CrossMark

MECHANISM IN AUTOIMMUNITY

The burden of the variability introduced by the HEp-2 assay kit and the CAD system in ANA indirect immunofluorescence test

M. Infantino[1] · F. Meacci[1] · V. Grossi[1] · M. Manfredi[1] ·
M. Benucci[2] · M. Merone[3] · P. Soda[3]

M. Infantino

Abstract According to the recent recommendations of the American College of Rheumatology, ANA Task Force, IIF technique should be considered the gold standard in antinuclear antibodies (ANAs) testing. To overcome the lack of standardization, biomedical industries have developed several computer-aided diagnosis (CAD) systems. Two hundred and sixty-one consecutive samples with suspected autoimmune diseases were tested for ANA by means of IIF on routinely HEp-2 assay kit (Euroimmun AG). Assignment of result was made if consensus for positive/negative was reached by at least 2 out of 3 expert physicians. ANA-IIF was also carried out using 3 CAD systems: Zenit G-Sight ($n = 84$), Helios ($n = 85$) and NOVA View ($n = 92$); human evaluation was repeated on the same substrate of each CAD system (Immco, Aesku and Inova HEp-2 cells, respectively). To anonymize the results, we randomly named these three systems as A, B and C. We ran a statistical analysis computing several measures of agreement between the ratings, and we also improved the evaluation by using the Wilcoxon's test for nonparametric data. Agreement between the human readings on routinely HEp-2 assay kit and human readings on CAD HEp-2 assay was substantial for A ($k = 0.82$) and B ($k = 0.72$), and almost perfect for C ($k = 0.89$). Such readings were statistically different only in case A. Comparing experts' readings with the readings of CAD systems, when the samples were prepared using CAD HEp-2 assay kits, we found almost perfect agreement for B and C ($k = 0.86$; $k = 0.82$) and substantial agreement for A ($k = 0.73$). Again, human and CAD readings were statistically different only in A. When we compared the readings of medical experts on routinely HEp-2 assay kit with the output of the CAD systems that worked using their own slides, we found substantial agreement for all the systems (A: $k = 0.62$; B: $k = 0.65$; C: $k = 0.71$). Such readings were not statistically different. The change of the assay kit and/or the introduction of a CAD system affect the laboratory reporting, with an evident impact on the autoimmune laboratory workflow. The CAD systems may represent one of the most important novel elements of harmonization in the autoimmunity field, reducing intra- and inter-laboratory variability in a new vision of the diagnostic autoimmune platform.

Keywords Indirect immunofluorescence · HEp-2 · Antinuclear antibodies · Automated classification system · Standardization

Abbreviations

ACR	American College of Rheumatology
ANA	Antinuclear antibodies
CAD	Computer-aided diagnosis
HEp-2	Human epithelial cells
IIF	Indirect immunofluorescence
IIM	Idiopathic inflammatory myopathies
FITC	Fluorescein isothiocyanate
SARD	Systemic rheumatic diseases
SjS	Sjögren's syndrome
SLE	Systemic lupus erythematosus
SSc	Systemic sclerosis

✉ M. Infantino
maria2.infantino@uslcentro.toscana.it

1 Immunology and Allergy Laboratory, S.Giovanni di Dio Hospital, Via Torregalli, 3, 50143 Florence, Italy

2 Rheumatology Unit, S.Giovanni di Dio Hospital, Via Torregalli, 3, 50143 Florence, Italy

3 Computer Systems & Bioinformatics Laboratory, Department of Engineering, University Campus Bio-Medico, Rome, Italy

 Springer

Introduction

The antinuclear antibodies (ANAs) are useful markers for the diagnosis, classification, prognosis and disease activity monitoring of ANA-associated rheumatic diseases (AARDs), such as systemic lupus erythematosus (SLE), systemic sclerosis (SSc), Sjögren's syndrome (SjS) and idiopathic inflammatory myopathies (IIM) [1, 2]. The detection of ANAs plays an important role in the global clinical management of AARDs patients, and it is a valuable resource to obtain the earliest possible diagnosis and to anticipate the clinical phenotype of individual patients. Due to the pivotal role of ANA testing a large reproducibility, specificity and accuracy are strongly required [3, 4], to limit false results and further inappropriate test requests. Traditionally the indirect immunofluorescence (IIF) on human epithelial cells (HEp-2) is used for ANA testing. IIF, which shows a multitude of native antigens, is a multiplex technique considered as a "natural array" that allows the detection of more than 30 different nuclear and cytoplasmic patterns [5]. Nevertheless, the IIF method is burdened by some limitations: the visual evaluation is time consuming, subjective, and it requires trained personnel and expert morphologists [6–11]. These disadvantages are the source of intra- and inter-laboratory discrepancies. Furthermore, the variability of IIF is strongly influenced by several issues inherent to the method, both biological and non-biological [12–22] (Table 1). During the last decades, given the growing request for ANA testing and the need of standardization, alternative techniques (e.g., enzyme-linked immunosorbent assay) and new available technologies (e.g., multiplex solid phases) have been proposed to replace IIF [23–33]. Interestingly, the innovation has led to a renaissance of the IIF method and, recently, the American College of Rheumatology (ACR) ANA Task Force has recommended the IIF on HEp-2 cells as the gold standard test for ANA detection [34]. ACR recommendations have led not only to a rapid increase of new techniques, but also to a challenge of the automated autoantibodies screening on HEp-2 cells [35, 36]. The actual autoimmunology laboratory scenario is characterized by a powerful driving force for an efficient workflow as a result of growing request of autoimmune diagnostic tests, regional restriction and limited reimbursement. There is a general agreement that technological innovation can help to tackle with these issues. In particular, these needs have motivated recent research efforts directed toward the development of computer-aided diagnosis (CAD) systems for reading and interpretation of ANA-IIF slides [37]. This research has also resulted in the availability of several commercial CAD systems, whose initial assessment has been recently described [38–45]. These systems are different in terms of

DNA counterstain, substrates throughput, run-time, types of recognized IIF ANA patterns and software [46, 47]. The use of a CAD system may strongly help to overcome most of IIF drawbacks and should be considered a potential reliable tool of standardization for a novel cost-effective autoimmune diagnostic. Furthermore, a CAD system may improve the laboratory quality certification by the introduction of internal quality control (ICQ) procedures and allowing an easy and reliable storage of images [48–51]. It can act as a newly global electronic data management platform able to relieve the increasing number of tests, and it can help the laboratory in time of personnel difficulty allowing to remotely analyze the images. Additionally, the introduction of a CAD system has the potential to reduce the inter-laboratory variability; indeed, for instance, it may help to reduce the nomenclature discrepancies between laboratories when describing IIF patterns and it can represent an objective criterion for a titer assignment. Besides, from a clinical point of view, the knowledge of IIF analytical variability is critical for the correct use and for the clinical interpretation of the ANA test. In fact an improper clinical interpretation of the test can lead to a misdiagnosis, inappropriate therapies and excessive health costs. A balance between available economic resources and growing health needs is the main goal of the recent advances in diagnostic technologies for AARDs [52–54]. In this respect, in this work, we focused our attention on the variability introduced in ANA indirect immunofluorescence test when a positive/negative classification has to be performed. As there is little data on inter-observer reading variability, this work aimed to study the burden introduced by two important issues, namely the HEp-2 assay kit and the CAD system. Indeed, we first assessed the variability among different kits when the ANA readings are performed according to the traditional visual inspection of the samples. As a second aim, we compared the outputs provided by the commercial CAD systems and those provided by the human readings. These combined data allowed assessing how much the substrate and the CAD system impact a routine workflow.

Materials and methods

Samples and visual interpretation

Two hundred and sixty-one consecutive samples with suspected autoimmune diseases were routinely screened for ANAs in the Laboratory of Immunology and Allergy of San Giovanni di Dio Hospital in IIF on HEp-2 cells substrate (Euroimmun AG, Luebeck, Germany). Mean age was 54 years (11–82 years), F/M ratio was 5/1. ANA-

Table 1 Biological, technical and operator-related variables influencing the ANA-IIF tests on HEp-2 cells

Variables		Description
Biological	HEp-2 cell culture	Cell morphology, culture technique, variable antigen expression, nuclear-to-cytoplasm ratio, actively dividing cells (mitosis number)
	Fixative	Alcohols, alcohols/acid, alcohols/acetone, formaldehyde
	Conjugate	Isotype specificity (mono/polyspecific, fab fragment), fluorescein isothiocyanate-to-protein (F/P) ratio, antibody-to-protein ratio, antibody content
	Assay reagents	pH and chemical characteristics
	Fluorescent dye	Concentration, adsorbing and emitting wavelengths, antibleaching agent
	Evans blue counterstain	Concentration
Technical	Preparation of slides	Manual, by an automated processing analyzer
	Microscope	Setting (magnification, filter), light intensity, light source (Hg vapor lamp, laser, diodes), camera signal-to-noise ratio, camera spectral sensitivity
	Dark room	Characteristics, setting
Operator related	Working cutoff dilution	Conventional (1:80), non-conventional (1:40, 1:160, etc.)
	Result transcription	Human oversight errors
	IIF pattern nomenclature	Different name for the same pattern, ambiguous description of patterns

IIF samples/slides were prepared at 1:80 titer using the automated pipetting device DAS AP16 IF Plus (DAS, Palombara Sabina, Rome, Italy) according to the conventional procedure. The slides were visually read using a fluorescence microscope (EUROStar II, Euroimmun AG, Luebeck, Germany) equipped with ocular $10\times$, an objective $40\times$, a 5 W LED whose wavelength of excitation light source ranges between 46 nm and 490 nm. Assignment of result was made if consensus for positive/negative was reached by at least 2 out of 3 expert physicians.

Automated ANA-IIF evaluation

ANA-IIF was carried out using 3 CAD systems and adopting the HEp-2 slides provided by each manufacturer. The CAD systems are: Zenit G-Sight, A. Menarini Diagnostics, Florence, Italy ($n = 84$); Helios, Aesku Diagnostics, Wendelsheim, Germany ($n = 85$); and NOVA View, Inova Diagnostics, San Diego, CA ($n = 92$). Consequently, HEp-2 slides were provided by Immco, Aesku and Inova, respectively. While testing the CAD systems in a fully automated fashion according to manufacturer's instructions and using the 1:80 titration, we had the assistance of a company specialist to avoid any wrong practice with the devices.

Once the slides were processed, and the positive/negative classifications were provided by the CAD systems, the three IIF experts performed visual evaluation of these samples.

To anonymize the results, we randomly named these three systems as A, B and C, and we kept blind the correspondence between these labels and the CAD systems. Their main features are summarized in the following.

Zenit G-Sight

Zenit G-Sight is an automated system designed for digitizing and displaying slides in immunofluorescence and for subsequent positive/negative and staining pattern classification of HEp-2 cells. It is also able to scan other IFA substrates such as ANCA, dsDNA, triple rat and mouse tissues, ICA and adrenal gland. The hardware consists of a microscope equipped with a motorized precision stage, a LED light-emitting source and a color camera. The image acquisition algorithm stitches the collected images using their relative positions, which are given by the known translations of the motorized stage. This provides a mosaic of the overall well area that can be used as a virtual microscope image. The autofocus algorithm makes use of an autoregulation procedure that, for every image of the mosaic, sets the gain and the exposure time of the camera. In this way, the appearance of image intensity is comparable between positive, weakly positive and negative samples. Each well undergoes positive/negative classification and staining pattern recognition. The former leverages on the analysis of signal intensity and distribution, whereas the latter can recognize homogeneous, nucleolar, speckled, centromere and mitochondrial patterns. The

system also provides a semiquantitative result for fluorescence intensity classification so that a sample is classified as negative, borderline and positive if its score is lower than 15 AU, between 15 and 20 AU, and larger than 20 AU, respectively. A cutoff level can be set to minimize the risk of false-negative classification. The manufacturer cutoff was 15 AU.

Helios

Helios is a platform that automatically performs the immunofluorescence pipetting and reading steps, integrated with positive/negative classification. It is based on an integrated autofocus epifluorescence microscope unit incorporating Nikon optics controlled by an own-engineered motor and LED light source. A unique characteristic is the approach used for mounting medium dispensation that, combined with the non-requirement for cover slip, enables complete processing without human intervention. The system works with the standard FITC (fluorescein isothiocyanate) fluorochrome, and no additional dye is needed.

The image acquisition stage automatically focuses and acquires the slide once the processing is completed; the user can define the desired numbers of images per well (between 1 and 10). The image classification module can provide positive/negative classification; it leverages on some image features such as the structure of the objects, the fluorescence signal intensity and the background/cells ratio. This module can work with samples exhibiting the following staining patterns: homogeneous, speckled, nucleolar, nucleolar dots, centromere, multiple nucleolar dots, cytoplasmic and cytoskeleton. The operator's check is needed to confirm results and to manually assign the pattern. Furthermore, the system can also estimate the endpoint titer for those wells classified as positive. This functionality allows reducing the number of well titrations needed to quantify the positivity level. The manufacturer cutoff was 48.

NOVA view

The NOVA View instrument consists of an automated and fully motorized inverted fluorescent microscope with LED light source, a CCD camera, a 40× objective and software running the CAD. NOVA View can acquire and interpret HEp-2, ANCA ethanol and formalin, dsDNA CLIFT slides. Slides processed by NOVA View have to be dyed with two conjugates: FITC and DAPI (4′,6-diamidino-2-phenylindole). This second dye is used as it strongly binds to A-T rich regions in DNA. The image processing algorithms use fluorescence information given by DAPI wavelengths to focus the samples, to locate and to segment the cells.

For each well in a slide, three to five images are acquired with both the DAPI and the FITC filter. The images of each cell must contain, at least, 25 interphase and 2 mitotic (metaphase) cells in total. Using FITC images, the system measures the average intensity in units named as light intensity units (LIUs), discriminating between positive and negative samples. The positive samples then undergo a staining pattern classification stage, which allows to distinguish between five basic fluorescent ANA patterns (homogeneous, speckled, centromere, nucleolar, nuclear dots). The operator's check is needed to confirm results. For wells containing positive reactions, the software can also estimate the endpoint titer (highest dilution that would give positive result) on the basis of nucleus LUI and detected pattern detected. The manufacturer cutoff was 48 LIUs.

Statistical analysis

The statistical analysis computed several measures of agreement between two ratings on fluorescence intensity classified as positive or negative. Our data basically consisted of two independent ratings provided by the medical expert and by the CAD system with respect to a dichotomous outcome: for this reason, we decided to compute the following indices, which are defined with respect to Table 2:

The overall agreement (oa): it is the proportion of cases for which raters 1 and 2 agree. It is defined as $oa = (a + d)/N$. This quantity is informative and useful, but, taken by itself, does not distinguish between agreement on positive ratings and agreement on negative ratings;

The positive and negative agreement denoted as pa and na, respectively. They measure the agreement relative to each category, and they are defined as $pa = 2a/(2a + b + c)$ and $na = 2d/(2d + b + c)$;

The Cohen's kappa (k) [55]: it is defined as $(oa - ca)/(1 - ca)$, where ca is hypothetical probability of chance agreement computed as $[(a + c)(a + b) + (b + d)(c + d)]/N^2$. Although the k value varies from a lower bound of -1 to an upper bound of 1, the usual region of

Table 2 Summary of binary ratings by two raters

	Rater 2		
	Positive	Negative	Total
Rater 1			
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	N

interest is $k > 0$. In the literature, the following guidelines for interpreting kappa values are used: $0 < k < 0.2$ implies slight agreement; $0.2 < k \leq 0.4$ implies fair agreement; $0.4 < k \leq 0.6$ implies moderate agreement; $0.6 < k \leq 0.8$ implies substantial agreement, and $0.8 < k \leq 1$ implies almost perfect agreement [56]. It is worth noting that Cohen's kappa copes with a limitation of oa score: oa can be high even when hypothetical raters randomly guess on each case according to prior probabilities. For instance, two raters would agree on the diagnosis if they simply guessed positive the large majority of times.

We also improved the evaluation process by using the Wilcoxon's test for nonparametric data, with 95 % confidence interval. Hence, p values lower than 0.05 were considered to indicate statistical significance. This test runs the pairwise comparison between two sets of readings, and it aims to detect significant differences between the two populations. In comparison with the parametric t test, it is safer since it does not assume normal distributions. Also, the outliers have less effect on the Wilcoxon test than on the t test [57].

Results

We analyzed the agreement between sets of fluorescence intensity classification. The readings we collected can be divided into three sets, namely:

- Set (i): it contains the fluorescence intensity classifications achieved as consensus of IIF experts in visual interpretation on samples prepared using the routinely HEp-2 assay kits, as described in Sect. 1.1;
- Set (ii): it contains the fluorescence intensity classifications as consensus of IIF experts in visual interpretation on samples prepared using the HEp-2 CAD kits, i.e., the kits provided by each CAD manufacturer (Sect. 1.2);
- Set (iii): it contains the fluorescence intensity classifications provided by each CAD system using its own substrate (Sect. 1.2).

These sets of readings were pairwise compared to measure the agreement, and the results are reported in Tables 3, 4 and 5. Table 3 reports the values of agreement between human readings on samples prepared using both the routinely HEp-2 assay kit and the assay kit provided by each CAD manufacturer. Hence, this table compares the readings in the aforementioned sets (i) and (ii). This comparison permitted us to evaluate the intra-observer variability in traditional visual interpretation induced by different HEp-2 assay kits. We observed that the overall positive and negative agreements ranged between 75.0% and 91.3; 61.8 and 92.0 %; 78.2 and 90.5 %, respectively.

Interestingly, while in case B and C, the values of the overall agreement, positive agreement and negative agreement were approximately equal; in case A, we observed a negative agreement larger than the positive. The different values of agreement reflected also in the Cohen's kappa: indeed, for assay kits A and B, the kappa value corresponded to substantial agreement, while in case C the larger scores got to almost perfect agreement. Note also that the readings on routinely and CAD assay kits were statistically different only in case A ($p < 0.001$).

Table 4 compares the IIF experts' readings with the readings of CAD systems, when the samples were prepared using CAD HEp-2 assay kits. Hence, this table compares the readings in the aforementioned sets (ii) and (iii). With these data, we evaluated the variability between the human and the automated readings when they work on the same substrate. We observed that the overall, positive and negative agreements ranged between 84.5 % and 89.4; 73.5 and 88.3 %; 80.6 and 90.3 %, respectively. In case A, we observed again a negative agreement larger than the positive. This observation reflected also in the Cohen's kappa values that in cases A corresponded to substantial agreement, whereas in cases B and C to almost perfect agreement. Human and CAD readings were statistically different only in case A ($p < 0.001$).

Table 5 compares the readings of medical experts on routinely HEp-2 assay kit with the output of the CAD systems that worked using their own slides. This corresponds to compare set (i) vs set (iii). In this case, the variability can be induced by both the use of different HEp-2 assay kits and by the use of a CAD system. We found that the overall positive and negative agreements ranged between 71.4 % and 76.1; 63.6 and 79.6 %; 71.1 and 76.5 %, respectively. Interestingly, in case A, we observed again a negative agreement larger than the positive, whereas in case of C the opposite situation happened. The use of different HEp-2 assay kits for humans combined with the use of a CAD system processing its own HEp-2 slides represented a case where the sources of variability were combined each other. This produced Cohen's kappa values that were lower than the others: in fact, they only implied substantial agreement between the readings. As a consequence, the human and automated readings were not statistically different for all the cases A, B and C ($p = 0.221$, $p = 0.221$, $p = 0.011$, respectively).

Discussion

Standardization in HEp-2 cell preparation and in the other factors responsible for the variability are the basis for a successful reading of IIF pattern in ANA diagnosis, therefore ensuring reliable results. The existing literature

Table 3 Agreement between the human readings on routinely HEp-2 assay kit and human readings on CAD HEp-2 assay

HEp-2 assay kit	Agreement			Cohen's k	P value
	Overall (%)	Positive (%)	Negative (%)		
A	75.0	61.8	81.4	0.62	<0.001
B	77.7	77.1	78.2	0.72	0.491
C	91.3	92.0	90.5	0.89	0.289

Table 4 Agreement between the human readings on CAD HEp-2 assay kit and CAD reading on its own kit

HEp-2 assay kit	Agreement			Cohen's k	P value
	Overall (%)	Positive (%)	Negative (%)		
A	84.5	73.5	89.1	0.73	<0.001
B	89.4	88.3	90.3	0.86	0.508
C	84.8	87.5	80.6	0.82	0.057

Table 5 Agreement between the human readings on routinely HEp-2 assay kit and CAD reading on its own kit

HEp-2 assay kit	Agreement			Cohen's k	P value
	Overall (%)	Positive (%)	Negative (%)		
A	71.4	63.6	76.5	0.62	0.221
B	71.8	70.0	73.3	0.65	0.221
C	76.1	79.6	71.1	0.71	0.011

on inter-observer reading variability does not focus on "real-life" autoimmunology laboratory. On the one side, it is well known in common laboratory practice that the ANA testing is subject to errors due to the HEp-2 assay kit used, working conditions and operators. On the other side, the previous studies showed great heterogeneity in the performance of the HEp-2 commercial assay kits. Furthermore, few data on how all the factors influencing the IIF of the ANA testing affect inter- and intra-observer reading variability are available [8, 9, 14, 37]. In this respect, to the best of our knowledge, this work is the first attempt to study the inter-observer reading variability of positive/negative fluorescence intensity classification in a "real-life" autoimmunology laboratory, focusing its attention on the burden introduced by the HEp-2 assay kits and by the use of a CAD system.

The fixative used for cell preparation is the major source of discrepancies [58, 59]. Copple et al. [14] comparing 5 assays for HEp-2 cell preparation in a SLE population showed that the global agreement between such assays was equal to 78 %. The authors claimed that the high variability in cell and nuclear morphology depended mostly on the fixative used. Another important cause of variability was the conjugate used. The guidelines published by the National Committee for Clinical Laboratory Standards offered a voluntary standard developed by consensus of the clinical laboratory testing community [60]. These guidelines compared antihuman IgG-specific conjugate, antihuman polyconjugate (G, A and M) and total IgG conjugate. The use of an antihuman IgG (Fc)-specific conjugate is preferred to enhance the positive predictive value of the ANA test, given that a polyconjugate or an IgG heavy/light chain conjugate might detect the IgM class antinuclear antibodies, which are usually clinically insignificant. According to the literature, our data confirmed how the choice of the HEp-2 assay kit may induce variability in IIF readings: indeed, the values of Cohen's kappa, when the IIF experts used both the routinely HEp-2 substrate and an alternative substrate (i.e., A, B, and C), showed that the agreement was satisfactory in two cases ($k = 0.618$ and $k = 0.718$) and in one case was almost perfect ($k = 0.894$). Interestingly, the rates of global, positive and negative agreement were larger than 90 % only in the third case, revealing the large impact of the assay kit in the decision-making process.

Several CAD systems for automated analysis of ANA testing have recently been developed, since the IIF test has to be considered the reference method for ANA determination, as ACR stated[34]. These systems adopt different materials and a different computational approach to

support the IIF experts in the reading phase. Indeed, they differ in the use of counterstain (DAPI, propidium iodide, none), in the use of substrate (most of them work only with one substrate), in the throughput, in the number of recognized staining patterns and, obviously, in the image classification approach adopted (e.g., features, learning algorithms). Although comparable performances between automated and conventional ANA-IIF testing for the interpretation of negative and positive samples have been reported in literature, discrepancies have been found when performing staining pattern recognition [46]. In fact, the CAD systems may fail in this latter task for the inherent difficulty to recognize mixed or "novel" fluorescence patterns and for misinterpretation issues arising when antibodies react with specific cell components. In addition, there is a strong need for an automated quantitative score assigned to the fluorescence intensity.

Our study focused on positive/negative classification since not all of the systems were able to recognize the same patterns and to predict image titer. On the one hand, when using the same assay kit (Table 4), we found an almost perfect concordance between CAD system and human readings for systems B and C ($k = 0.818$; $k = 0.857$), which reduced to substantial for system A ($k = 0.729$). On the other hand, the results of the experiments reported in Table 5 permitted us to observe the effects of introducing a CAD system in a laboratory that routinely uses an assay kit different from the assay kit of the CAD system. We found that their combined effect worsened the performance of agreement, thus confirming their burden. This work wants to be the first burdens comparison between the variability introduced by the assay kit and by the CAD system in IIF ANA testing.

Our data comparing visual and automated systems partially agreed with the global literature [42, 61, 62], and our concordance rates were found to be lower than those reported by Gorgi et al. [63] (CyclopusCADImmuno® versus visual IIF reading: overall agreement 90.5 %; $k = 0.8$). These differences could be expected since, unlike Gorgi et al. who considered AARDs cohorts and healthy donors, we used a routinely consecutive population of patients with suspected autoimmune diseases. This choice introduced more samples with borderline fluorescence intensity, although it is well known that CAD systems have lower recognition performance on such samples [64, 65]. The basis of a "real-life" study is also to work on a large population in order to consider possible human mistakes in the global IIF variability, as they can be rarer when considering small study population. This motivated the differences we found with respect to the results reported in the French national external quality assessment (EQA) on ANA detection [8]. In fact, this work could not be considered as a "real-life" study because it included mainly

strong titer sera and not borderline positive samples and because human tendency of a morphologist at microscope is to pay more attention for a small population or for an EQA sample, if compared to an unknown routine sample. This focus on few samples or on "special samples," as we could consider the EQA automatically may exclude human distraction mistakes and makes a study far from the normal working conditions.

Moreover, this work could be considered a "real-life" study for one more relevant issue: readers were people working in a laboratory, while in some previous studies IIF readers were not laboratory personnel but manufacturer laboratories or clinicians [9, 37]. Involving manufacturer laboratories or clinicians could introduce a bias because, for instance, a clinician who knows clinical features of the patients generally will find more easily what she/he is looking for at microscope.

With reference to the performance of the CAD systems we measured in this work, it is worth observing that we set the cutoff at the values provided by the manufacturer since we wanted to compare the different CAD systems performances at the same conditions. Nevertheless, the performance of each CAD system could be optimized by modifying the manufacturer cutoff according to sensitivity/specificity balance (clinical efficacy) opted by the laboratory. Considering that nowadays the ANA testing is no longer only a specialist request from immunology and rheumatology clinicians, leading to a decreased pre- and posttest probability and furthering misdiagnosis, it should be optimal to have a preference for the specificity in the CAD system cutoff setting, in order to reduce the false-positive results. With reference to positive/negative discrimination, the results achieved in this work suggested us that in a laboratory the change of assay kit and/or the introduction of a CAD system affect the laboratory reporting, with an evident impact on the autoimmune laboratory routinely reading.

Conclusion

Standardization of ANA testing is far from being completed. In fact the IFA method is labor intensive, subjective and prone to reader bias. A proposal of tight procedures both for users (working protocol) and for manufacturers (assay kits production) and a large use of international standards and independent calibrators could make easily the standardization process. Therefore, an international standardization of the HEp-2 assay kit (fixative used, conjugate etc.) and the introduction of a CAD system may represent two important elements of this process.

Actually CAD systems are defined as a virtual operator suggesting a very probable result that has to be validated to video by an expert in a second step of work. They cannot still replace human evaluation in any step of the work. Given the promising performance of the CAD systems in ANA positive/negative discrimination, our data lead us to hope that in the near future these systems, with further fine-tuning and/or further development, may become a reliable tool of IIF ANA screening. The operator will be involved only in the second step of workflow for the characterization of the positive samples and for the titration via serial dilutions. As a result, they will reduce operator time of work improving the diagnostic efficacy. In conclusion, the CAD systems may represent one of the most important novel elements of harmonization in the autoimmunity field, reducing intra- and inter-laboratory variability in a new vision of the diagnostic autoimmune platform.

References

1. Solomon DH, Kavanaugh AJ, Schur PH, American College of Rheumatology Ad Hoc Committee on Immunologic Testing Guidelines. Evidence based guidelines for the use of immunologic tests: antinuclear antibody testing. Arthritis Rheum. 2002;47(4):434–44.
2. Fritzler MJ. Challenges to the use of autoantibodies as predictors of disease onset, diagnosis and outcomes. Autoimmun Rev. 2008;7(8):616–20.
3. Tozzoli R, Bizzaro N, Tonutti E, Villalta D, Bassetti D, Manoni F, Piazza A, Pradella M, Rizzotti P, Italian Society of Laboratory Medicine Study Group on the Diagnosis of Autoimmune Diseases. Guidelines for the laboratory use of autoantibody tests in the diagnosis and monitoring of autoimmune rheumatic diseases. Am J Clin Pathol. 2002;117:316–24.
4. Brand F, Martin F, Philipp S, Rößler J, Hansen B, Anderer U. Difference in fluorescence pattern of cytoplasmic and nuclear antigens in cultivated human cells dependent on the applied fixation procedure. Cell Prolif. 2005;38:205.
5. Wiik A, Hoier-Madsen M, Forslid J, Charles P, Meyrowitsch J. Antinuclear antibodies: a contemporary nomenclature using HEp-2 cells. J Autoimmun. 2010;35(3):276–90.
6. Hiemann R, Büttner T, Krieger T, Roggenbuck D, Sack U, Conrad K. Challenges of 417 automated screening and differentiation of non-organ specific autoantibodies 418 on HEp-2 cells. Autoimmun Rev. 2009;9:17–22.
7. Fenger M, Wiik A, Høier-Madsen M, Lykkegaard JJ, Rozenfeld T, Hansen MS, et al. Detection of antinuclear antibodies by solid-phase immunoassays and immunofluorescence analysis. Clin Chem. 2004;50(11):2141–7.
8. Pham BN, Albarede S, Guyard A, Burg E, Maisonneuve P. Impact of external quality assessment on antinuclear antibody detection performance. Lupus. 2005;14:113–9.
9. Bizzaro N, Tozzoli R, Tonutti E, Piazza A, Manoni F, Ghirardello A, et al. Variability between methods to determine ANA, anti-dsDNA and anti-ENA autoantibodies: a collaborative study with the biomedical industry. J Immunol Methods. 1998;219:99–107.
10. Song L, Hennink EJ, Young IT, Tanke HJ. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. Biophys J. 1995;68:2588–600.
11. Marcolongo R, Ruffatti A, Morozzi G. Presentazione linee guida del Forum Interdisciplinare per la Ricerca sulle Malattie Autoimmuni (F.I.R.M.A.). Reumatismo. 2003;55:9–21.
12. Kavanaugh A, Tomar R, Reveille J, Solomon DH, Homburger HA, American College of Pathologists. Guidelines for clinical use of the antinuclear antibody test and tests for specific autoantibodies to nuclear antigens. Arch Pathol Lab Med. 2000;124(1):71–81.
13. Sack U, Conrad K, Csernok E, Frank I, Hiepe F, Krieger T, et al. Autoantibody detection using indirect immunofluorescence on HEp-2 cells, German EASI (European Autoimmunity Standardization Initiative). Ann N Y Acad Sci. 2009;1173:166–73.
14. Copple SS, Giles SR, Jaskowski TD, Gardiner AE, Wilson AM, Hill HR. Screening for IgG antinuclear autoantibodies by HEp-2 indirect fluorescent antibody assays and the need for standardization. Am J Clin Pathol. 2012;137(5):825–30.
15. Burlingame RW, Peebles C. Detection of antibodies. In: Pollard KM, editor. Autoantibodies and autoimmunity: molecular mechanisms in health and disease. Weinheim, Germany: Wiley-VCH; 2006. p. 159–88.
16. Egner W. The use of laboratory tests in the diagnosis of SLE. J Clin Pathol. 2000;53(6):424–32.
17. Fenger M, Wiik A, Høier-Madsen M, Lykkegaard JJ, Rozenfeld T, Hansen MS, et al. Detection of antinuclear antibodies by solid-phase immunoassays and immunofluorescence analysis. Clin Chem. 2004;50:2141–7.
18. Bradwell AR, Hughes RG, Karim AR. In: Detrick B, Hamilton RG, Folds JD, editors. Immunofluorescent antinuclear antibody tests in Manual of Clinical Laboratory Immunology. 7th ed. Washington: ASM press; 2006. p. 995–1006.
19. Emlen W, O'neill L. Clinical significance of antinuclear antibodies: comparison of detection with immunofluorescence and enzyme-linked immunosorbent assays. Arthritis Rheum. 1997;40(9):1612–8.
20. Tonutti E, Bassetti D, Piazza A, Visentini D, Poletto M, Bassetto F, et al. Diagnostic accuracy of ELISA methods as an alternative screening test to indirect immunofluorescence for the detection of antinuclear antibodies. Evaluation of Five Commercial Kits. Autoimmunity. 2004;37:171–6.
21. Tan EM, Smolen JS, McDougal JS, Butcher BT, Conn D, Dawkins R, et al. A critical evaluation of enzyme immunoassays for detection of antinuclear autoantibodies of defined specificities. I. Precision, sensitivity, and specificity. Arthritis Rheum. 1999;42:455–64.
22. Swaak AJ. Diagnostic significance of antinuclear antibodies in clinical practice. Ned Tijdschr Geneeskd. 2000;144(13):585–9.
23. Nifli AP, Notas G, Mamoulaki M, Niniraki M, Ampartzaki V, Theodoropoulos PA, et al. Comparison of a multiplex, bead-based fluorescent assay and immunofluorescence methods for the detection of ANA and ANCA autoantibodies in human serum. J Immunol Methods. 2006;311(1–2):189–97.
24. Smith J, Onley D, Garey C, Crowther S, Cahir N, Johanson A, et al. Determination of ANA specificity using the UltraPlex platform. Ann N Y Acad Sci. 2005;1050:286–94.
25. Ghirardello A, Bendo R, Rampudda ME, Bassi N, Zampieri S, Doria A. Commercial blot assays in the diagnosis of systemic rheumatic diseases. Autoimmun Rev. 2009;8(8):645–9.
26. Somers K, Govarts C, Stinissen P, Somers V. Multiplexing approaches for autoantibody profiling in multiple sclerosis. Autoimmun Rev. 2009;8(7):573–9.
27. Liu H, Norman GL, Shums Z, Worman HJ, Krawitt EL, Bizzaro N, et al. PBC screen: an IgG/IgA dual isotype ELISA detecting multiple mitochondrial and nuclear autoantibodies specific for primary biliary cirrhosis. J Autoimmun. 2010;35(4):436–42.

28. Lea P, Keystone E, Mudumba S, Kahama A, Ding SF, Hansen J, et al. Advantages of multiplex proteomics in clinical immunology: the case of rheumatoid arthritis: novel IgXPLEX: planar microarray diagnosis. Clin Rev Allergy Immunol. 2011;41(1): 20–35.

29. Bernardini S, Infantino M, Bellicampi L, Nuccetelli M, Afeltra A, Lori R, et al. Screening of antinuclear antibodies: comparison between enzyme immunoassay based on nuclear homogenates, purified or recombinant antigens and immunofluorescence assay. Clin Chem Lab Med. 2004;42(10):1155–60.

30. Tonutti E, Bassetti D, Piazza A, Visentini D, Poletto M, Bassetto F, et al. Diagnostic accuracy of ELISA methods as an alternative screening test to indirect immunofluorescence for the detection of antinuclear antibodies. Evaluation of five commercial kits. Autoimmunity. 2004;37(2):171–6.

31. Sinclair D, Saas M, Williams D, Hart M, Goswami R. Can an ELISA replace immunofluorescence for the detection of antinuclear antibodies? The routine use of anti-nuclear antibody screening ELISAs. Clin Lab. 2007;53(3–4):183–91.

32. Lopez-Hoyos M, Rodriguez-Valverde V, Martinez-Taboada V. Performance of antinuclear antibody connective tissue disease screen. Ann N Y Acad Sci. 2007;1109:322–9.

33. Op De Beeck K, Vermeesch P, Verschueren P, Westhovens R, Marien G, Blockmans D, et al. Detection of antinuclear antibodies by indirect immunofluorescence and by solid phase assay. Autoimmun Rev. 2011;10(12):801–8.

34. Meroni PL, Schur PH. ANA screening: an old test with new recommendations. Ann Rheum Dis. 2010;69:1420–2.

35. Hiemann R, Buttner T, Krieger T, Roggenbuck D, Sack U, Conrad K. Challenges of automated screening and differentiation of non-organ specific autoantibodies on HEp-2 cells. Autoimmun Rev. 2009;9:17–22.

36. Rigon A, Buzzulini F, Soda P, Onofri L, Arcarese L, Iannello G, et al. Novel opportunities in automated classification of antinuclear antibodies on HEp-2 cells. Autoimmun Rev. 2011;10: 647–52.

37. Rigon A, Soda P, Zennaro D, Iannello G, Afeltra A. Indirect Immunofluorescence in autoimmune diseases: assessment of digital images for diagnostic purpose. Cytomet B Clin Cytom. 2007;72B(6):472–7.

38. Hiemann R, Büttner T, Krieger T, Roggenbuck D, Sack U, Conrad K. Challenges of automated screening and differentiation of non-organ specific autoantibodies on HEp-2 cells. Autoimmun Rev. 2009;9:17–22.

39. Hiemann R, Hilger N, Michel J, Nitschke J, Böhm A, Anderer U, et al. Automatic analysis of immunofluorescence patterns of HEp-2 cells. Ann N Y Acad Sci. 2007;1109:358–71.

40. Egerer K, Roggenbuck D, Hiemann R, et al. Automated evaluation of autoantibodies on human epithelial-2 cells as an approach to standardize cell-based immunofluorescence tests. Arthrit Res Ther. 2010;12(2):R40.

41. Melegari A, Bonaguri C, Russo A, Luisita B, Trenti T, Lippi G. A comparative study on the reliability of an automated system for the evaluation of cell-based indirect immunofluorescence. Autoimmun Rev. 2012;11(10):713–6.

42. Kivity S, Gilburd B, Agmon-Levin N, Carrasco MG, Tzafrir Y, Sofer Y, et al. A novel automated indirect immunofluorescence autoantibody evaluation. Clin Rheumatol. 2012;31(3):503–9.

43. Voigt J, Krause C, Rohwäder E, Saschenbrecker S, Hahn M, Danckwardt M, et al. Automated indirect immunofluorescence evaluation of antinuclear autoantibodies on HEp-2 cells. Clin Dev Immunol. 2012;2012:651058.

44. Bossuyt X, Cooreman S, De Baere H, Verschueren P, Westhovens R, Blockmans D, et al. Detection of antinuclear antibodies by automated indirect immunofluorescence analysis. Clin Chim Acta. 2013;16(415):101–6.

45. Bonroy C, Verfaillie C, Smith V, Persijn L, De Witte E, De Keyser F, et al. Automated indirect immunofluorescence antinuclear antibody analysis is a standardized alternative for visual microscope interpretation. Clin Chem Lab Med. 2013;51(9): 1771–9.

46. Meroni PL, Bizzaro N, Cavazzana I, Borghi MO, Tincani A. Automated tests of ANA immunofluorescence as throughput autoantibody detection technology: strengths and limitations. BMC Med. 2014;12:38.

47. Bizzaro N, Antico A, Platzgummer S, Tonutti E, Bassetti D, Pesente F, Study Group on Autoimmune Diseases of the Italian Society of Laboratory Medicine, Italy, et al. Automated antinuclear immunofluorescence antibody screening: a comparative study of six computer-aided diagnostic systems. Autoimmun Rev. 2013;13:292–8.

48. Bonroy C, Verfaillie C, Smith V, Persijn L, De Witte E, De Keyser F, et al. Automated indirect immunofluorescence antinuclear antibody analysis is a standardized alternative for visual microscope interpretation. Clin Chem Lab Med. 2013;51:1771–9.

49. Bossuyt X, Cooreman S, De Baere H, Verschueren P, Westhovens R, Blockmans D, et al. Detection of antinuclear antibodies by automated indirect immunofluorescence analysis. Clin Chim Acta. 2013;415:101–6.

50. Bizzaro N, Antico A, Platzgummer S, Tonutti E, Bassetti D, Pesente F, Study Group on Autoimmune Diseases of the Italian Society of Laboratory Medicine, Italy, et al. Automated antinuclear immunofluorescence antibody screening: a comparative study of six computer-aided diagnostic systems. Autoimmun Rev. 2013;13:292–8.

51. Maenhout TM, Bonroy C, Verfaillie C, Stove V, Devreese K. Automated indirect immunofluorescence microscopy enables the implementation of a quantitative internal quality control system for anti-nuclear antibody (ANA) analysis. Clin Chem Lab Med. 2014;52(7):989–98.

52. Zafrir Y, Gilburd B, Carrasco MG, Kivity S, Sánchez-Castañón M, López-Hoyos M, et al. Evaluation of an automated chemiluminescent immunoassay kit for antinuclear antibodies in autoimmune diseases. Immunol Res. 2013;56(2–3):451–6.

53. Fritzler MJ, Fritzler ML. The emergence of multiplexed technologies as diagnostic platforms in systemic autoimmune diseases. Curr Med Chem. 2006;13(21):2503–12.

54. Utz PJ. Multiplexed assays for identification of biomarkers and surrogate markers systemic lupus erythematosus. Lupus. 2004;13(5):304–11.

55. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.

56. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1997;33:159–74.

57. Derrac J, García S, Molina D, Herrera F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput. 2011;1(1):3–18.

58. Monce NM Jr, Cappel VL, Saqueton CB. A comparison of two fixatives on IFA HEp-2 slides for the detection of antinuclear antibodies. J Immunoassay. 1994;15:55–68.

59. Hahm D, Anderer U. Establishment of HEp-2 cell preparation for automated analysis of ANA fluorescence pattern. Cytomet A. 2006;69:178–81.

60. Clinical and Laboratory Standards Institute (CLSI). Quality assurance of laboratory tests for autoantibodies to nuclear antigens: (1) Indirect fluorescent assay for microscopy and (2) microtiter enzyme immunoassay methods. Approved guideline; second edition. Wayne, PA: CLSI; 2006. CLSI document I/LA2-A2.

61. Copple SS, Jaskowski TD, Giles R, Hill HR. Interpretation of ANA indirect immunofluorescence test outside the

darkroomusing NOVA view compared tomanual microscopy. J Immunol Res. 2014;2014:149316.

62. Bertin D, Jourde-Chiche N, Bongrand P, Bardin N. Original approach for automated quantification of antinuclear autoanti-bodies by indirect immunofluorescence. Clin Dev Immunol. 2013;2013:182172.

63. Gorgi Y, Dhaouadi T, Sfar I, Haouami Y, Abdallah TB, Raso G, et al. Comparative study of human and automated screening for

antinuclear antibodies by immunofluorescence on HEp-2 cells. Int J Stat Med Res. 2015;4(3):270–6.

64. Soda P, Iannello G, Vento M. A multiple experts system for classifying fluorescence intensity in antinuclear autoantibodies analysis. Pattern Anal Appl. 2009;12(3):215–26.

65. Foggia P, Percannella G, Soda P, Vento M. Benchmarking HEp-2 cells classification methods. IEEE Trans Med Imaging. 2013;32(10):1878–89.