

Challenges in the Standardization of Autoantibody Testing: a Comprehensive Review

Renato Tozzoli¹ · Danilo Villalta² · Nicola Bizzaro³

Published online: 8 August 2016
© Springer Science+Business Media New York 2016

Abstract Standardization and harmonization are complementary tools to achieve higher testing quality in laboratory medicine. Both are of great relevance and are strongly needed in autoimmune diagnostics, due to the impressive advance in basic research and technological development observed in this diagnostic field in recent years that has led to the introduction of many new tests and new analytical methods. It is, therefore, essential that this strong innovative thrust is translated into clinical practice in a coordinated way to avoid confusion and the risk of potentially harmful errors for the patient. However, while standardization of antibody assays is a very complex task, harmonization of procedures and behaviors is a more feasible target and should necessarily include all the phases of the total testing process—in the pre-analytical phase, appropriateness of test requests, harmonization of autoantibody terminology, and adoption of uniform nomenclature for laboratory tests; in the analytical phase, harmonization of measurements, and sharing of test profiles and diagnostic algorithms; and in the post-analytical phase, harmonization of data reporting, and criteria for interpreting immunoserological results, especially harmonization of units, reference intervals, decision limits, and definition and notification of critical values. We here provide and discuss some examples of harmonization initiatives related to anti-nuclear antibodies, TSH

receptor, and anti-thyroid peroxidase antibodies and to antibodies associated with autoimmune hepatitis and with celiac disease. These initiatives could be the starting steps to achieve a wider consensus and a closer interaction among stakeholders in the path of autoimmune diagnostics harmonization to enhance clinical effectiveness and provide greater patient safety.

Keywords Harmonization · Autoimmune diagnostics · Celiac disease · Gluten sensitivity · ANA · Reflex test · TSH receptor antibodies · Thyroid antibodies · Autoimmune hepatitis

Introduction

In laboratory medicine, the terms standardization and harmonization are frequently used interchangeably to define the condition in which laboratory results are comparable among different measurement procedures over time and space [1], but they define distinct though closely linked concepts based on traceability principles [2].

Although these concepts have the same final goal, the term “standardization” refers to the condition in which calibration is traceable to a reference measurement procedure calibrated with an appropriate reference material, while “harmonization” is a more general term and includes the standardized condition as well as the condition in which results are consistent and comparable despite the absence of a reference procedure.

Reference materials and procedures are available for quite a few analytes in clinical chemistry but are lacking for autoantibody assays for several reasons: autoantibodies are complex and structurally heterogeneous molecules due to post-translational modification and are present in biological fluids in different types due to oligoclonality. For these reasons, in the field of autoimmune disease diagnostics (laboratory

✉ Nicola Bizzaro
nicola.bizzaro@as3.sanita.fvg.it

¹ Laboratory of Clinical Pathology, S. Maria degli Angeli Hospital, Pordenone, Italy

² Clinical Allergy and Immunology, S. Maria degli Angeli Hospital, Pordenone, Italy

³ Laboratory of Clinical Pathology, San Antonio Hospital, via Morgagni 18, 33028 Tolmezzo, UD, Italy

autoimmunology) as in other areas of molecular diagnostics, only harmonization is currently feasible.

In autoimmune diagnostics, the most pressing drivers for harmonization are patient centered and organization centered [2]. The first category is the most important because differences in clinical practice and in application of diagnostic thresholds are common for several autoantibodies measured in the diagnosis of systemic or organ-specific autoimmune disorders, and these differences may lead to increased patient risks. For example, there is no doubt that differences in test reference values may lead to confusion and reduced patient safety. In addition, performing tests that provide the best analytical and diagnostic performance and avoiding tests that offer little incremental information would save money and improve the risk to benefit ratio.

Over the past 25 years, autoimmune diagnostics has gone through an evolutionary period. There have been many achievements in pathophysiology as well as steady advancement in the development of diagnostic technologies [3]. To meet both clinical need and the growing demand for autoantibody testing, automation of autoimmune diagnostics has spread into the autoimmunology laboratory, as an extension of a similar progress in several other areas of the clinical laboratory [4]. This great change has highlighted even more strongly the urgent need to harmonize all aspects of the total testing process (TTP) related to diagnostic autoantibody, including the pre-analytical, analytical, and post-analytical phases. This concept has been defined as the complete picture of harmonization in laboratory medicine, as outlined in the definition of harmonization provided by the Clinical and Laboratory Standards Institute (CLSI) [5, 6].

Harmonization Projects in Autoimmune Testing

As in other field of laboratory medicine, in autoimmune diagnostics, harmonization initiatives and projects are proliferating. The increased awareness of the relevance of the global picture in laboratory practice has led to several attempts to improve upon the different steps of the TTP: in the pre-analytical phase, the appropriateness of test requests, including harmonization of autoantibody terminology and adoption of uniform nomenclature for laboratory tests [7]; in the analytical phase, harmonization of measurements—referring to any process enabling equivalence of reported values produced by different measurement procedures for the same measurand; and finally, in the post-analytical phase, harmonization of data reporting and interpretation of immunoserological results, especially standardization of units for reporting test results and harmonization of reference intervals and decision limits.

The scope of this review is the presentation and discussion of some harmonization initiatives and projects, completed by scientific associations or groups of experts, related to various

autoantibodies and their diagnostic features, namely, anti-nuclear antibody, thyrotropin (TSH) receptor autoantibodies, anti-thyroid peroxidase antibody measurements, antibodies in autoimmune hepatitis, and anti-transglutaminase antibody reference intervals and decision limits.

Harmonization in Anti-nuclear Antibody Testing

The anti-nuclear antibody (ANA) harmonization process involves many factors including the starting dilution, the choice of the clinically relevant titer, pattern classification, and diagnostic algorithms (such as when and which confirmatory tests should be performed in the presence of a positive ANA result or even in the presence of a negative ANA test when an autoimmune rheumatic disease is strongly suspected). As harmonization can also be defined by the concept that we should “speak the same language,” the correct definition of the test name is an important aspect.

We and others [8, 9] have recently drawn attention to the fact that the term anti-nuclear antibodies, in its literal meaning, cannot be considered technically correct because it is not exhaustive of the spectrum of autoantibodies recognized by the indirect immunofluorescence (IIF) method on HEp-2 cells, which detects the presence of a set of autoantibodies directed against various cell structures including nuclear constituents, components of the nuclear membrane, mitotic apparatus, cytoplasmic organelles, and cell membrane. We believe that a new term, “*antibodies to intracellular antigens*,” might more accurately describe the wide spectrum of antibodies recognized by the diagnostic test. However, the acronym ANA, because of its established and universal use, might not easily be replaced. We suggest that it can be maintained, but the lab report should refer to the test as “ANA—antibodies to intracellular autoantigens” in order to state explicitly that the search includes autoantibodies against all cellular constituents, not just those present in the nucleus.

This proposal may resolve the only apparently semantic discussion whether a cytoplasmic or mitotic apparatus staining pattern is to be considered ANA positive and overcome the problem of reporting cytoplasmic patterns as ANA negative, with possible misinterpretation of test results. Indeed, if such a pattern is reported as negative, the additional information in the report on pattern and titer may go unnoticed because clinicians tend to pay less attention to ANA-negative results [10].

Similarly, the acronym ENAs for “extractable nuclear antigens” is too restrictive, and we suggest that it should be replaced by the term “antibodies to intracellular specific antigens” because today it is possible to detect autoantibodies to a greater number of antigenic specificities, including nuclear non-extractable antigens and cytoplasmic antigens.

Uniform terminology is needed also in the description of the ANA-IIF patterns [11]. Toward that end, the initiative of the International Consensus on ANA Patterns (ICAP) [10, 12] looks very helpful especially because the consultation is readily and freely available online at the www.ANAPatterns.org website and can be done very quickly, even during the reading at the microscope stage. With this online program, it is possible to get information on the different patterns, the target antigens corresponding to the autoantibodies that may produce that given pattern, and their clinical associations. The inevitable limit of this and other similar initiatives is that the patterns represented are always very clear and emblematic, while the patterns found in the daily diagnostic workup are not so. At least half of the ANA patterns observed during the routine workup are not typical and cannot easily be framed in one of the patterns proposed by ICAP. Nevertheless, this initiative is widely welcome and, as stated by the same authors, this is only the first step taken by the ICAP toward a more comprehensive analysis of the entire spectrum of patterns identified by the ANA-IIF test.

As for the initial dilution of the sample, there is now sufficient agreement [8, 9] that the threshold cutoff for ANA should no longer be fixed at 1:40 as suggested in the 1990s [13, 14]. Accumulated evidence has made clear that the best compromise between sensitivity and specificity of the ANA test be at least 1:80. In fact, at a titer below 1:80, the proportion of true positives that are lost is very low, while false positives are >30 % [13]. In addition, the huge increase over the last 20 years in requests for ANA tests as a screening test by non-rheumatologists and especially by general practitioners led to the occurrence of false positives at unacceptable rates. Furthermore, the choice of 1:80 as the optimal screening dilution is consistent with the results obtained by Tan et al. [13] on more than 22,000 healthy individuals, showing that this titer corresponds to the 95 percentile of healthy controls recommended by the EASI group [8]. It is also worth mentioning that 1:80 is the screening dilution adopted by all manufacturers of recently developed computer-aided systems for automated reading and interpretation of ANA [15]. The use of these automated systems is expected to improve the harmonization of the reading of ANA. In particular, two important benefits are expected: greater agreement in discriminating between positive and negative ANA samples and lower imprecision in the definition of antibody titer/concentration [15].

The clinically significant value (decision limit) is higher and is set at 1:160 because this titer is associated with both the probability of a positive confirmatory test (anti-dsDNA and anti-ENA) and the probability that the patient is suffering from an autoimmune rheumatic disease [16–18]. Unless clinical findings suggest execution of second-level tests even when ANA is positive at low titer or negative at a titer below 1:160, it does not make sense to continue with the diagnostic investigation; rather the patient should be kept under observation.

As already mentioned, the attempt at harmonization in autoimmune diagnostics concerns not only the nomenclature and the cutoff values but also, and perhaps most importantly, the adoption of algorithms that guarantees the best diagnostic efficacy according to internationally accepted guidelines and optimal use of limited financial resources.

One of the algorithms that are making progress is ANA-reflex. The term “reflex test” is used to indicate a “cascade” diagnostic approach in which if an initial test (first level) is positive, testing continues with one of the new in-depth tests (second level). The rationale of the ANA-reflex is to simplify the patient workup: a single visit to the doctor’s surgery, a single visit to the laboratory, and a more rapid clinical diagnosis.

However, the ANA-reflex test is much more complex than other reflex tests (such as TSH reflex or PSA reflex) because patterns observable on HEp-2 cell substrates are so numerous (>50) that interpretation by the pathologist is required along with choice of the most suitable confirmatory tests according to the IIF pattern. Hence, evaluation of the ANA pattern and titer is fundamental to the execution of second-level tests. For certain patterns, such as homogeneous, speckled, fine grainy (Scl70-like), nucleolar, centromeric or speckled cytoplasmic, the identification of precise autoantibody markers is considered essential, while for others it is not deemed necessary because they do not constitute classification criteria for any systemic autoimmune rheumatic disease (Table 1 indicates which reflex tests should be executed based on the pattern type observed on the HEp-2 cells).

In our opinion, the ANA-reflex test request should always be accompanied by clinical information. Although this will be a major challenge to introduce into the ANA testing requirements, some signs and symptoms could independently justify the execution of second level tests [19]. The exact nature of the signs and symptoms to associate to the ANA-reflex test request should be decided in conjunction with the referring rheumatologists. Clinical findings that could warrant second-level tests even in the case of low titer ANA positivity or ANA negativity are shown in Table 2. Harmonization of laboratory behaviour in these cases will enable a faster and more complete diagnostic process and, at the same time, prevention of waste of resources by avoiding unnecessary confirmatory tests.

Harmonization of Nomenclature of TSH Receptor Autoantibodies and of the Reporting of Results

TSH receptor (TSHR) autoantibodies (TRABs) are the pathogenic and diagnostic hallmarks of autoimmune hyperthyroidism (e.g., Graves’ disease (GD)). Three varieties of TRAb are now recognized in GD patients: stimulating antibodies, blocking antibodies, and apoptotic antibodies. Stimulating antibodies are pathogenic and lead to hyperthyroidism; blocking antibodies prevent the functional activity of the thyroid and

Table 1 ANA-reflex test procedure based on titer and fluorescence pattern

ANA-IIF pattern on HEp-2 cells	Reflex test(s)
Nuclear homogeneous $\geq 1:160$	Anti-specific intracellular antigens (anti-ENA) and anti-dsDNA/nucleosomes/histones
Nuclear speckled $\geq 1:160$	Anti-dsDNA and anti-specific intracellular antigens (anti-ENA) possibly including anti-RNA polymerase III
Fine grainy (Scl70-like) $\geq 1:160$	Anti-specific intracellular antigens (anti-ENA) possibly including anti-PM/Scl
Cytoplasmic speckled $\geq 1:160$	Anti-specific intracellular antigens (including anti-tRNA synthetases and anti-P ribosomal)
Pleomorphic (any titer)	Anti-PCNA
Centromere	No confirmation necessary if high titers. Execute specific test for anti-CENP-B only in dubious cases (low titer or centromeric pattern not clearly recognizable)
Few nuclear dots; low titer nucleolar ($< 1:160$); NuMa-like or MSA 1 (mitotic spindle antigen 1); spindle fibers (anti-HsEg5); intercellular bridge (MSA 2); CENP-F (MSA 3); cytoplasmic GW bodies; polar/Golgi-like; fibrillar cytoplasmic	No further test is available or necessary

Anti-specific intracellular antigens (ENA) include Ro52 and Ro60, La, U1RNP, Sm, Scl70, and Jo1

cause hypothyroidism; apoptotic antibodies are directed against the cleavage region of the TSHR and are able to induce apoptosis of thyrocytes [20].

After Duncan Adams’ historical discovery of the long-acting thyroid stimulator (LATS) as a cause of hyperthyroidism, a long list of bioassay and immunoassay (IMA) methods has been described for detecting TRAb (reviewed in [21]). Bioassays measure the functional activity of TRAbs (stimulating, blocking or apoptotic), while immunoassays measure the binding of autoantibodies to the receptor without functional discrimination. Some methods can detect only a few of these autoantibodies, while other methods can detect all of them. Furthermore, some recently developed immunometric methods are capable of specifically measuring functional autoantibodies, particularly stimulating antibodies [22, 23]. This

condition has produced a plethora of confusing and confounding terms and abbreviations [24], based on the characteristics of the biochemical and immunological reactions of the different methods (Table 3).

To contribute to the harmonization of the results obtained with the new assay methods, either biological or immunometric, and in consideration of the fact that the new reference preparation (“2nd International Standard for Thyroid-Stimulating Antibody,” National Institute of Biological Standards and Control, NISBC code 80/204) merges with, but does not replace, the previous standard (NISBC code 90/672), we consider it appropriate to introduce a new classification system and new terminology for anti-TSH receptor antibodies (Table 4). We propose using this new nomenclature in lab reports, where the method used (biological

Table 2 Clinical findings that could warrant second-level tests even if ANAs are negative or positive at low titer

Clinical manifestation	Reflex test(s)
Persistent oral or ocular dryness	Anti-intracellular specific antigens (anti-ENA)
Raynaud’s phenomenon and/or photosensitivity (or malar rash) and/or leucopenia and/or arthritis	Anti-dsDNA and anti-specific intracellular antigens (anti-ENA)
Raynaud’s phenomenon and ANA positivity with nucleolar pattern at elevated titers ($\geq 1:320$)	Anti-PM/Scl, anti-fibrillarin, anti-Th/To, and anti-RNA polymerase III
Significantly increased CPK	Anti-specific intracellular antigens (anti-ENA) and myositis-associated antibodies
Persistent arthritis	Anti-citrullinated peptide antibodies and rheumatoid factor
ANA and for SLE-associated specific antibodies (i.e., dsDNA, Sm, RNP, Ro60), with a clinical history of thrombotic events and/ or polyabortion	Anti-phospholipid antibodies (anti-cardiolipin, anti-beta2 glycoprotein I, lupus anticoagulant)

Anti-specific intracellular antigens (ENA) include Ro52 and Ro60, La, U1RNP, Sm, Scl70, and Jo1

Table 3 Terms and abbreviations used for TSH receptor autoantibodies

Term	Abbreviation
Long-acting thyroid stimulator	LATS
LATS protector	LATS-P
TSH-binding inhibiting immunoglobulins	TBIs
Thyroid-stimulating antibodies	TSAbs
Thyroid-stimulating immunoglobulins	TSI
Thyroid-blocking antibodies	TBAbs
Thyroid-blocking immunoglobulins	TBI
TSH-stimulation blocking antibodies	TSBAb
TSH-blocking antibodies	TSBABs
TSH receptor antibodies	TSHRABs
Stimulating TSH-receptor antibodies	S-TSHRABs
Blocking TSH-receptor antibodies	B-TSHRABs
Neutral TSH-receptor antibodies	N-TSHRABs

or immunometric) and the characteristics of the assay should also be specified. In particular, for immunometric methods, specifications should be given for the type of receptor used for the solid phase and for the analytical platform; for biological methods, specs are needed for the type of cell used and the type of instrument system; and for each of the methods, the international reference preparation (80/204 or 90/672) should be specified. This proposal would not automatically standardize the results of the TRAb assay but would allow the clinician to better interpret data obtained by different methods in different laboratories. In all, these changes would contribute to the harmonization of the results obtained with the new assay methods (biological and immunometric) and will highlight all TRAb or their individual functional variety.

Harmonization of Upper Reference Limits of Thyroid Peroxidase Antibodies

Autoantibodies against thyroid peroxidase (TPOAbs) are markers and early indicators of autoimmune thyroid diseases and have an important predictive role in healthy subjects, in pregnant women, and in high-risk patients [25, 26].

In recent years, refinements in the preparation of TPO antigen for optimal coating in solid phase assays and in the

Table 4 New proposed nomenclature for the classification of TSH receptor autoantibodies

Autoantibody	Abbreviation
Total TSH receptor autoantibodies	T-TRABs
Stimulating TSH receptor autoantibodies	S-TRABs
Blocking TSH receptor autoantibodies	B-TRABs
Apoptotic (cleavage) TSH receptor autoantibodies	A-TRABs

selection of polyclonal and monoclonal antibodies have led to third-generation (3G), automated, quantitative IMAs with improved sensitivity and specificity for the measurement of TPOAbs [27–29]. However, notwithstanding these analytical improvements and the use of the same reference preparation (NIBSC code 66/387), efforts must be made in the definition of the upper reference limit (URL) in order to correctly classify patients with autoimmune thyroid disease (AITD). Estimation of the URL for TPOAbs is a very critical issue, arising mainly from the uncertainty associated with the procedures used to correctly define the reference population.

Using 3G automated commercial immunoassays, we demonstrated that the URLs of TPOAb are method and gender dependent [30, 31]. When we tested 120 healthy males and 120 healthy females, selected according to the recommendations of the National Academy of Clinical Biochemistry (NACB) [32] with the 12 most diffuse IMA methods applied in automated analyzers, we found wide differences in experimental URLs, ranging from 1.0 to 29 IU/mL both in males and in females (Table 5).

We also found a significant difference between experimental TPOAb URLs and those given in the package insert of the analysis kits. In most cases, experimental TPOAb values were lower than those proposed by the manufacturers, with a delta value ranging from 10 to 60 %. This finding was in line with two previous studies on the same topic [29, 31]. In our opinion, these discrepancies may be linked to racial differences among subjects, or, more likely, to non-standardized criteria in the selection of the reference subjects, possibly resulting in the enrollment of individuals with subclinical AITD and with abnormal levels of TPOAb.

Another relevant consideration emerging from our study [30, 31] was the dependence of URLs on the analytical characteristics of the methods. Indeed, while some methods have an overall average median value <3 IU/mL, other methods have higher values up to 10 IU/mL. These results seem largely dependent on the different analytical sensitivity or limit of detection (LoD) which ranges from 0.05 to 9.3 IU/mL. There are no clear reasons for these discrepancies, which indicate poor harmonization between methods that are both automated and use the same reference preparation. The high inter-method variability might be caused by the different coating preparation of the TPO antigen (purified native or recombinant), which affects the exposure of the immunodominant epitopes recognized by the polyclonal antibodies present in sera of AITD patients [33, 34] (epitopic fingerprint), where improper exposure may lead to poor recognition in some cases.

Hence, despite the attempt of harmonization among methods, the heterogeneity of the LoD observed among methods requires greater standardization of the analytical performance with an urgent invitation to the biomedical industry companies to pay attention to the TPO antigen preparation as the possible source of variability between different assays and

Table 5 Analytical performance of 12 automated methods/platforms for the measurement of TPOAbs in a reference male and female healthy population (LoD values stated by the manufacturers)

Method	LoD (IU/mL)	Median (IU/mL)	Experimental URL (IU/mL) Males	Experimental URL (IU/mL) Females	Manufacturer URL (IU/mL)	Classification
Siemens Immulite XPi	5.00	6.6	29	29	35	3G
Siemens Advia Centaur XP	9.30	10.0	20	25	20	3G
Roche Diagnostics Cobas e411	5.00	5.5	18	28	34	3G
Diasorin Liaison XL	0.12	1.8	14	18	16	3G
SNIBE Maglumi 2000 Plus	0.38	2.2	25	25	30	3G
Thermo Scientific Phadia EliA 250	4.00	2.8	8	10	25	3G
Abbott Architect ci4100	0.21	0.6	2	2	6	4G
Thermo Scientific Brahms Kryptor	1.80	2.4	6	7	10.5	4G
Fujirebio Lumipulse G 1200	0.60	2.2	5	5	6.9	4G
Beckman Coulter Unicel Dx C880i	0.25	0.8	7	9	9	4G
Tosoh Bioscience AIA-2000	0.05	0.4	1.3	1.5	3.2	4G
Tosoh Bioscience AIA-CL2400	0.05	0.2	1	1	1.5	4G

LoD low limit of detection (analytical sensitivity), 3G third generation, 4G fourth generation, nd not declared

to reevaluate test procedures using approved protocols and the guidelines of the Clinical and Laboratory Standard Institute. We also propose that the method used for detection of TPOAbs, the instrument platform, and its analytical sensitivity (LoD) all be clearly indicated in the lab report.

At this moment, it could be useful to use a classification of IMA methods distinguishing third- and fourth-generation IMAs, depending on whether even only one among the three parameters considered (LoD, median values of the distribution of reference group and URL defined as positivity cutoff) is higher or lower than 2.0, 3.0, and 15.0 IU/mL, respectively. This new tentative classification might enable harmonization in the interpretation of TPOAb results, avoiding considering IMA methods with low (3G) or high (4G) analytical sensitivity as equivalent.

Finally, another important finding of our study was the difference of the TPOAb reference values between sexes: for some of the methods, medians were not significantly different between groups of males and groups of females, while the medians were very different for other methods. Based on these results, we propose an integration of the dated guidelines of NACB [32], recommending the use of two distinct groups of reference individuals, one composed of male healthy subjects and one of female healthy subjects. This proposal may help overcome inaccuracy resulting from the use of a male reference group for a disease (autoimmune thyroiditis) that mainly affects the female sex and may further help harmonize URLs for TPOAb.

Harmonization in the Diagnosis of Autoimmune Hepatitis

Autoimmune hepatitis (AIH) is an immune-mediated liver disease characterized by elevated transaminase levels,

hypergammaglobulinemia, serum autoantibodies, and histological interface hepatitis, with a variety of clinical presentations ranging from asymptomatic liver abnormalities to acute severe hepatitis or even acute liver failure [35]. The heterogeneity of the clinical presentation can make difficult the diagnosis of AIH even in experienced hands. In 1993, the International Autoimmune Hepatitis Group (IAHG) proposed a scoring system to help standardization of patient selection for research purposes [36]. The scoring system was subsequently revised in 1999 [37] and also was applied to daily clinical practice, although it was not primarily designed and tested for this purpose. Notwithstanding this limitation, the updated scoring system proved useful to accommodate deficiencies or inconsistencies in clinical presentation and to support diagnosis in difficult cases [38]. However, because these criteria are cumbersome and insufficiently validated, the IAIHG recently devised a simplified scoring system for wider use in routine clinical practice [39]. Using clinical judgment as the standard, the simplified scoring systems yielded lower sensitivity (95 vs 100 %) but higher specificity (90 vs 73 %) and higher overall accuracy (92 vs 82 %) than the 1999 scoring system [40].

The new scoring systems helped improve comparability of diagnoses of AIH from different medical centers, but some limitations have not been overcome. For example, they are inappropriate for determining the presence of AIH in patients with primary biliary cirrhosis (PBC) and are not validated in patients with acute severe liver failure or in patients with graft dysfunction after liver transplantation [38].

Serum autoantibodies play a pivotal role in the diagnosis of AIH and represent a relevant component of the scoring systems, particularly ANA, anti-smooth muscle (SMA), anti-liver kidney microsomes type 1 (anti-LKM-1), and anti-

soluble liver antigen/liver–pancreas antigen (anti-SLA/LP). Unfortunately, autoantibody testing is not sufficiently standardized and therefore may lead to inadequate scoring values. For this reason, in 2004, the IAIHG Committee for Autoimmunity Serology drew up a consensus statement containing guidelines for appropriate and effective autoantibody testing in AIH [41]. In this statement, the committee recommended using the IIF method with rodent tissue sections dried in air without further fixation, both for ANA and anti-LKM-1. However, some of these recommendations cannot be followed in most autoimmunology laboratories, where commercially available tissue substrates treated with fixatives and HEp-2 cell lines are used for SMA/anti-LKM-1 and ANA detection, respectively.

As far as ANA is concerned, if the screening is performed on HEp-2 cells, the values are higher than when screening is done on tissue sections, and as a consequence, the score attributed to the ANA in the scoring systems is not applicable. The IAIHG suggests that if results from HEp-2 cells are used, they should be halved. However, this rule has not been validated by comparative studies and cannot be applied without distinction in all conditions and for all the ANA patterns. This remains an unresolved problem in the attempt to harmonize AIH diagnosis.

SMA with F-actin specificity are commonly regarded as more specific markers of type 1 AIH, but a reference method for their identification is not yet available. Enzyme-linked immunosorbent assay (ELISA) and IIF methods using vascular smooth muscle (VSM47) and rat intestinal epithelial cell lines recently have been proposed for anti-F-actin detection. The promising and quite comparable results obtained in pivotal studies [42–44] have to be confirmed by more extensive studies before the implementation of these methods in daily practice.

Recently, the identification of the molecular targets of some autoantibody specificities present in AIH, such as cytochrome P4502D6 (anti-LKM-1), formimino-transferase cyclodeaminase (anti-LC-1) and Sep (O-phosphoserine) tRNA/Sec (selenocysteine) tRNA synthase (anti-SLA/LP), has led to the establishment of various immunoassays based on the use of recombinant or purified antigens. Nowadays, they are widely used in clinical laboratories, but data on the variability among methods and among assays are lacking. Therefore, studies comparing the results of different methods and different assays for detection of these autoantibodies are needed to confirm if their results are harmonized or not.

Finally, the starting dilution for IIF tests in children is another debated question. The consensus statement reports that “for subjects up to the age of 18 years, any level of autoantibody reactivity in serum is infrequent, so that positivity at dilutions of 1/20 for ANA and SMA and even 1/10 for anti-LKM-1 is clinically relevant. Hence, the laboratory should report any level of positivity from 1/10, with the result interpreted within the clinical context and the age of the patient.” Apart from the

abovementioned considerations on the results of ANA obtained using HEp-2 cell lines, substantial differences in autoantibody titers between adults and children are not supported by strong evidence. In addition, using a line blot assay, we have recently demonstrated that anti-LKM-1 and anti-LC-1 antibody prevalence and concentration do not differ in adults and children [44].

In conclusion, the scoring systems, and especially the simplified scoring system which is more suitable to clinical application, have been useful tools in the harmonization process of AIH diagnosis. However, for autoantibody testing, some recommendations should be revised in light of the more recent methods used in routine clinical practice. Comparative studies evaluating the diagnostic accuracy of different methods/assays are still needed.

Harmonization in the Diagnosis of Celiac Disease

Celiac disease (CD) is an immune-mediated disorder elicited by gluten in genetically susceptible individuals. Immunoglobulin A (IgA) anti-tissue transglutaminase (tTG) antibody assay is the preferred single test for detection of CD, and its high accuracy (sensitivity and specificity around 95 %) has been demonstrated both in primary care settings and in referral cohorts [45, 46]. While some studies have shown comparable diagnostic utility in the screening of CD of the most often used commercial anti-tTG-IgA assays, controversies relating to their predictive positive value (PPV) for CD exist [47–52]. Although many studies have demonstrated that, in most assays, a higher titer of antibody generally correlates with a higher PPV for CD (Marsh 3 histology) [50, 53–55], allocation of a harmonized anti-tTG IgA decision cutoff remains problematic. In fact, as a consequence of the lack of standardization between anti-tTG IgA assays in the absence of an international reference preparation and of a reference measurement procedure, antibody units and reference ranges are arbitrary, method specific, and not commutable.

In the effort to find a strategy for avoiding biopsy in children, and in the absence of an acceptable standard to define the specific level at which anti-tTG IgA antibodies predict the disease, the updated European Society of Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) guidelines for the diagnosis of CD [56] suggest the use of a specific multiple of the URL in algorithms for symptomatic or asymptomatic children as an attempt to harmonize the results obtained with different assays. In particular, it is suggested that a histological assessment may be omitted in symptomatic patients who have anti-tTG IgA levels 10 times above URL as verified by EMA positivity, and who also are positive for HLA-DQ2 and/or HLA-DQ8 heterodimer. The threshold of >10 URL was deduced by the results of some primary studies showing that in these cases, the likelihood for villous atrophy (Marsh 3) is very high or absolute [39, 42, 43]. However, such

an arbitrary cutoff value cannot be generalized as a statement which is valid for all commercial assays in all screened cohorts, even if published studies are internally valid [57]. Recently, Suh-Lailam et al. [58] showed that the use of multiples of URL does not improve commutability among anti-tTG IgA assays. Similar results were obtained by Beltran et al. [52], who showed a wide disparity of results and poor consensus, both between methods and between laboratories evaluating the values returned by laboratories participating in the UK National External Quality Assurance Scheme (UK NEQAS) for CD serology. In addition, they showed that the normalization to URL does not harmonize results between anti-tTG IgA assays. These findings highlight the differences among commercial assays and show that common multiples and stratified thresholds cannot be used unless true comparability between the tests exists. In other words, a generalized recommendation about fixed threshold decision points only will make sense when anti-tTG IgA values are standardized, which is traceable to a high-order primary reference material. While we are waiting for an international standard or reference preparation for anti-tTG IgA measurement, we have to consider that, at present, each anti-tTG IgA assay is unique and its characteristics must be considered when establishing diagnostic pathways in the laboratory as well as communicating results to clinicians. So, data derived from local audits are needed to determine performance characteristics and optimal thresholds of the anti-tTG IgA assay in use in each laboratory.

If a main effort has to be done to harmonize and standardize the analytical process of anti-tTG IgA measurement, it is likewise important to consider that harmonization of CD diagnostic includes also other aspects, with special emphasis to test profiles (i.e., when testing for EMA, for anti-deamidated gliadin antibodies and for HLA DQ2/DQ8), report formats, and criteria for interpretations of results. An important step toward the harmonization of some of these aspect is represented by the recent guidelines proposed by national or international scientific societies [56, 59–61], even if some items (i.e., the level of IgA to define IgA deficiency) still remain a matter of debate and further studies are needed to obtain evidence for final decisions.

Harmonization in the Diagnosis of Non-Celiac Gluten Sensitivity

The diagnosis of non-celiac gluten sensitivity (NCGS) in patients with persistent intestinal and/or extra-intestinal complaints is still based on exclusion of CD and of wheat allergy [62]. As NCGS is an emergent clinical condition, there is a high need to harmonize the procedure leading to confirmation of suspected cases. Unfortunately, there is no specific serologic marker for the diagnosis of NCGS. Recently, an

international expert group on gluten-related disorders developed recommendations on how a diagnosis of NCGS should be confirmed [63]. In addition to negative serum anti-tTG and anti-endomysium antibodies, negative duodenal biopsy (Marsh 0–1) and negative-specific IgE and prick tests to wheat, the Salerno Experts' Criteria suggest that a close and standardized monitoring of the patient during elimination and reintroduction of gluten is the most specific diagnostic approach and can be used as the diagnostic hallmark of NCGS. However, since many of these patients are already on the gluten-free diet (GFD) at the first visit, the Salerno experts counsel that a simplified diagnostic procedure, including only the effect of reintroducing gluten after a period of treatment with the GFD, may be adopted in these patients. Most important from a practical point of view, in clinical practice, a single-blind procedure could be sufficient, while for research purposes, a double blind placebo-controlled challenge remains the first choice [63].

In conclusion, harmonization of autoimmune diagnostics is necessary for clinical and diagnostic appropriateness, for economic reasons (to optimize use of resources), and for patient safety (especially when a patient visits different laboratories). As standardization and harmonization go hand in hand, efforts at improving sensitivity and specificity and reducing analytical imprecision of diagnostic tests are certainly an important goal to be pursued by the laboratory and by the manufacturing companies, but the main goal is patient outcome. We must ask ourselves, even when using a highly reliable and possibly standardized antibody assay, how many diagnoses would we miss if we do not use a good diagnostic algorithm? Improved harmonization of diagnostic behavior would be the challenge for forthcoming years.

References

1. Miller WG, Eckfeldt JH, Passarelli J, Rosner W, Young IS (2014) Harmonization of test results: what are the challenges; how can we make it better? *Clin Chem* 60:923–927
2. Plebani M (2015) Harmonization in laboratory medicine: requests, samples, measurements and reports. *Crit Rev Clin Lab Sci* 15:1–13
3. Tozzoli R, Bizzaro N (2012) The clinical autoimmunologist and the laboratory autoimmunologist: the two sides of the coin. *Autoimmun Rev* 11:766–770
4. Tozzoli R, D'Aurizio F, Villalta D, Bizzaro N (2015) Automation, consolidation, and integration in autoimmune diagnostics. *Autoimmun Highlights* 6:1–6
5. Tate JR, Johnson R, Sikaris K (2012) Harmonization in laboratory testing. *Clin Biochem Rev* 33:121–122
6. Plebani M (2013) Harmonization in laboratory medicine: the complete picture. *Clin Chem Lab Med* 51:741–751
7. Legg M, Swanepoel C (2012) The Australian pathology units and terminology standardization project—an overview. *Clin Biochem Rev* 33:103–108

8. Agmon-Levin N, Damoiseaux J, Kallenberg C, Sack U, Witte T, Herold M et al (2014) International recommendations for the assessment of autoantibodies to cellular antigens referred to as anti-nuclear antibodies. *Ann Rheum Dis* 73:17–23
9. Cinquanta L, Bizzaro N, Villalta D, Morozzi G, Tonutti E, Bagnasco M et al (2015) Guidelines for the laboratory use of autoantibody tests in the diagnosis and monitoring of autoimmune rheumatic diseases. Revision 2015. *Riv It Med Lab [Italian]* 11:205–224
10. Damoiseaux J, von Muhlen CA, Garcia-De La Torre I, Carballo OG, de Melo CW et al (2016) International consensus on ANA patterns (ICAP): the bumpy road towards a consensus on reporting ANA results. *Autoimmun Highlights* 7:1
11. Wiik AS, Høier-Madsen H, Forslid J, Charles P, Meyrowitsch J (2010) Antinuclear antibodies: a contemporary nomenclature using HEp-2 cells. *J Autoimmun* 35:276–290
12. Chan EKL, Damoiseaux J, Carballo OG et al (2015) Report of the first international consensus on standardized nomenclature of anti-nuclear antibody HEp-2 cell patterns 2014–2015. *Front Immunol* 6:412, PMID:26347739
13. Tan EM, Feltkamp TEW, Smolen JS et al (1997) Range of antinuclear antibodies in ‘healthy’ individuals. *Arthritis Rheum* 40:1601–1611
14. Tozzoli R, Bizzaro N, Tonutti E, Villalta D, Bassetti D, Manoni F et al (2002) Guidelines for the laboratory use of autoantibody tests in the diagnosis and monitoring of autoimmune rheumatic diseases. *Am J Clin Pathol* 117:316–324
15. Bizzaro N, Antico A, Platzgummer S, Tonutti E, Bassetti D, Pesente F et al (2013) Automated antinuclear immunofluorescence antibody screening: a comparative study of six computer-aided diagnostic systems. *Autoimmun Rev* 13:292–298
16. Dawkins RL, Martinez OP, Freitas EM, Hollingsworth PN (1998) Diagnosis of autoimmune disease. In: Rose NR, Mackay IR (eds) *The autoimmune diseases*, 3rd edn. Academic, San Diego, pp 821–831
17. Bizzaro N, Wiik A (2004) Appropriateness in anti-nuclear antibody testing: from clinical request to strategic laboratory practice. *Clin Exp Rheumatol* 22:349–355
18. Wiik AS, Bizzaro N (2012) Missing links in high quality diagnostics of inflammatory systemic rheumatic diseases. It is all about the patient! *Autoimmun Highlights* 3:35–49
19. Tonutti E, Bizzaro N, Morozzi G, Radice A, Villalta D, Cinquanta L et al (2016) The ANA-reflex test as a model for improving clinical appropriateness in autoimmune diagnostics. *Autoimmun Highlights* 7:9. doi: 10.1007/s13317-016-0080-3
20. Morshed SA, Davies TF (2015) Graves’ disease mechanisms: the role of stimulating, blocking, and cleavage region TSH receptor antibodies. *Horm Metab Res* 47:727–734
21. Tozzoli R, Bagnasco M, Villalta D (2014) Thyrotropin receptor antibodies. In: Shoenfeld Y, Meroni PL, Gershwin ME (eds) *Autoantibodies*, 3rd edn. Elsevier, Amsterdam, pp 375–383
22. Frank CU, Braeth S, Dietrich JW, Wanjura D, Loos U (2015) Bridge technology with TSH receptor chimera for sensitive direct detection of TSH receptor antibodies causing graves’ disease: analytical and clinical evaluation. *Horm Metab Res* 47:880–888
23. Tozzoli R, D’Aurizio F, Villalta D, Giovannella L (2016) Evaluation of the first fully automated immunoassay method for the measurement of stimulating TSH receptor autoantibodies in Graves’ disease. *Clin Chem Lab Med* doi: 10.1515/cclm-2016-0197
24. Tan K, Loh TP, Sethi S (2013) Lack of standardized description of TRAb assays. *Endocrine* 43:732–733
25. Czarnocka B, Cocks Eschler D, Godlewska M, Tomer Y (2014) Thyroid antibodies: thyroid peroxidase and thyroglobulin antibodies. In: Shoenfeld Y, Meroni PL, Gershwin ME (eds) *Autoantibodies*, 3rd edn. Elsevier, Amsterdam, pp 365–373
26. Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JI et al (2012) Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocr Pract* 18:988–1028
27. Tozzoli R, Bonaguri C, Melegari A, Antico A, Bassetti D, Bizzaro N (2013) Current state of diagnostic technologies in the autoimmunology laboratory. *Clin Chem Lab Med* 51:129–138
28. Bizzaro N, Tozzoli R, Villalta D (2014) Autoimmune diagnostics: the technology, the strategy and the clinical governance. *Immunol Res* 61:126–134
29. D’Aurizio F, Tozzoli R, Villalta D, Pesce G, Bagnasco M (2015) Immunoassay of thyroid peroxidase autoantibodies: diagnostic performance in automated third generation methods. A multicentre evaluation. *Clin Chem Lab Med* 53:415–421
30. Tozzoli R, D’Aurizio F, Ferrari A, Castello R, Metus P, Caruso B et al (2016) The upper reference limit for thyroid peroxidase autoantibodies is method-dependent: a collaborative study with biomedical industries. *Clin Chim Acta* 452:61–65
31. D’Aurizio F, Metus P, Polizzi Anselmo A, Villalta D, Ferrari A, Castello R et al (2015) Establishment of the upper reference limit for thyroid peroxidase autoantibodies according to the guidelines proposed by the National Academy of clinical biochemistry: comparison of five different automated methods. *Autoimmun Highlights* 6:31–37
32. Baloch Z, Carayon P, Comte-Devolx B, Demers LM, Feldt-Rasmussen U, Henry JF et al (2003) Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. *Thyroid* 13:3–126
33. Godlewska M, Gora M, Buckle AM, Porebski BT, Kemp EH, Sutton BJ et al (2014) A redundant role of human thyroid peroxidase propeptide for cellular, enzymatic and immunological activity. *Thyroid* 24:371–382
34. Nielsen CH, Brix TH, Gardas A, Banga JP, Hegedus L (2008) Epitope recognition patterns of thyroid peroxidase autoantibodies in healthy individuals and patients with Hashimoto’s thyroiditis. *Clin Endocrinol (Oxf)* 69:664–668
35. Vergani D, Mieli-Vergani G (2011) Cutting edge issues in autoimmune hepatitis. *Clin Rev Allerg Immunol* 42:309–321
36. Johnson PJ, McFarlane IG (1993) Meeting report: international autoimmune hepatitis group. *Hepatology* 18:998–1005
37. Alvarez F, Berg PA, Bianchi FB, Bianchi L, Burroughs AK, Cancado EL et al (1999) International autoimmune hepatitis group report: review of criteria for diagnosis of autoimmune hepatitis. *J Hepatol* 31:929–938
38. Czaja AJ (2016) Diagnosis and management of autoimmune hepatitis: current status and future directions. *Gut Liver* 10:177–203
39. Hennes EM, Zeniya M, Czaja AJ, Pares A, Dalekos GN, Krawitt EL et al (2008) Simplified criteria for the diagnosis of autoimmune hepatitis. *Hepatology* 48:169–176
40. Czaja AJ (2008) Performance parameters of the diagnostic scoring systems for autoimmune hepatitis. *Hepatology* 48:1540–1548
41. Vergani D, Alvarez F, Bianchi FB, Conçado ELR, Mackay IR, Manns MP et al (2004) Liver autoimmune serology: a consensus statement from the committee for autoimmune serology of the International Autoimmune Hepatitis Group. *J Hepatol* 41:677–683
42. Villalta D, Bizzaro N, Da Re M, Tozzoli R, Komorowski L, Tonutti E (2008) Diagnostic accuracy of four different immunological methods for the detection of anti-F-actin autoantibodies in type 1 autoimmune hepatitis and other liver-related disorders. *Autoimmunity* 41:105–110
43. Toh BH, Taylor R, Pollok V, Dearden S, Gill CC, Buchner C et al (2010) Actin-reactive discriminated from non-actin-reactive smooth muscle auto-antibody by immunofluorescence reactivity with rat epithelial cell line. *Pathology* 42:463–469
44. Villalta D, Girolami E, Alessio MG, Sorrentino MC, Tampoia M, Brusca I et al (2016) Autoantibody profiling in a cohort of pediatric and adult patients with autoimmune hepatitis. *J Clin Lab Anal* 30:41–46

45. Van der Windt DA, Jellema P, Mulder CJ, Kneepkens CM, van der Horst HE (2010) Diagnostic testing for celiac disease among patients with abdominal symptoms: a systematic review. *J Am Med Assoc* 303:1738–1746
46. Giersiepen K, Lelgemann M, Stuhldreher N, Ronfani L, Husby S, Koletzko S et al (2012) Accuracy of diagnostic antibody tests for coeliac disease in children: summary of an evidence report. *J Pediatr Gastroenterol Nutr* 54:229–241
47. Carroccio A, Vitale G, Di Prima L, Chifani N, Napoli S, La Russa C et al (2002) Comparison of anti-transglutaminase ELISAs and anti-endomysial antibody assay in the diagnosis of celiac disease: a prospective study. *Clin Chem* 48:1546–1550
48. Reeves GE, Squance ML, Duggan AE, Murugasu RR, Wilson RJ, Wong RC et al (2006) Diagnostic accuracy of coeliac serological tests: a prospective study. *Eur J Gastroenterol Hepatol* 18:493–451
49. Weiss B, Bujanover Y, Avidan B, Fradkin A, Weintraub I, Shainberg B (2004) Positive tissue transglutaminase antibodies with negative endomysial antibodies: low rate of celiac disease. *Isr Med Assoc J* 6:9–12
50. Hill PG, Holmes GK (2008) Coeliac disease: a biopsy in not always necessary for diagnosis. *Aliment Pharmacol Ther* 27:572–577
51. Li M, Yu L, Tiberti C, Bonamico M, Taki I, Miao D et al (2009) A report on the international transglutaminase autoantibody workshop for celiac disease. *Am J Gastroenterol* 104:154–163
52. Beltran L, Koenig M, Egner W, Howard M, Butt A, Austin MR et al (2013) High-titre circulating tissue transglutaminase-2 antibodies predict small bowel villous atrophy, but decision cut-off limits must be locally validated. *Clin Exp Immunol* 176:190–198
53. Donaldson MR, Book LS, Leiferman KM, Zone JJ, Neuhausen SL (2008) Strongly positive tissue transglutaminase antibodies are associated with Marsh 3 histopathology in adult and pediatric celiac disease. *J Clin Gastroenterol* 42:256–260
54. Alessio MG, Tonutti E, Brusca I, Radice A, Licini L, Sonzogni A et al (2012) Correlation between IgA tissue transglutaminase antibody ratio and histological finding in celiac disease. *J Pediatr Gastroenterol Nutr* 55:44–49
55. Singh P, Kurray L, Agnihotri A, Das P, Verma AK, Sreenivas V et al (2015) Titers of anti-transglutaminase antibody correlate well with severity of villous abnormalities in celiac disease. *J Clin Gastroenterol* 49:212–217
56. Husby S, Koletzko S, Korponay-Szabó IR, Mearin ML, Phillips A, Shamir R et al (2012) European society for pediatric gastroenterology, hepatology, and nutrition guidelines for the diagnosis of celiac disease. *J Pediatr Gastroenterol Nutr* 54:136–160
57. Egner W, Shrimpton A, Sargur R, Patel D, Swallow K (2012) ESPGHAN guidance on coeliac disease 2012: multiples of ULN for decision making do not harmonize assay performances across centres. *J Pediatr Gastroenterol Nutr* 55:733–735
58. Suh-Lailam BB, Davis KW, Tebo AE (2016) Immunoassays for the detection of IgA antibodies to tissue transglutaminase. Significance of multiples of upper limit of normal and inter-assay correlations. *Clin Chem Lab Med* 54:257–264
59. Rubio-Tapia A, Hill ID, Kelly CP, Calderwood AH, Murray JA (2013) ACG clinical guidelines: diagnosis and management of celiac disease. *Am J Gastroenterol* 108:656–676
60. Bai JC, Fried M, Corazza GR, Schuppan D, Farthing M, Catassi C et al (2013) World gastroenterology organization global guidelines on celiac disease. *J Clin Gastroenterol* 47:121–126
61. Hill P, Austin A, Forsyth J, Holmes G (2015) British Society of Gastroenterology guidelines on the diagnosis and management of celiac disease. *Gut* 64:691–692
62. Sapone A, Bai JC, Ciacci C, Dolinsek J, Green PH, Hadjivassiliou M et al (2012) Spectrum of gluten-related disorders: consensus on new nomenclature and classification. *BMC Med* 10:13
63. Catassi C, Elli L, Bonaz B, Bouma G, Carroccio A, Castillejo G et al (2015) Diagnosis of non-celiac gluten sensitivity (NCGS): the Salerno experts' criteria. *Nutrients* 7:4966–4977