalibration. After the recalibration we evaluated the effect of using a common cut-off for the interpretation of the results (positive/nega-tive). Fourteen samples had the same interpretation when all assays were considered, and 20 samples had the same interpretation when the seven best-correlating assays were considered.

Thus, the between-assay variability was decreased, but remained significant with respect to the within-assay vari-ability (<10%). It is clear that a considerable part of the remaining variability is due to the different behaviour of individual samples in the different assays, i.e. differences in assay selectivity or susceptibility to interference. In terms of the clinical interpretation of the results (positive, equivocal, negative) there is not one assay that is really an outlier.
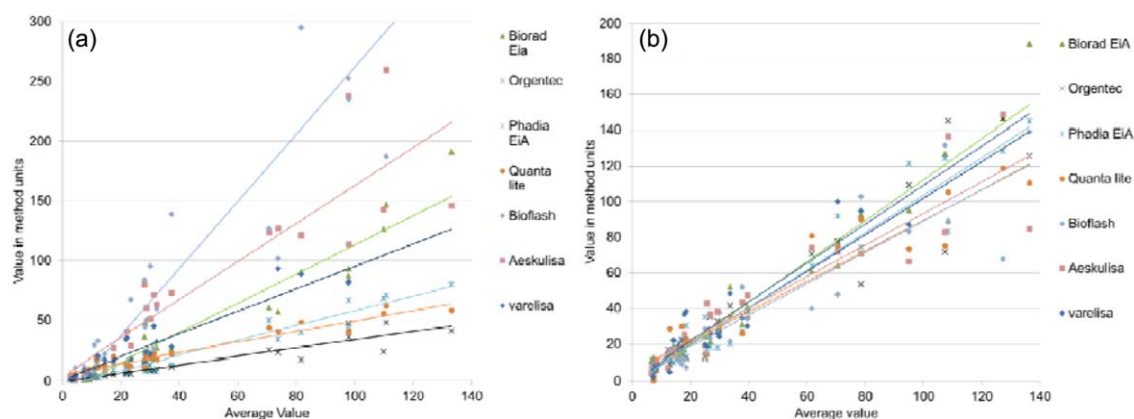
Fig. 4. Evaluation of the feasibility of the harmonization of immunoglobulin IgG myeloperoxidase-anti-neutrophil cytoplasmic antibodies (MPO-ANCA) assays. (a) Results for 30 clinical samples obtained with seven assays plotted against the average value per sample. (b) Results for 30 clinical samples after applying a correction factor for each assay based on the slope of the Passing–Bablok regression for each individual assay with respect to the average value per sample.

There is also not a clear pattern for the samples that give varying results, i.e. it is not always the same assay that gives a positive or negative bias. Therefore, the use of a common calibrant should make it possible to put the assays on the same scale and improve the comparability of results, but not all variability for individual patient samples can be removed.

A different type of study would be needed in order to study the reason for poor correlation between some assays. A better understanding of the link between the biochemical behaviour of the samples and the clinical background could help in making anti-MPO tests more useful. However, this would require the analysis of a larger number of clinical samples with known clinical background.

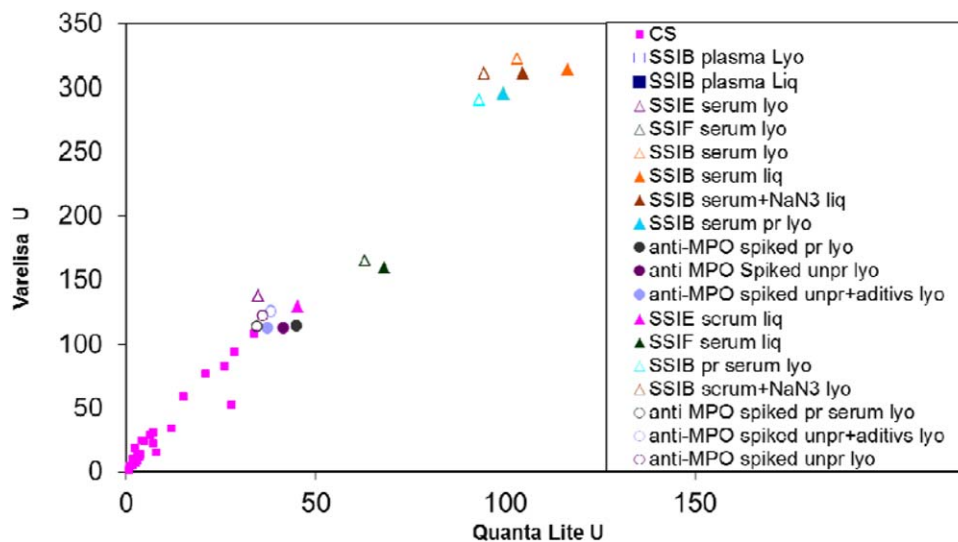### Development of candidate reference materials

In order to be used as a calibrator, a reference material has to fulfil some criteria: the batch of the material must be homogeneous, the material must be stable over a long time-period and, most importantly, the material must be commutable with clinical routine samples [11]. Serum reference materials certified for human IgG content have been produced successfully in the past at the Institute for Reference Materials and Measurements (IRMM) [13]. The processing procedures optimized for total IgG served as a basis for the development of candidate reference materials for IgG MPO-ANCA.

Two different types of starting materials were used: plasmapheresis materials collected from patients who underwent plasmapheresis treatment and affinity-purified IgG MPO-ANCA. A total of 18 candidate reference material types were produced from the starting materials (Table 1): one plasmapheresis material as such (SSIB), three serum materials produced from plasma collected from three different patients (SSIB, SSIE and SSIF), one serum material with sodium azide added, one processed patient serum

(SSIB), two sera collected from healthy subjects spiked with purified IgG MPO-ANCA, respectively, containing or not containing additives, and one processed serum spiked with purified IgG MPO-ANCA. As lyophilized formats have certain advantages in terms of storage and distribution logistics as well as for the material stability, the candidate reference materials were each prepared in both liquid and lyophilized form. All materials were stored at −70°C after processing.

Figure 5 shows one comparison of the results for clinical samples and candidate reference materials measured with Quanta Lite and with Varelisa. The values for material SSIB and different processed forms of SSIB are high; in fact, above the upper limit of the set of clinical samples used in the study. Results for the plasmapheresis materials SSIE and SSIF, as well as for the spiked material, were more in the middle of the results for the clinical samples. Freeze-drying has little influence on the values; values of couples of materials that were, respectively, stored frozen-liquid or freeze-dried and when reconstituted gave very similar results. The freeze-dried materials systematically give a somewhat higher value, which can be explained by the fact that the volume of water added in order to reconstitute the freeze-dried materials is not exactly the same as the amount of water removed by freeze-drying. Similarly, processing of the serum (mainly delipidation and dialysis) and the addition of preserving agents has no significant effect on the measured values. The same general trend was found for other combinations of assays, leading to the conclusion that the different processing steps do not affect significantly the activity of the IgG MPO-ANCA as measured with the different assays.

The suitability of processed materials was verified further in the second study. Five candidate reference materials were produced: SSIB serum (material code MPO-4), SSIB processed serum (material code MPO-6), SSIB processed

Fig. 5. Example of results from the first commutability study. Plotted are the measurements results for clinical samples (CS) and candidate reference materials for immunoglobulin (Ig)G myeloperoxidase-anti-neutrophil cytoplasmic antibodies (MPO-ANCA). Patient samples are shown as pink squares; liquid-frozen and lyophilized candidate reference materials based on processed serum are shown as triangles; liquid-frozen and lyophilized candidate reference materials based on purified IgG MPO-ANCA spiked into different serum matrices are shown as circles.

serum diluted 1 : 6 (v/v) in serum collected from healthy donors (material code MPO-6D1), SSIB processed serum diluted 1:6 (v/v) in Zenalb® 4·5 (Bio Products Laboratory, Hertfordshire, UK), a 45 g/l of human albumin solution used for infusion (MPO-6D2) and one serum collected from healthy individuals spiked with purified IgG-MPO-ANCA (material code MPO-7). The first four candidate reference materials were produced in a lyophilized form, whereas the fifth was produced in a liquid frozen form.

The discussion of the results for the reference material will focus upon those for MPO-6D2 and MPO-7. The reason is that MPO-6D2 presents a format that is very similar to what the final certified reference material is expected to be (processed, stabilized serum produced from high-titre plasmapheresis material, with a concentration at the upper end of the measurement interval of most assays), while MPO-7 is the type of calibration solution that would be produced by the Joint Research Centre (JRC-IRMM) for the value assignment of the certified reference material (serum spiked with affinity-purified IgG MPO-ANCA).

Immunoassay response for dilutions of candidate reference materials

If reference materials are used for calibration it is important that dilutions of the materials follow the response curve for clinical samples in the different assays to avoid bias. This can be tested by comparing the response curve for dilutions of the candidate RMs and clinical samples, or by comparing the response for dilutions of RMs to the response for different clinical samples in assay comparisons. In Fig. 3 the results for dilutions are shown for MPO-6D2 and MPO-7. For certain assays the dilution profile is highly linear (Aesku,

Orgentec, Bioplex). For others it is curved (e.g. Quanta Lite, Bioflash, Phadia EliA). This different behaviour (linear *versus* non-linear) in the different assays is not necessarily a problem if the profile corresponds to that found for clinical samples with different concentrations of IgG MPO-ANCA. The results from the commutability studies (see below) indicate that this is the case.

Commutability of candidate reference materials

In the first study the concentration of anti-MPO IgG in the candidate reference materials based on SSIB was very high compared to the clinical samples, and above the upper limit of the measurement interval of most of the assays used. Therefore, the RM samples were subjected to an additional 1 : 2 (v/v) dilution in assay diluent. As the values for the candidate reference materials were still at the extreme of the interval covered by clinical samples, the design was not ideal for assessing commutability. Therefore, although the results looked promising (see Fig. 5), data were not evaluated numerically in terms of commutability. Instead, a second commutability study was performed using a more limited selection of candidate reference materials with lower concentrations of IgG MPO-ANCA.

In the second commutability the samples were measured with four different types of assays: eight manual ELISAs, one multiplex based assay, one fluorescence assay and a chemiluminescence assay.

One clinical sample gave very high values, out of range, for eight of 11 assays, and the results were therefore not used. For two assays, Euroimmune and Varelisa, the concentrations of undiluted reference materials were too high, and no direct information on commutability could be

Table 5. Summary of results for the assessment of the commutability*.

	Wieslab C	Phadia EliA	Euroimmun	Varelisa	Orgentec	Quanta Lite	IMMCO	Bio-Rad EIA	Bioflash	Aesku
Bioplex 2200	–	–	–	–	–	–	–	–	–	–
Wieslab C		–	–	NC	–	–	–	–	–	–
Phadia EliA			dC	dC	7C	7C	–	dC	7+, 6D2–	dC
Euroimmun				–	NC	dC	–	–	–	dC
Varelisa					dC	dC	–	dC	dC	dC
Orgentec						C	–	dC	–	7C
Quanta Lite							C	dC	7C	dC
IMMCO								7C	7–, 6D2+	–
Bio-Rad EIA									7C	dC
Bioflash										–

*Summary of the results for the assessment of the commutability for reference materials (RMs) myeloperoxidase (MPO)−6D2 and MPO-7 or dilutions thereof. A dash (–) indicates that commutability was not assessed, as the Pearson's correlation coefficient was below 0·8. C indicates commutability of both MPO-7 and MPO-6D2. When the undiluted candidate RMs were outside the measurement interval dC indicates commutability for all dilutions in the measurement interval. '7+, 6D2−' indicates commutability for MPO-7 and non-commutability for MPO-6D2, and '7−, 6D2+' indicates the inverse. NC indicates non-commutability of the undiluted materials, and 7C indicated that undiluted MPO-7 was commutable, as well as dilutions of MPO-6D2.

obtained for these assays. This is a consequence of the fact that these assays have been optimized for sensitivity. The commutability of candidate reference materials and of dilutions was verified for the 27 combinations of assays where the Pearson's correlation coefficient was 0·8 or higher. This cut-off was chosen because at correlation coefficients below 0·8 the prediction interval for the clinical samples is so broad that the fact that the candidate reference material is found within this interval is no longer meaningful in terms of the required assay performance.

Figure 2 shows representative examples of results from the second commutability study. For 10 assay comparisons the values of the undiluted reference material MPO-7 was within the measurement intervals for both assays (Table 5). In nine of these cases both MPO-7 was found to be commutable, i.e. the results for the candidate reference materials were within the 95% prediction interval of the linear regression for the clinical samples (e.g. Fig. 2a,b,c). In five comparisons the undiluted MPO-6D2 was within the measurement interval. In four of those cases it was commutable (e.g. Fig. 2a). In 17 cases the commutability of dilutions was evaluated. For 15 of these 17 comparisons the dilutions that were in the measurement interval were all found to be commutable (e.g. Fig. 2d,e,h). In two of these 17 comparisons one or more of the dilutions were outside the 95% prediction interval (e.g. Fig. 2f,g).

Commutability can also be analysed with other approaches, for example using a Passing–Bablok or Deming regression, and evaluation of the results for the reference material with respect to the 95% prediction interval. These approaches gave the same overall picture, with some deviations due mainly to the calculation of the prediction interval being different.

Overall, the commutability of the two reference materials MPO-6D2 and MPO-7, including their dilutions, is acceptable. In a few cases a candidate reference material or a dilution is found to be non-commutable, but this was never consistent. Using a 95% prediction interval means that in a certain percentage of cases the candidate reference material is expected to be found non-commutable even if it behaves in the same way as patient samples.

MPO-6D2 is based on serum produced from a plasmapheresis material that was further processed in the same way as the reference material for IgG, ERM-DA470/IFCC. The latter material is known to be stable for at least 6 years for IgG. The dilution with the albumin solution did not affect the commutability negatively. Therefore, the processed serum can be diluted using the albumin solution in order to obtain a reference material with an IgG MPO-ANCA concentration that is suitable.

Discussion

Biochemical testing for autoimmune antibodies is evolving, with new assay principles and multiplexing being introduced and automation becoming more generalized. There is also an evolution towards the provision of quantitative results.

This raises the issue of the reliability of these quantitative results. A laboratory result must be of value to the requesting clinician who will need appropriate information to interpret it. The specificity (negativity in health) and sensitivity (positivity in disease) of the test(s) must be considered to ensure that tests are capable of contributing to the diagnosis, prognosis or monitoring of disease. The result will need to be related to a reference range or a cut-off value appropriate to the test and to the patient being investigated. Increasingly, tests and their results are incorporated into protocols or best practice guidelines. In all these situations there is an assumption that the results for the same sample, for the same test analysed in different laboratories,

will be equivalent (within clinically meaningful limits); unfortunately, for many analytes, including autoimmune serology testing, this is simply not the case.

The results of these studies confirm, first, that there is a large variability in the results for IgG MPO-ANCA testing, despite the fact that the different assays participating in the study all performed well in terms of precision. Even more worrying is the fact that the assays do not only give different quantitative, but also different qualitative, results for the samples tested here. The diagnostic or qualitative results are the same for all assays for only 10 of 30 samples. These samples are routine samples for which IgG MPO-ANCA testing has been ordered.

Assay results are expressed in different units even when targeting the same analyte, and thus have different scales. However, the between-assay differences (as evidenced by the slopes of correlation lines) vary up to factor 5 even when the same units are used. Also, for certain combinations of assays there is a non-linear correlation of the results. These two aspects (non-linear correlation and correlation lines with slopes different from 1) are associated closely with the calibration of the assays, and could be improved using a common calibrator. A third factor causing variability of results from different assays is the different behaviour of individual samples with respect to the different assays. When results from two assays are compared, the results for individual samples show a scatter around the correlation line. This scatter is larger than what would be expected from the within-run imprecision alone, as this is below 10% for all assays (and sample averages over six measurements are compared). The Pearson's correlation coefficients for assay comparisons range from 0·29 to 0·95, from essentially a lack of correlation to an excellent correlation. For a subgroup of seven assays there is a reasonably good correlation for all combinations of assays, showing that they measure closely related properties of the patient samples. For one assay (Wieslab C) it is known that the different geometry of the assay causes differences in the numerical values compared to other assays, although the clinical significance is comparable. For another assay (Bioplex) dilutions are required for most of the samples, as this assay seems very much geared towards sensitivity. The low correlation between results from certain assay combinations is also associated with differences in assay selectivity, due to differences in reagents used and the geometry of the assay. Patient samples contain antibodies with different selectivity, affinity and avidity for the target, which will give a different response in assays. Assays with different selectivity may still have comparable clinical significance, or give complementary information. A large-scale evaluation of patient samples with clinical background information could provide more insight.

There is a potential for improvement by using a common calibrator and selecting a common cut-off, as shown by the results of the recalibration. After recalibration the average between assay CVs for the seven best correlating assays is reduced from 87 to 32%, and the number of samples with the same qualitative interpretation is 20 out of 30. However, particularly or borderline samples it is not possible to fully harmonize the classification over the different methods, even when using harmonized cut-offs.

A reference material certified for IgG MPO-ANCA represents an important step forward in the process of standardization of autoimmune serology testing. Two batches of candidate reference material formats were produced using different sources of raw materials: serum produced from plasmapheresis materials and affinity-purified IgG MPO-ANCA spiked into serum from healthy people. The influence of different processing steps on the activity of IgG MPO-ANCA was tested. Commutability studies were performed in order to assess whether the IgG MPO-ANCA present in those formats behaves in a similar manner to the autoantibodies present in routine patient samples. We have selected a commutable final format to produce a reference material for IgG MPO-ANCA. This reference material is not only positive in all IgG MPO-ANCA assays, but also behaves in a manner comparable to the majority of routine samples in terms of interassay properties and dilution behaviour. It is intended to be used for calibration of quantitative assays for IgG MPO-ANCA, as well as for quality control. As such it would be used mainly by manufacturers of IgG MPO-ANCA assays, but could also be used directly in clinical laboratories.

The use of a certified reference material for standardizing IgG MPO-ANCA assays could reduce the between-assay variation, as it provides a common scale. It could also help in making the assay results linear with respect to the reference material and to each other, and support the control of lot-to-lot variation of reagents. However, there will still be variation due to different responses of the assays to the biochemical properties of the samples and antibodies present as shown, for example, by Trevisin *et al.* [15]. It may be possible to increase the understanding of the relationship between the behaviour of the samples and the disease status of the patients. In the longer term this could lead to a more refined use of IgG MPO-ANCA testing.

In conclusion, the first steps towards the production of a candidate reference material for standardizing IgG MPO-ANCA testing have been made, and a candidate reference material for IgG MPO-ANCA has been produced.

Disclosure

None.

Author contributions

D. H. and I. Z. wrote the paper; I. Z., J. S., H. S. and P. M. designed the study; D. H. and E. T. performed the study; D. H., I. Z. and E. M. analysed the data.

References

1 Csernok E, Moosig F. Current and emerging techniques for ANCA detection in vasculitis. Nat Rev Rheumatol 2014; **10**:494–500.

2 Kallenberg CGM. Pathogenesis of PR3-ANCA associated vasculitis. J Autoimmun 2008; **30**:29–36.

3 Chan EKL, Fritzler MJ, Wiik A *et al.* AutoAbSC.Org–Autoantibody Standardization Committee in 2006. Autoimmun Rev 2007; **6**:577–80.

4 Mukhtyar C, Guillevin L, Cid MC *et al.* EULAR recommendations for the management of primary small and medium vessel vasculitis. Ann Rheum Dis 2009; **68**:310–7.

5 Kaul R, Johnson K, Scholz H, Marr G. Performance of the Bio-Plex 2200 Autoimmune Vasculitis kit. Autoimmun Rev 2009; **8**:224–7.

6 Sinico RA, Radice A, Corace C, DI Toma L, Sabadini E. Value of a new automated fluorescence immunoassay (EliA) for PR3 and MPO-ANCA in monitoring disease activity in ANCA-associated systemic vasculitis. Ann NY Acad Sci 2005; **1050**:185–92.

7 Mahler M, Radice A, Yang W *et al.* Development and performance evaluation of novel chemiluminescence assays for detection of anti-PR3 and anti-MPO antibodies. Clin Chim Acta 2012; **413**:719–26.

8 Merlini G, Blirup-Jensen S, Johnson AM, Sheldon J, Zegers I. Standardizing plasma protein measurements worldwide: a challenging enterprise. Clin Chem Lab Med 2010; **48**:1567–75.

9 Schimmel H, Zegers I, Emons H. Standardization of protein biomarker measurements: is it feasible? Scand J Clin Lab Invest 2010; **242** (Suppl.):27–33.

10 Meroni P, Biggioggero M, Pierangeli S, Sheldon J, Zegers I, Borghi M. Standardization of autoantibody testing: a paradigm for serology in rheumatic diseases. Nat Rev Rheumatol 2014; **10**:35–43.

11 Vesper H, Emons H, Gnezda M *et al.* C53-A: characterization and qualification of commutable reference materials for laboratory medicine: approved guideline. Wayne, PA: Clinical and Laboratory Standards Institute, International Federation of Clinical Chemistry and Laboratory Medicine, 2010.

12 Miller WG, Myers GL, Rej R. Why commutability matters. Clin Chem 2006; **52**:553–4.

13 Zegers I, Schreiber W, Linstead S *et al.* Development and preparation of a new serum protein reference material: feasibility studies and processing. Clin Chem Lab Med 2010; **48**:805–13.

14 Dawson B, Trapp RG. Basic and clinical biostatistics, 4th edn. New York: Lange Medical Books/McGraw-Hill, 2004.

15 Trevisin M, Pollock W, Dimech W *et al.* Antigen-specific ANCA ELISAs have different sensitivities for active and treated vasculitis and for nonvasculitic disease. Am J Clin Pathol 2008; **129**:42–53.