

Annals of Clinical Biochemistry: An international journal of biochemistry and laboratory medicine

<http://acb.sagepub.com/>

Uncertainty of measurement: an immunology laboratory perspective

Sarah C Beck and Robert J Lock

Ann Clin Biochem published online 26 August 2014

DOI: 10.1177/0004563214551066

The online version of this article can be found at:

<http://acb.sagepub.com/content/early/2014/10/09/0004563214551066>

Published by:



<http://www.sagepublications.com>

On behalf of:

Association for Clinical Biochemistry and Laboratory Medicine



Netherlands Society for Clinical Chemistry and Laboratory
Japan Society of Clinical Chemistry

Additional services and information for *Annals of Clinical Biochemistry: An international journal of biochemistry and laboratory medicine* can be found at:

Email Alerts: <http://acb.sagepub.com/cgi/alerts>

Subscriptions: <http://acb.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Oct 9, 2014

[Accepted Manuscript](#) - Aug 26, 2014

[What is This?](#)



Uncertainty of measurement: an immunology laboratory perspective

Sarah C Beck¹ and Robert J Lock²

Abstract

'Measurement uncertainty of measured quantity values' (ISO15189) requires that the laboratory shall determine the measurement uncertainty for procedures used to report measured quantity values on patients' samples. Where we have numeric data measurement uncertainty can be expressed as the standard deviation or as the co-efficient of variation. However, in immunology many of the assays are reported either as semi-quantitative (i.e. an antibody titre) or qualitative (positive or negative) results. In the latter context, measuring uncertainty is considerably more difficult. There are, however, strategies which can allow us to minimise uncertainty. A number of parameters can contribute to making measurements uncertain. These include bias, precision, standard uncertainty (expressed as standard deviation or coefficient of variation), sensitivity, specificity, repeatability, reproducibility and verification. Closely linked to these are traceability and standardisation. In this article we explore the challenges presented to immunology with regard to measurement uncertainty. Many of these challenges apply equally to other disciplines working with qualitative or semi-quantitative data.

Keywords

Immunoassay, Instrumentation, Statistics

Accepted: 20th August 2014

Introduction

“That's right”, shouted Vroomfondel, “We demand rigidly defined areas of doubt and uncertainty!”¹ Douglas Adams, Hitchhiker's Guide to the Galaxy, 1979.

Definitions of uncertainty are not all helpful but all imply an element of doubt caused by a situation in which something is not known. Measurement uncertainty can be quantified statistically where we have numerical data. However, in immunology many of the assays are reported either as semi-quantitative (i.e. an antibody titre) or qualitative (positive or negative) results. In the latter context, measuring uncertainty is considerably more difficult. There are, however, strategies which can allow us to minimise uncertainty.

The approach advocated herein for a range of immunology assays addresses both sorts of data.

Measurement uncertainty

Measurement of uncertainty provides a quantitative assessment of the quality of a test result. The old Clinical Pathology Accreditation (CPA) standard

¹Department of Immunology, Peterborough and Stamford's NHS Foundation Trust, Peterborough, UK

²Immunology and Immunogenetics, North Bristol NHS Trust, Bristol, UK

Corresponding author:

Robert J Lock, Immunology and Immunogenetics, North Bristol NHS Trust, Bristol, UK.

Email: robert.lock@nbt.nhs.uk

advocated that we 'determine the uncertainty of results... where possible'.¹ The current international standard (ISO 15189)² has a tighter requirement, instructing us to consider 'measurement uncertainty of measured quantity values'. As noted earlier, in immunology we must also understand, and attempt to minimise, the measurement uncertainty of the many assays producing a purely qualitative result.

The concept of measurement uncertainty was originally devised for use in measurements in physics. The total error of a measurement is a combination of random error (measured by the standard deviation, SD) and systematic error (bias). The measurement uncertainty principle is that bias should be corrected for by reference to standards, leaving the SD as the basic parameter for measuring uncertainty. As we will see, the lack of standards makes this much harder to isolate as the sole source of error in the biological sciences.

Causes of measurement uncertainty can occur at any stage of the laboratory process, i.e. preanalytical, analytical or postanalytical. Preanalytical specimen requirements and collection criteria are defined by the laboratory. Users of the service should have access to the requirements through the local laboratory users' guide. This should be reinforced through training and teaching, for example in induction sessions at the start of employment, and through user group interaction.

The bulk of our discussion focuses on the analytical phase. A number of parameters can contribute to making measurements uncertain. These include bias, precision, standard uncertainty (expressed as SD or coefficient of variation [CV]), sensitivity, specificity, repeatability, reproducibility and verification. Closely linked to these are traceability and standardisation.

Standardisation

Standardisation of assays for the determination of autoimmunity is an international problem and at present there is no unified international solution to resolve these problems. This is due to the fact that antibodies are difficult to define unequivocally due to their molecular structure, weight and biological variability and that they reside within a complex matrix (serum) and therefore cannot be directly measured.³ Furthermore, the quantification of an autoantibody relies on its functional ability to bind antigen, with consequent variation between samples and standards of affinity, avidity and epitope specificity. In consequence, the uncertainty of measurement with these assays is high. Variability and therefore uncertainty may be introduced at many levels, including the antigen source (whole tissue, cell extract, purified protein, recombinant protein), antibody

detected (isotype, affinity, concentration), antibody detection system (polyclonal, monoclonal, affinity, conjugation [enzyme, fluorochrome] and methodological variations [incubation time, volume, choice of substrate]). As a result we have to ensure the maximum adherence to quality control measures in order to minimise subsequent assay uncertainty.

Autoantibody standards

There is a limited number of autoantibody reference materials (Table 1) from the World Health Organisation (WHO) and the Centers for Disease Control which is available to laboratories. The purpose of these is to provide a uniform base from which other secondary standards can be derived. Secondary autoantibody standards may be derived by the local laboratory but in practice this is more commonly performed by the manufacturers, who provide assay standards with their tests, some of which may have been derived based on international standards where available.⁴ For many of these standards only the assigned reactivity is specified and not a quantitative value. For many autoantibodies this is entirely adequate as presence of an antibody is considered as support for an autoimmune diagnosis. For example, where there is biochemical evidence of Addison's disease (e.g. low serum sodium, low cortisol) the presence of antibodies to adrenal cortex, most commonly demonstrated by indirect immunofluorescence (IIF), will confirm the autoimmune nature of the disease. Similarly, the presence of antibodies to ovarian theca, also demonstrated by IIF, will provide support for an autoimmune aetiology for premature ovarian failure.

Protein standards

International calibration material is available for 12 serum proteins (ERM-DA470k/IFCC) including immunoglobulins G, A and M; complement components C3c and C4. Most, if not all, commercial companies derive their secondary commercial standards for specific proteins from ERM-DA470k/IFCC. This makes it possible for laboratories worldwide to compare these specific protein results over time and between hospitals and countries. Without such defined materials standardisation of assays is impossible as each manufacturer can define their own arbitrary target units.

Even with certified reference preparations harmonisation of results between manufacturers may not be achievable. A recent commutability study for caeruloplasmin, using the original and new reference preparations (ERM-DA470 and ERM-DA472/IFCC, respectively), failed to show standardisation for this measurand across multiple analytical platforms all of

Table 1. Examples of reference materials available for use in clinical immunology laboratories.

| Analyte | Reference material | Assigned units | Source |
|--|---|--|-----------|
| ANA (homogeneous/rim pattern) antinative DNA (dsDNA) | ANA#1 | None assigned | CDC |
| Antinuclear factor serum (homogenous) human | 66/233 (W 1064), 1 st International reference preparation 1971 | 100 IU/ampoule | WHO/NIBSC |
| ANA/ENA – (speckled pattern) anti-SS-B/La | ANA#2 | None assigned | CDC |
| ANA (speckled pattern) | ANA#3 | None assigned | CDC |
| ANA/ENA – antinuclear ribonucleoprotein (nRNP) autoantibody | W 1063, 1 st International reference preparation 1983 | None assigned | WHO/NIBSC |
| ANA/ENA – anti-U1 RNP (nuclear RNP) | ANA#4 | None assigned | CDC |
| ANA/ENA – anti-Sm | ANA#5 | None assigned | CDC |
| ANA – nucleolar pattern FANA; antifibrillar (U3 RNP) | ANA#6 | None assigned | CDC |
| ANA – anti-SS-A/Ro | ANA#7 | None assigned | CDC |
| ANA – centromere pattern FANA | ANA#8 | None assigned | CDC |
| ANA/ENA – anti-Scl-70 (DNA topoisomerase I) | ANA#9 | None assigned | CDC |
| ENA – Jo1 (histidyl-tRNA transferase) | ANA#10 | None assigned | CDC |
| ANA – PM-Scl | ANA#11 | None assigned | CDC |
| Cytoplasmic – ribosomal P | ANA#12 | None assigned | CDC |
| Human IgG anticardiolipin monoclonal HCAL (anti-beta-2 GPI) | ANA#13 | 1.1 ug/ml | CDC |
| Human IgM anticardiolipin monoclonal EY2C9 (anti-beta-2 GPI) | ANA#14 | | CDC |
| MPO-ANCA | ANA#15 | | CDC |
| PR3 – ANCA | ANA#16 | | CDC |
| Antismooth muscle serum (anti actin) | W 1062, 1 st International reference preparation 1983 | None assigned | WHO/NIBSC |
| Islet cell antibodies | 97/550, 1 st Reference Reagent, 1999 | Islet cell antibody 20 U/ampoule Anti-GAD55 100 U/ampoule Anti-IA-2 100 U/ampoule | WHO/NIBSC |
| Primary biliary cirrhosis serum (antimitochondrial antibody) | 67/183 | 100 U/ampoule | NIBSC |
| Rheumatoid arthritis serum (rheumatoid factor) | W 1066, 1 st International reference preparation 1973 | 100 IU/ampoule | WHO/NIBSC |
| Rheumatoid arthritis serum (rheumatoid factor) | 64/002, 1 st British Standard | 100 IU/ampoule | NIBSC |
| Human complement components C1q, C4, C5, factor B | W 1032, 1 st International reference preparation 1980 | | WHO/NIBSC |

(continued)

Table 1. Continued.

| Analyte | Reference material | Assigned units | Source |
|--|--|---|-----------|
| and whole functional complement | | C1q 100 IU/ampoule; C4 100 IU/ampoule; C5 100 IU/ampoule; CH50 100 IU/ampoule | |
| C-reactive protein | 85/506, 1 st International Standard 1986 | 0.049 IU/ampoule | WHO/NIBSC |
| Human serum immunoglobulin E | 75/502, 2 nd International reference preparation 1980 | 5000 IU/ampoule | WHO/NIBSC |
| Immunoglobulins G, A and M, complement components C3c and C4 | ERM-DA470k/IFCC ^a | IgG 9.17 ± 0.18 g/L IgA 1.8 ± 0.05 g/L IgM 0.723 ± 0.027 g/L C3c 1.00 ± 0.04 g/L C4 0.162 ± 0.007 g/L | IRMM |
| Human serum immunoglobulins G, A and M (IgG, IgA and IgM) | 67/086, 1 st International reference preparation 1970 | IgG 100 IU/ampoule IgA 100 IU/ampoule IgM 100 IU/ampoule | WHO/NIBSC |

ANA: antinuclear antibody; CDC: Centers for Disease Control, Atlanta, USA; CRP: C-reactive protein; ENA: extractable nuclear antigen; IRMM: Institute for Reference Materials and Measurement, Belgium; NIBSC: National Institute of Biological Standards and Control, Potters Bar, UK; WHO: World Health Organisation.

^aDA470k, although now the international reference material for many proteins, including immunoglobulins and complement, is not generally available to diagnostic laboratories. Many commercial preparations (secondary standards) are available, calibrated against this material.

which claimed to be traceable to the same calibration preparation (ERM-DA470).⁵

Traceability

For many autoantibody methods there is limited traceability of the standards, going only as far as manufacturers' in-house selected measurement procedures or master calibrators. Often these are assigned an arbitrary numeric value. Users should be advised that results are not directly comparable with results generated by different measurement procedures.³ There are numerous examples of divergence in the numeric values assigned to quantified autoantibody tests (e.g. antitissue transglutaminase,⁶ antithyroid peroxidase⁷). In the case of antithyroid peroxidase the two antigen sources (purified or recombinant) differ in glycosylation and are likely to express different conformational epitopes. In some patients this leads to very different binding of the autoantibodies, resulting in very different quantified results.

Even where there is good traceability there is often no agreed standard. There is currently much interest in IgG4 quantification in IgG4-related disease.⁸ However, there is no internationally agreed standard. The two main methods used in the UK (The Binding Site Group Ltd. and Siemens Healthcare Diagnostics) are traceable to two different standards. The Binding Site calibration material is traceable to the CRM470 whereas the Siemens calibration material is traceable to WHO67/97.^{9,10} This results in a situation where

the IgG4 results from the Siemens assay are twice those of the Binding Site assay.¹¹

Methodological standardisation

Published consensus documents exist for the use of some assays but most of these only relate to appropriate use of generic tests for disease diagnosis/monitoring rather than prescription of which analytical system to use (e.g. ESPGHAN (European Society for Pediatric Gastroenterology, Hepatology and Nutrition) for coeliac disease,¹² NICE (National Institute for Health and Clinical Excellence) guidance for coeliac disease¹³ and Rheumatoid arthritis¹⁴ and international consensus document of antineutrophil cytoplasm antibody (ANCA) testing and reporting¹⁵).

Having a good understanding of the analytical aspects of assays, including knowledge of the biology of the antibodies that we are detecting and of the clinical significance of results, is imperative in ensuring that results are given the correct interpretation. In particular, the user should appreciate the difficulty in controlling such variables as those introduced by changes in the antigen (e.g. animal or human, purified or recombinant), the sensitivity of the detection system (e.g. for rheumatoid factor agglutination (coated cells, gelatin beads or latex), ELISA (enzyme-linked immunosorbent assay) or fluorescence immunoassay) and the antibody mixture in the serum being analysed (avidity, affinity, antibody class (IgG, IgA, IgM or a combination of these) and epitope specificity).

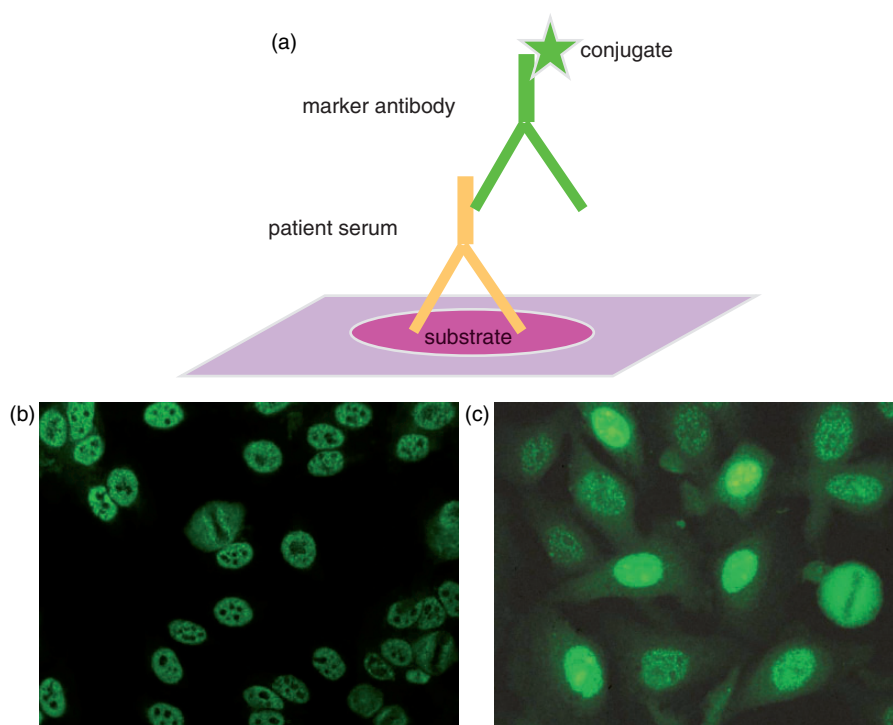


Figure 1. (a) Typical design for indirect immunofluorescence assays. (b) Appearance of antinuclear antibody on Hep2 cells (speckled pattern; nucleoli are spared; no cytoplasmic stain). (c) Appearance of antinuclear antibody on Hep2000 cells (speckled pattern variable owing to overexpression of Ro60 in some cells; nucleoli variable; no cytoplasmic stain).

Applications in clinical immunology

Immunology assays fall into three categories: qualitative assays, semi-quantitative assays and quantitative assays. Each will be discussed with reference to what factors make the results uncertain.

Qualitative assays

Qualitative assays include all IIF tests (e.g. antinuclear antibody (ANA), ANCA). The principle of these assays and the microscopic appearance of a typical assay are shown in Figure 1. These are subjective tests and therefore by their nature likely to have a high degree of uncertainty. Variability of reagents can be down to choice including the substrate (e.g. rodent tissue versus human cell lines for ANA), fixatives (e.g. pure ethanol or ethanol plus proprietary additives for ANCA), serum dilution, conjugation (fluorescence versus enzyme) and the quality of the secondary antibody (affinity, avidity, antibody class detected). Even the setup of UV (ultraviolet) microscopes (mercury versus light-emitting diode light sources, magnification, numerical aperture) will have an impact on the results obtained.

Whilst reagents are available as kits for some assays, for many we have to select individual components, each

of which must be optimised to maximise the sensitivity and specificity, for example by altering sample or reagent concentrations or changing incubation times. The laboratory may be directed by the manufacturer as to what is optimal for the autoantibody of interest (e.g. a 1 in 20 initial dilution of serum for multiple autoantibody screening on rodent liver/kidney/stomach composite sections [NOVALite, INOVA Diagnostics, Inc., San Diego, USA]). Determination of positivity/negativity is set based on the use of controls as well as operator consensus opinion and/or expert operator judgement.

Qualitative assays – substrate choice and uncertainty

As there is no standardisation of substrate preparation, differences between manufacturers can be seen. Substrate preparation can vary in choice of cell line and whether they are transfected with additional antigens, for example use of HEP2 cells versus HEP2000 for ANA determination. HEP 2000 cells are transfected with, and hyper-express, SS-A/Ro60 antigens for ANA detection (Figure 1(b) and (c)). Potential variations in tissue section preparation include cross-section of tissue, orientation, fixation and whether multiple tissues are arranged as a composite block, i.e. subsections of individual wells (e.g. rodent liver,

kidney and stomach slides). What is deemed best for an individual lab may depend on laboratory and clinical expertise and also personal preference. In this context it is worth noting that rodent tissue is relatively poor in Ro60 antigen and some patient serum samples with only this specificity may be negative on rodent composite blocks.

Semi-quantitative assays – substrate choice and uncertainty

Semi-quantitative assays include immunoblot methods. In these assays, substrates can be recombinant and/or purified antigens which are bound to nitrocellulose or other membranes. The most common commercial format has a series of antigens of interest immobilised on a single strip. Examples would include panels for myositis-related antibodies (e.g. Jo-1, PL7, PL12, OJ, EJ, etc.) (e.g. Euroline Myositis profiles, Euroimmun, Luebeck, Germany) or for paraneoplastic antibodies (e.g. Purkinje cell, antineuronal nuclear antibodies 1 & 2, amphiphysin, Ma2) (e.g. Euroline neuronal antigen profile, Euroimmun, Luebeck, Germany). There is no standardisation of antigens for these assays and the sensitivity and specificity vary accordingly. Reagents are usually produced as a kit and therefore have a manufacturer validated methodology. In this context (ISO15189),² a validated assay is one for which extensive studies have been undertaken to provide objective evidence of the performance characteristics of the assay (including, for example accuracy, precision, sensitivity and specificity). Note that it is still incumbent upon the user of the assay to verify the manufacturer's claims and this may require a significant resource input in terms of both time and reagents. All immunoblots include positive and negative controls on each membrane to authenticate each test.

ELISA methods that generate a ratio result are also categorised as semi-quantitative. Again, there is great variation in the substrates used. Screening ELISAs such as those used for ANA or extractable nuclear antigen (ENA) detection may contain a variety of purified or recombinant antigens, the exact combination of which will vary from manufacturer to manufacturer. These assays utilise a single calibrator (set by manufacturer), the optical density (OD) of which is multiplied by a factor (given by the manufacturer specific to each lot number) to give an OD cut-off by which all patient ODs must be divided to generate a ratio.

Other examples of semi-quantitative assays in immunology include assays for functional components of the immune system. These include complement (classical pathway CH50, alternate pathway), assays for neutrophil function and lymphocyte proliferation.

Again, these rely largely on single point calibration, usually a fresh random normal sample.

Quantitative assays

Quantitative assays include all ELISA-based autoimmunity and allergy tests (including those that are measured as quantitative but may only be reported as qualitative, e.g. ENA profile assays). Some examples of quantitative assays, and their associated limitations, are discussed below.

Substrate choice and uncertainty

Source of antigen was noted earlier as being one area for substantial variation between assays. For a few assays (e.g. cyclic citrullinated peptide) the patented antigen is supplied from a single source under licence to other manufacturers. Thus, differences between manufacturers can be reduced due to the limited variation in antigen preparations available. Yet even here variation has been introduced at a later point, with some manufacturers using other citrullinated antigens (e.g. Mutated citrullinated vimentin) thereby bypassing the patent issue. Whilst clinically equally valid these two antigen sources identify overlapping but different patient populations.¹⁶ The analytical differences between these assays are now being reflected in the UK NEQAS (National External Quality Assessment Service) scheme for anticitrullinated protein antibodies.¹⁷

Variability of substrate is even greater for other autoantigens. We offer one example. Anticardiolipin antibodies, as seen in the antiphospholipid syndrome, rely on the binding of a co-factor, beta-2-glycoprotein-1 (β GP1), which results in a neo-antigen to which the antibodies in turn bind. Current guidelines require the addition of β GP1 either bovine or human and that this may be by addition of purified human protein or bovine serum.¹⁸ The type of solid phase, concentration of antigen and method of coating are also recognised to contribute to the variation in substrate but are not addressed by the guideline.

Standards and uncertainty of autoantibody assays

ELISA-based assays present a challenge as there are few international standards for autoantibody preparations. Examples would include the 'Harris standards' (three generations have been available from Louisville APL Diagnostics, US traceable to the original Harris standards), the original anticardiolipin calibrators which were samples containing specified concentrations of antibodies¹⁹ and the WHO designated serum Wo80 for anti-dsDNA (Wo80) which unfortunately is no longer available. As a result, there is potentially huge

assay uncertainty which is shown by the variability of dynamic ranges and positive cut-off values between manufacturers for the same assay, as seen in many external quality assessment (EQA) returns in immunology. This is noted even in assays with an international reference preparation (e.g. anticardiolipin^{17,20}).

This is further highlighted by EQA schemes where most only require qualitative responses for some assays which are routinely reported quantitatively or semi-quantitatively. For example, for quantified anticardiolipin antibodies the interlaboratory CV for is often high (>25%).²¹ In a report on UK NEQAS data in 1995 the authors concluded 'On the basis of the evidence there seems little room for confidence in the concordance of numeric values for cardiolipin antibodies, and only slightly more confidence in their semiquantitative interpretation'. With the slow progress in agreement on international standards, there is little to suggest that this situation has improved in the intervening two decades. Changing to a qualitative approach achieved higher interlaboratory concordance than the numeric data in the UK NEQAS phospholipid scheme.⁴

Quantitative autoantibody assays are measured as IgG, IgA or IgM autoantibodies depending upon the isotype of clinical interest. Some assay systems (e.g. Phadia ImmunoCAP 250 EliA, Thermo Fisher Scientific Inc., Uppsala, Sweden) use immunoglobulin calibrators which are traceable via an unbroken chain of calibrations to the International Reference Preparation (IRP) 67/86 of Human Serum Immunoglobulins A, G and M from the WHO; measured in $\mu\text{g/L}$. A lot specific correction factor is applied to each batch of autoantibody specific assay wells and is used to automatically convert the mass unit results ($\mu\text{g/L}$) to the arbitrary units (e.g. U/mL) that are reported. All autoantibodies are measured in arbitrary units (U/mL or IU/mL) as there are no equivalents of SI units. Use of such correction factors must rely on manufacturer specific criteria to be appropriately set. Clearly, this does not always work effectively. Examination of EQA data even for the measurement of serum proteins shows some manufacturers to have introduced reproducible and consistent bias. Variability between lots may result in variability of correction factors which potentially may affect end results generated. This may be highlighted in lot specific shifts observed in cumulative internal quality control (IQC) results.

Owing to the lack of standardisation in autoantibody measurement, different manufacturers utilise different arbitrary units and positive cut-off values. Positive cut-offs provided by manufacturers ought to have a good evidence base but should be verified by laboratories. Whilst it is advised to adjust/set up in-house reference ranges based on local population cohorts this is often not feasible particularly in a district

general hospital setting. Limited patient numbers and the ethical issues of large normal control studies may mean that small cohort studies would lead in inaccurate results. For many 'data mining' is one feasible way to overcome this. Data mining, also called knowledge discovery, is a term borrowed from computer science whereby raw data in large databases is turned into useful information. One example of this applied to immunology is the use of probability plots to isolate normal from abnormal populations within a database.²²

Standards and uncertainty of assays for allergy testing allergen specific IgE and total IgE

Total serum IgE concentrations are reported in international units or nanograms per millilitre (1 IU/mL = 2.44 $\mu\text{g/L}$). Total IgE is cross-standardised with the WHO 75/502 human reference IgE serum verified by periodic proficiency surveys.²³ As with total IgE, some commercial specific IgE antibody assays are calibrated using heterologous interpolation against the WHO 75/502 human IgE reference serum, thereby enabling a uniform system of reporting as international units of IgE per millilitre. This value can be converted into mass units of IgE per volume (e.g., micrograms of IgE per litre) because the IgE calibration curve is standardised against the WHO Human IgE International Reference Preparation 75/502. Note that this ignores the problem that to measure total IgE we use reagents that react with the constant region of the IgE molecule (Fc, fragment crystallisable), giving us some hope of measurement in mass units. In contrast, specific IgE testing relies on the active binding site of the patient's IgE (located in the Fab region [fragment antibody binding]) which will differ in epitope specificity and affinity/avidity from patient to patient. This must have an effect on the amount of specific IgE bound in a given system. In addition, source and preparation of allergen (purified, recombinant) is chosen by the manufacturers which will introduce further variation. Fortunately, the small number of manufacturers delivering specific IgE testing (five represented on UKNEQAS) should limit the problem.

Total serum mast cell tryptase concentrations are reported in $\mu\text{g/L}$. The UK NEQAS for Tryptase EQA scheme shows all participants utilise the Phadia ImmunoCAP methodology albeit on the various platforms available. Calibrators are calibrated against a manufacturer's internal tryptase reference material, which has been purified from human lung according to the protocol published by Schwartz.²⁴ As all three platforms and their reagents are manufactured by the same supplier the preparation of the reference material across platforms should be

standardised. However, it is clear that, despite there being a single supplier, there is significant bias according to the platform used.¹⁷ Nevertheless between laboratory reproducibility, which should be similar to ELISA (between batch variation 8–12%⁴), is acceptable at around 10%.¹⁷

Analytical controls to reduce risk

Given all the analytical variability described earlier, how can we ensure that our immunoassays are generating results as accurately and reliably as possible? Below are examples of what we believe to be necessary actions that should be applied to the different methodologies. While the examples given are from immunology, much of the advice could be equally well applied to other laboratory disciplines.

1. For qualitative assays

- (a) Validation methods include confirmation of clinical sensitivity and specificity where possible. This requires a significant input of resources in terms of reagents and time (both scientific and clinical) and may not be achievable in all centres. In this case it is essential that the laboratory thoroughly research the literature to understand how the assay performs. Comparison between manufacturers may be necessary. Indeed, this is vital when replacing one assay with another in the laboratory's repertoire. This is essential if the end users are to be aware of the impact of the change on their patients, either at diagnosis or for monitoring response to therapy. Analytical tools and appropriate statistical evaluation for this are limited for qualitative assays but can be applied more extensively for quantitative assays.
- (b) Use of IQC material. Positive and negative serum samples should be included in each batch. Sensitivity can be monitored by including a positive sample in serial dilutions (usually two-fold serial dilutions) to give a titre. The end result should be plus or minus one increment (e.g. if the expected titre is 1 in 320 the acceptable range would be 1 in 160 to 1 in 640).
- (c) Enrolment in EQA schemes, where they exist. CPA accreditation¹ requires that where no such schemes exist, alternative mechanisms of comparisons should be employed such as sample exchange arrangements.
- (d) Clinical authorisation with appropriate clinical commentary. Note that full interpretation would require knowledge of the positive and negative predictive values of the assay in the context of your local population.

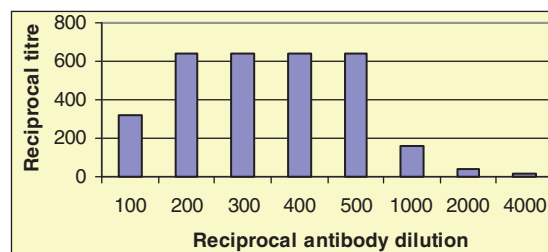


Figure 2. The principle of optimising conjugated antisera for use in indirect immunofluorescence. See text for detail. In this example the 1 in 100 dilution shows prozone. The appropriate dilution for use is 1 in 400.

- (e) Use of CE (Conformité Européene) marked reagents not necessarily as part of a standard kit. This requires regular re-evaluation, e.g. of appropriate dilution of fluorescein isothiocyanate-conjugated antisera. If the antiserum is too concentrated one risks a prozone effect (reduced stain, the mechanism of which is uncertain), wastes money and potentially increases background stain. If too dilute there is a loss of sensitivity. The standard approach is to use a known positive serum of defined titre to create a checkerboard with the conjugated antiserum. The appropriate dilution is the last but one before the fall off in sensitivity. This allows long-term storage without risking a drop in sensitivity. An example of this is shown in Figure 2.
 - (f) Expertise in IIF reading.
 - (g) Audit of IIF reading with comparison of user agreement and regular revalidation of competency exercises. Practically, this first requires the collection of a panel of serum samples with defined specificities (for example a panel for antinuclear antibodies would include homogeneous, speckled, nucleolar, centromere, mitotic spindle apparatus, etc.). These are then used to create a test slide which should be assessed 'blind' by all staff involved in the regular reading of IIF. Similar panels may be created for other specificities, e.g. ANCA, skin antibodies. Definition of 'correct result' must be established whether this be based on an expert opinion or based on consensus. Where any reader is considered out of consensus then re-education is the key to quality improvement.
 - (h) Regular maintenance of analysers, diluters and UV microscope.
- ### 2. For semi-quantitative assays
- (a) As above a–d and h plus.
 - (b) Use of CE marked commercial kits. It is possible to use non-CE marked kits provided that adequate evidence of assay maintenance is

- available, however this may be prohibitive for most routine laboratories.
3. For quantitative assays
 - (a) As above plus.
 - (b) Use of manufacturer validated reference ranges, internally verified as appropriate.
 - (c) Scrutiny of IQC performance to accept assay runs with appropriate use of Westgard rules, modified where appropriate.⁴ Regular review of IQC performance by monitoring coefficient of variation (CVs). This can also include cumulative CV analysis. Internal assessment of within batch and between batch CV. The expected CV will vary according to the technology used with, for example 3–5% for nephelometry or 8–12% for ELISA.⁴ For flow cytometric assays one could expect even higher, depending on the antigen under investigation (for example we find CD4+ T-cells CV 6-10%; CD16+ NK cells CV 15-22% [RJL – unpublished data]).
 - (d) Appropriate root cause analysis investigations of IQC/EQA failure.

Overcoming clinical risk given measurement uncertainty with immunoassays

Given the known lack of standardisation, international calibrators or reference preparations available for immunoassays, it is clear that control procedures such as internal and external control and assessment should be scrutinised to minimise the potential risk.

'Sense checking' is very important to limit illogical results. This should be done at the assay authorisation stage. For example if a complete run of ANA or anti-TTG (anti-tissue transglutaminase) elicited no positive results this should flag as an oddity. Thus, if the expected positivity rate was approximately 10% for each (a typical figure, but which will vary depending on the population studied), one would statistically expect positive results in each run. For example, assuming 10% prevalence and a batch size of 100 samples, it is very unlikely only one would be positive ($p=0.0007$). It is important that results are looked at in combinations and not in isolation by suitably trained and qualified staff who have a sound analytical and clinical understanding of the assays. For example a dsDNA antibody should not be reported with a negative ANA without further analysis. This might include checking by alternate methods with higher specificity. Some tests require correlation with tests that may be performed in other disciplines, such as total IgA concentration with low IgA anti-TTG results

or antimitochondrial antibody with liver function tests. This is good practice and, with the advent of combined blood sciences departments and integrated information technology solutions should become increasingly more common.

The use of clinical sensitivity and specificity gives good indications for the clinical significance of results for different clinical conditions. Where possible these should be publicised to users. However, many widely quoted clinical specificities particularly must be used with caution. The issue of technological drift must be considered.²⁵ Many of the clinical associations were defined using the original high specificity 'gold standard' assays. With the greatly improved sensitivity of modern methods there is often an adverse effect on the clinical specificity of auto-antibodies for specific clinical conditions. For example high affinity dsDNA antibodies detected by the original radioimmunoassay (using the Farr technique) were highly specific for systemic lupus erythematosus but not sensitive.²⁶ Current highly sensitive ELISA assays to dsDNA also detect low affinity antibodies which may be found in other disease states and healthy, particularly elderly, individuals therefore reducing specificity. Sole reliance on data from manufacturers is not advised but large in-house studies to accurately determine clinical sensitivity and specificity may not be feasible. Furthermore, differences in local populations affect the pretest probabilities of having the disease under investigation and thus affect the determination of positive and negative predictive values. These locally derived parameters may be more useful in interpretation than the sensitivity and specificity alone.

It is not recommended that any Immunology test results alone are used to make a definitive diagnosis without correlation with the clinical context. Many parameters are prognostic indicators but again should not be used in isolation to direct clinical management. In the case of potentially useful serial markers (e.g. dsDNA antibodies, anti-glomerular basement membrane or ANCA titres) the use of reference change values (a statistical tool reviewed by Fraser²⁷) is likely to be very inaccurate given the assay uncertainty as well as the heterogeneity of the clinical conditions. However, delta checking for the antibody concentrations can be useful and enables suitable comments to be added to results. One clear example is the use of paraproteins as markers of response to therapy in plasma cell proliferative disorders.²⁸ Clinicians are advised to look at trends with results rather than absolute value differences.

Good professional practice requires us to define our uncertainty. This means knowing our assays, their limitations and not generating results blindly without putting them into context. Immunoassays may have a long way to go in standardisation but we can ensure we are providing results and interpretation to the best of our ability.

Acknowledgements

This article was prepared at the invitation of the Clinical Sciences Reviews Committee of the Association for Clinical Biochemistry and Laboratory Medicine.

Declaration of conflicting interests

None declared.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical approval

Not applicable

Guarantor

RJL.

Contributorship

SCB and RJL both wrote, edited and agreed the final version.

References

- Standards for the Medical Laboratory. Clinical Pathology Accreditation (UK) Ltd. PD-LAB-Standards v2.02 Nov 2010 <http://www.cpa-uk.co.uk/files/pdlabst.pdf> (accessed 23 August 2013).
- British Standards Institution. Medical Laboratories – Requirements for quality and competence (ISO 15189:2012). Publisher BSI Standards Limited 2012.
- White GH. Metrological traceability in clinical biochemistry. *Ann Clin Biochem* 2011; 48: 393–409.
- Lock RJ. My approach to internal quality control in a clinical immunology laboratory. *J Clin Pathol* 2006; 59: 681–684.
- Zegers I, Beetham R, Keller T, et al. The importance of commutability of reference materials used as calibrators: The example of ceruloplasmin. *Clin Chem* 2013; 59: 1322–1329.
- Egner W, Shrimpton A, Sargur R, et al. ESPGHAN guidance on coeliac disease 2012: Multiples of ULN for decision making do not harmonise assay performance across centres. *J Pediatr Gastroenterol Nutr* 2012; 55: 733–735.
- Wood WG and Hanke R. Comparability of method results and performance in a national external quality assessment scheme between 1993 and 2003 using thyroid associated antibodies as examples. *Clin Lab* 2004; 50: 209–221.
- Stone JH, Zen Y and Deshpande V. IgG4-related disease. *N Engl J Med* 2012; 366: 539–551.
- Schauer U, Stemberg F, Rieger CH, et al. IgG subclass concentrations in certified reference material 470 and reference values for children and adults determined with the binding site reagents. *Clin Chem* 2003; 49: 1924–1929.
- Klein F, Skvaril F, Vermeeren R, et al. The quantification of human IgG subclasses in reference preparations. *Clin Chim Acta* 1985; 150: 119–127.
- Khosroshahi A, Cheryk LA, Carruthers MN, et al. Brief report: Spuriously low serum IgG4 concentrations caused by the prozone phenomenon in patients with IgG4-related disease. *Arthritis Rheumatol* 2014; 66: 213–217.
- Husby S, Koletzko S, Korponay-Szabo IR, et al. ESPGHAN Working Group on Coeliac Disease Diagnosis; ESPGHAN Gastroenterology Committee. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition Guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012; 54: 136–160.
- NICE clinical guideline 86. *Coeliac disease: Recognition and assessment of coeliac disease*. London: National Institute for Health and Clinical Excellence, 2009, www.nice.org.uk/CG86 (accessed 23 August 2013).
- NICE Clinical Guidelines, No. 79 Rheumatoid Arthritis National Clinical Guideline for Management and Treatment in Adult. National Collaborating Centre for Chronic Conditions (UK). London: Royal College of Physicians (UK); February 2009.
- Savige J, Gillis D, Benson E, et al. International consensus statement on testing and reporting of antineutrophil cytoplasmic antibodies (ANCA). *Am J Clin Pathol* 1999; 111: 507–513.
- Yousefghahari B, Alhoeei S, Soleimani-Amiri MJ, et al. Comparison of sensitivity and specificity of anti-CCP and anti-MCV antibodies in an Iranian cohort of patients with rheumatoid arthritis. *Caspian J Intern Med* 2013; 4: 702–706.
- UK NEQAS for Immunology, Immunochemistry and allergy. Annual report 2012 http://www.immqas.co.uk/downloads/Annual_Report_2012_V1.pdf (accessed 30 Sept 2013).
- Lakos G, Favaloro EJ, Harris EN, et al. International consensus guidelines on anticardiolipin and anti- β 2-glycoprotein 1 testing. Report from the 13th international congress on antiphospholipid antibodies. *Arthritis Rheum* 2012; 64: 1–10.
- Devreese KM. Standardization of antiphospholipid antibody assays Where do we stand? *Lupus* 2012; 21: 718–721.
- Miyakis S, Lockshin M, Atsum T, et al. International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). *J Thromb Haemost* 2006; 4: 295–306.
- Ward AM and White PA. Measuring cardiolipin antibodies. *BMJ* 1995; 310: 1472–1473.
- Lock RJ and Unsworth DJ. Immunoglobulins and immunoglobulin subclasses in the elderly. *Ann Clin Biochem* 2003; 40: 143–148.
- Bernstein IL, Li JT, Bernstein DI, et al. Allergy diagnostic testing: An updated practice parameter. *Ann Allergy Asthma Immunol* 2008; 100: S1–S148.
- Schwartz LB. Mediators of human mast cell and human mast cell subsets. *Ann Allergy* 1987; 58: 226–237.

25. Lock RJ and Unsworth DJ. Antibodies to extractable nuclear antigen Has technological drift affected clinical interpretation? *J Clin Pathol* 2001; 54: 187–190.
26. Wold RT, Young FE, Tan EM, et al. Deoxyribonucleic acid antibody: A method to detect its primary interaction with deoxyribonucleic acid. *Science* 1968; 23: 806–807.
27. Fraser CG. Reference change values. *Clin Chem Lab Med* 2012; 50: 807–812.
28. Bird J, Behrens J, Westin J, et al. On behalf of the Haemato-oncology Task Force of the British Committee for Standards in Haematology, UK Myeloma Forum and Nordic Myeloma Study Group. UK Myeloma Forum (UKMF) and Nordic Myeloma Study Group (NMSG): Guidelines for the investigation of newly detected M-proteins and the management of monoclonal gammopathy of undetermined significance (MGUS). *Brit J Haematol* 2009; 147: 22–42.