

A CRITICAL EVALUATION OF ENZYME IMMUNOASSAYS FOR DETECTION OF ANTINUCLEAR AUTOANTIBODIES OF DEFINED SPECIFICITIES

I. Precision, Sensitivity, and Specificity

ENG M. TAN, JOSEF S. SMOLEN, J. S. McDOUGAL, BRIAN T. BUTCHER, DOYT CONN,
ROGER DAWKINS, MARVIN J. FRITZLER, THOMAS GORDON, JOHN A. HARDIN,
JOACHIM R. KALDEN, ROBERT G. LAHITA, RAVINDER N. MAINI, NAOMI F. ROTHFIELD,
RUUD SMEENK, YOSHINARI TAKASAKI, WALTHER J. VAN VENROOIJ, ALLAN WIIK,
MERLIN WILSON, and JAMES A. KOZIOL

Objective. To determine the performance characteristics of enzyme-based immunoassay (EIA) kits for the detection of antinuclear and other autoantibodies of defined specificities.

Supported by research grants to individual authors from government funding agencies or charitable foundations in their respective countries. The grants are too numerous to list, and this is a collective acknowledgment by the authors that grant support was used in part to perform this study.

Eng M. Tan, MD: The Scripps Research Institute, La Jolla, California, and Chair, Antinuclear Antibody Standardization Subcommittee, Standardization Committee of the International Union of Immunological Societies, World Health Organization, Arthritis Foundation, and Centers for Disease Control and Prevention; Josef S. Smolen, MD: University of Vienna, Ludwig Boltzmann Institute for Rheumatology, and Lainz Hospital, Vienna, Austria; J. S. McDougal, MD: Centers for Disease Control and Prevention, Atlanta, Georgia; Brian T. Butcher, PhD, Doyt Conn, MD: Arthritis Foundation, Atlanta, Georgia; Roger Dawkins, MD: Royal Perth Hospital, Perth, Australia; Marvin J. Fritzler, MD: University of Calgary, Calgary, Alberta, Canada; Thomas Gordon, MD: Flinders Medical Centre, Bedford Park, Adelaide, Australia; John A. Hardin, MD: Medical College of Georgia, Augusta; Joachim R. Kalden, MD: University of Erlangen-Nuremberg, Erlangen, Germany; Robert G. Lahita, MD, PhD: St. Luke's/Roosevelt Hospital, New York, New York; Ravinder N. Maini, MD: Kennedy Institute of Rheumatology, London, UK; Naomi F. Rothfield, MD: University of Connecticut, Farmington; Ruud Smeenk, MD: Netherlands Red Cross Blood Transfusion Service, Amsterdam, The Netherlands; Yoshinari Takasaki, MD: Juntendo University School of Medicine, Tokyo, Japan; Walther J. van Venrooij, PhD: University of Nijmegen, Nijmegen, The Netherlands; Allan Wiik, MD: Statens Seruminstitut, Copenhagen, Denmark; Merlin Wilson, MD: Tulane University, New Orleans, Louisiana; James A. Koziol, PhD: The Scripps Research Institute, La Jolla, California.

Address reprint requests to Eng M. Tan, MD, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037.

Submitted for publication April 20, 1998; accepted in revised form August 12, 1998.

Methods. Nine manufacturers of EIA kits to detect antibodies of defined specificities participated in a study in which they received coded sera from the Centers for Disease Control and Prevention. These coded sera contained different dilutions of antibody of one specificity mixed with sera containing antibodies of other specificities. The manufacturers were asked to use their standard technology to determine antibody content and send the data to a committee of the International Union of Immunological Societies for analysis. The data were analyzed for sensitivity and specificity in the detection of anti-double-stranded DNA (anti-dsDNA), anti-single-stranded DNA, antihistone, anti-Sm, anti-U1 RNP, anti-SSA/Ro, anti-SSB/La, anti-Scl-70 (DNA topoisomerase I), antacentromere, and anti-Jo-1 antibodies. In addition, replicate samples were included in the coded sera to evaluate the precision of each EIA method.

Results. Lack of sensitivity and specificity was most evident in the anti-dsDNA and anti-Sm kits, although 2 kits for anti-dsDNA achieved acceptable sensitivity and specificity. Generally, anti-SSA/Ro, anti-SSB/La, anti-Scl-70, antacentromere, and anti-Jo-1 kits performed well. Many false-positive results were obtained with a multiple myeloma serum containing cryoprecipitates, but multiple myeloma sera without cryoprecipitates presented no problem in the EIA system. Precision, based on evaluation of replicate samples, varied from very good to poor.

Conclusion. No single manufacturer was clearly superior to others in terms of their products' overall

sensitivity, specificity, and precision. Areas that needed improvement were in kits for the detection of antibodies to dsDNA and to Sm antigen. Some EIA kits achieved good sensitivity and specificity. Individual manufacturers were informed of the performance of their respective kits so they could take measures to correct perceived deficiencies and thus improve the reliability of a group of important diagnostic assays used in the evaluation of systemic rheumatic diseases.

The presence of autoantibodies reactive with antigens located in the nucleoplasm, nucleolus, and cytoplasm has been documented in many different disease conditions (1–3). The frequency of autoantibodies is especially high in the systematic rheumatic diseases, such as systemic lupus erythematosus (SLE), scleroderma, mixed connective tissue disease, and Sjögren's syndrome. Detection of such autoantibodies has gained increasing acceptance by clinicians who use the information to aid or confirm diagnosis and in treatment followup.

Immunofluorescence microscopy (IFM) using human tissue culture cells such as HEp-2 as the substrate has been a widely used and standard assay for the detection of most of these autoantibodies. The method requires personal "hands-on" manipulation by the laboratory technologist, and the interpretation is dependent on the experience and knowledge of the microscopist. Both of these factors can lead to variability and lack of reproducibility between different laboratories. This was documented in a recent study on the range of antinuclear antibodies (ANA) in "healthy" individuals, in which the interlaboratory coefficients of variation ranged from 36% to 51% (4).

Because of the lack of reproducibility and inherent problems associated with IFM, alternative methods such as enzyme immunoassay (EIA) have been under investigation. The EIA method for detection of antibodies, which was first introduced in 1972 (5), eliminates the important variable of the expertise of the operator viewing the specimens under the microscope as well as other variables, such as differences in quality of instrumentation. Recent studies have compared the IFM ANA method with certain EIA kits that have been designated as ANA "screening" assays (6,7). These studies led to somewhat different conclusions with regard to the concordance between IFM ANA testing and EIA ANA testing. The differences might be due to noncomparable patient populations, as was pointed out by Emlen and O'Neill (7), or to the fact that the screening EIA ANA kits manufactured by different

companies are coated with differing numbers and amounts (concentrations) of nuclear, nucleolar, and cytoplasmic antigens. In addition, IFM ANA kits are capable of detecting a large repertoire of antigen-antibody systems, whereas the EIA ANA screening systems are capable of detecting only those antigen-antibody systems for which they were designed.

The ANA Subcommittee of the International Union of Immunological Societies (IUIS) Standardization Committee has been instrumental in establishing a bank of 10 different sera that have been designated as reference reagents for autoantibodies of different antigenic specificities (8–10). These reference reagents, containing antibodies of defined specificities, can be used to critically evaluate the performance of commercial EIA test kits that are marketed for the detection of single antigen-antibody systems.

Many manufacturers and distributors of EIA ANA kits were approached regarding their interest in participating in a study to evaluate the performance of their own EIA ANA kits. The specific aims of the study were to determine 1) the sensitivity and specificity of EIA ANA kits, 2) whether mixtures of antibodies of different specificities would interfere with detection, 3) whether multiple myeloma sera containing high concentrations of immunoglobulins would cause problems in the assays, and 4) the intralaboratory reproducibility of EIA ANA kits.

MATERIALS AND METHODS

Twenty companies that were known purveyors of EIA kits were approached in order to determine their interest and willingness to participate in this study. The manufacturers were advised that they would be sent coded sera containing mixtures of the Arthritis Foundation/Centers for Disease Control and Prevention (AF/CDC) reference reagents, and that they were to use their own test kits to analyze the antibody content of these sera and report the data, preferably in optical density (OD) units (or in lieu of that, in their own arbitrary units). They were informed that this study was designed to critically evaluate the performance of EIA-based methods for detection of autoantibodies, and that the data would be published as a comprehensive evaluation of this methodology without divulgence of the specific performance of any individual manufacturer. Initially, 11 manufacturers agreed to participate, but 2 manufacturers subsequently withdrew. The 9 participating manufacturers were (in alphabetical order): Cambridge Life Sciences (Cambs, UK), Elias (Freiburg, Germany), Helix Diagnostics (Sacramento, CA), Immunoconcepts (Sacramento, CA), Imtec Immunodiagnostika (Zepernick, Germany), Inestar (Stillwater, MN), Inova Diagnostics (San Diego, CA), MBL (Nagoya, Japan), and Shield Diagnostics (Dundee, Scotland).

Table 1. Key to serum samples: relative volumes of reagents in each sample*

Sample	CDC1 (dsDNA)	CDC2 (SSB/La)	CDC4 (U1 RNP)	CDC5 (Sm)	CDC7 (SSA/Ro)	CDC9 (Scl-70)	CDC3 (Speckled)	CDC6 (Nucleolar)	CDC8 (Centromere)	CDC10 (Jo-1)	MM (Fz)	MM (Ba)	MM (Fr)	NHS
A	4×	2×	1×											
B	2×			1×	4×									
C	1×		4×		2×									
D		4×	2×	1×										
E		2×		4×	1×									
F	2×	1×			4×									
G		1×	4×	2×										
H			1×	4×	2×									
I	1×		2×	4×										
J		2×		1×	4×									
K						4×	1×	2×						
L						2×	1×	4×						
M						1×	4×	2×						
N							4×	1×	2×					
O						1×		2×	4×					
P						2×	1×		4×					
Q										7×				
R											7×			
S												7×		
T													7×	
U														7×

* CDC = Centers for Disease Control and Prevention; dsDNA = double-stranded DNA; MM = multiple myeloma; NHS = normal human serum.

Design of test samples. Serum samples (Table 1) were prepared by a research technician (Martha Byrd, CDC). For the sake of brevity, the AF/CDC standard reagents are designated CDC1, CDC2, etc. The relative volumes of the 21 samples are shown as multiples of 4×, 2×, and 1× the unit volume. For example, the vertical column of CDC1 (Table 1) shows that this reference reagent was used in different relative volumes of 4×, 2×, and 1×, and the horizontal line for sample A shows that this sample contained 4× unit volumes of CDC1, 2× unit volumes of CDC2, and 1× unit volume of CDC4. With this system, the sensitivity, specificity, and dose-response of different test kits could be evaluated, and it could be ascertained whether antibodies of other specificities would interfere with each other in the EIA system.

The study was designed so that mixtures of antibodies would contain different proportions of anti-double-stranded DNA (anti-dsDNA), anti-SSB/La, anti-U1 RNP, anti-Sm, and anti-SSA/Ro (see first 5 columns of Table 1), because such mixtures might be expected in diseases such as SLE. Mixtures of anti-Scl-70 (CDC9), antinucleolar antigens (CDC6), and anticentromere (CDC8) might be seen in scleroderma. In addition, use of antibody of a defined specificity (e.g., anti-SSB/La [CDC2] at 4×, 2×, and 1× relative volumes) made it possible to examine whether EIA could be used for quantitation of antibody content.

As seen in Table 1, multiple myeloma sera, which were used undiluted (7×), were included in samples R, S, and T. These multiple myeloma sera were a gift from Dr. H. Spiegelberg (University of California, San Diego) and contained an average of 35 mg/ml of IgG. Sera Fz and Ba were of the IgG1 subclass, and Fr was of the IgG3 subclass. Serum Ba was known to contain cryoprecipitates, but the other 2 did not. These multiple myeloma sera were included in the study to

determine whether high concentrations of immunoglobulin or cryoprecipitates might lead to false-positive results. Serum CDC10, containing anti-Jo-1, and normal human serum were also used undiluted.

Each participating manufacturer received 26 samples for analysis: 1 aliquot from each of samples A–U, plus 5 replicate samples. For each laboratory, 2 of the 5 replicate samples were randomly selected from samples A–J, 2 were randomly selected from samples K–P, and 1 was randomly selected from samples Q–U. A coding scheme was adopted to ensure that the manufacturers' laboratory personnel remained blinded to the identity of the serum samples. Only the biostatistician (JAK) and the technician at the CDC knew details of the randomization and coding scheme.

Report forms were prepared for use by the participants. For each test sample, manufacturers' laboratory personnel were requested to determine (in duplicate) optical densities at serum dilutions of 1:100, 1:400, 1:1,600, and at the manufacturer's recommended dilution (if different). In addition, manufacturers were asked to indicate whether their kits gave positive or negative results for each antibody at the recommended dilution. Findings and interpretations could be omitted if a manufacturer's kit was not designed to identify a particular autoantibody. For example, none of the test kits was designed to detect antibodies to fibrillarin (contained in CDC6), and only 2 manufacturers had test kits designed to detect antibodies to centromere antigens (CDC8).

Individual characteristics of the ANA reference sera. Rigorously defined reference sera are the gold standards for autoantibody determinations. The CDC reference sera have been widely used as standards since they were introduced in 1982, and the methods used to define and establish the standards have been published (8,10). For example, the anti-

Table 2. Operational definition of antibody content of Arthritis Foundation/Centers for Disease Control and Prevention reference sera

Reference serum	Antibody*									
	dsDNA	ssDNA	Histone	Sm	U1 RNP	SSA/Ro	SSB/La	Scl-70	Centromere	Jo-1
1	+	+	+	+(weak)						
2										
3				+(weak)	+	+(weak)	+			
4					+					
5				+						
6										
7							+			
8									+	
9								+		
10										+

* dsDNA = double-stranded DNA; ssDNA = single-stranded DNA.

dsDNA standard (CDC1) was established using Millipore filter assay methods, the Farr radioimmunoassay, and *Crithidia luciliae* immunofluorescence assays (8). To confirm these determinations, a separate study was undertaken, in which each of the 10 reference sera (CDC1–CDC10) was sent to the 9 manufacturers. The manufacturers were asked to

report the presence or absence of the ANAs listed above, as well as the corresponding OD values at their respective recommended dilution levels. Seven of the 9 complied with this request. From these results, together with those obtained previously (7,9), a consensus was established regarding the specific autoantibody profiles of the reference sera; this is shown in Table 2. Operating characteristics (sensitivity and specificity) of the various kits were determined relative to these standards.

Data analysis. Intralaboratory variability was assessed with coefficients of variation (CV). Within each manufacturer's laboratory, CV were calculated from the OD (or units/ml) at the recommended serum dilution, using the replicate results with the duplicate test samples. Box plots were prepared from the CV within each laboratory. In addition, the sets of observed CV were used to partition the 9 manufacturers into relatively homogeneous subgroups in terms of precision (reproducibility), with smaller CV connoting greater precision. This partitioning was effected using the method of Cox and Spjotvoll (11), as applied with the nonparametric Kruskal-Wallis procedure (12) for comparison of CV across laboratories.

For every kit, operating characteristics (i.e., sensitivity and specificity for each ANA) were determined using the pooled data of all test samples at the recommended dilution level. The manufacturers' interpretations were adopted for designation of a positive assay result. A minority of the manufacturers also reported results that were called "borderline" or "weakly positive." For purposes of statistical analysis, it was decided to include as positive only those that were called "definitely positive." Overall sensitivity and specificity for each manufacturer were calculated by pooling individual sensitivity and specificity values across the ANAs tested by at least 8 of the manufacturers. This analysis provides a global sense of the sensitivity or specificity of a particular kit relative to others. As before, the Cox-Spjotvoll method as applied with standard chi-square statistics (since assay outcomes here are dichotomous) was used to partition the 9 manufacturers into relatively homogeneous subgroups, in terms of overall sensitivity and overall specificity. Spearman's nonparametric rank correlation (13) was used to assess associations between the orderings of

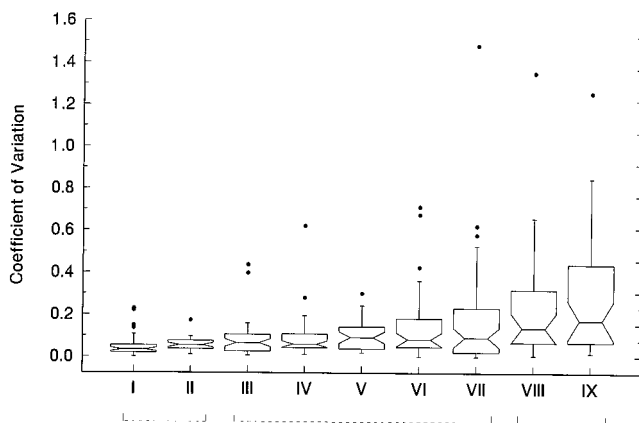


Figure 1. Grouped box plots of manufacturer reproducibility. Each box summarizes the reproducibility of results from a particular manufacturer, as determined from coefficients of variation from the optical densities (or units/ml) at that manufacturer's recommended serum dilution, using all of the replicate results with the 5 duplicate test samples. In each box, the median is depicted by a horizontal line segment within the rectangle, and the upper and lower quartiles of the data are depicted by the top and bottom of the rectangle. The notches in the rectangle approximate a 95% confidence region about the median. The vertical lines at each end of the box approximate a central 95% confidence region for the entire sample values. Closed circles denote observed values outside these limits. Also indicated at the bottom of the figure is an ordering of the manufacturers into 3 subgroups on the basis of reproducibility: a group of 2 manufacturers with relatively high reproducibility (I and II), a group with relatively moderate reproducibility (III–VII), and a group with relatively poor reproducibility (VIII and IX). This ordering was derived with the method of Cox and Spjotvoll (11) (see Materials and Methods).

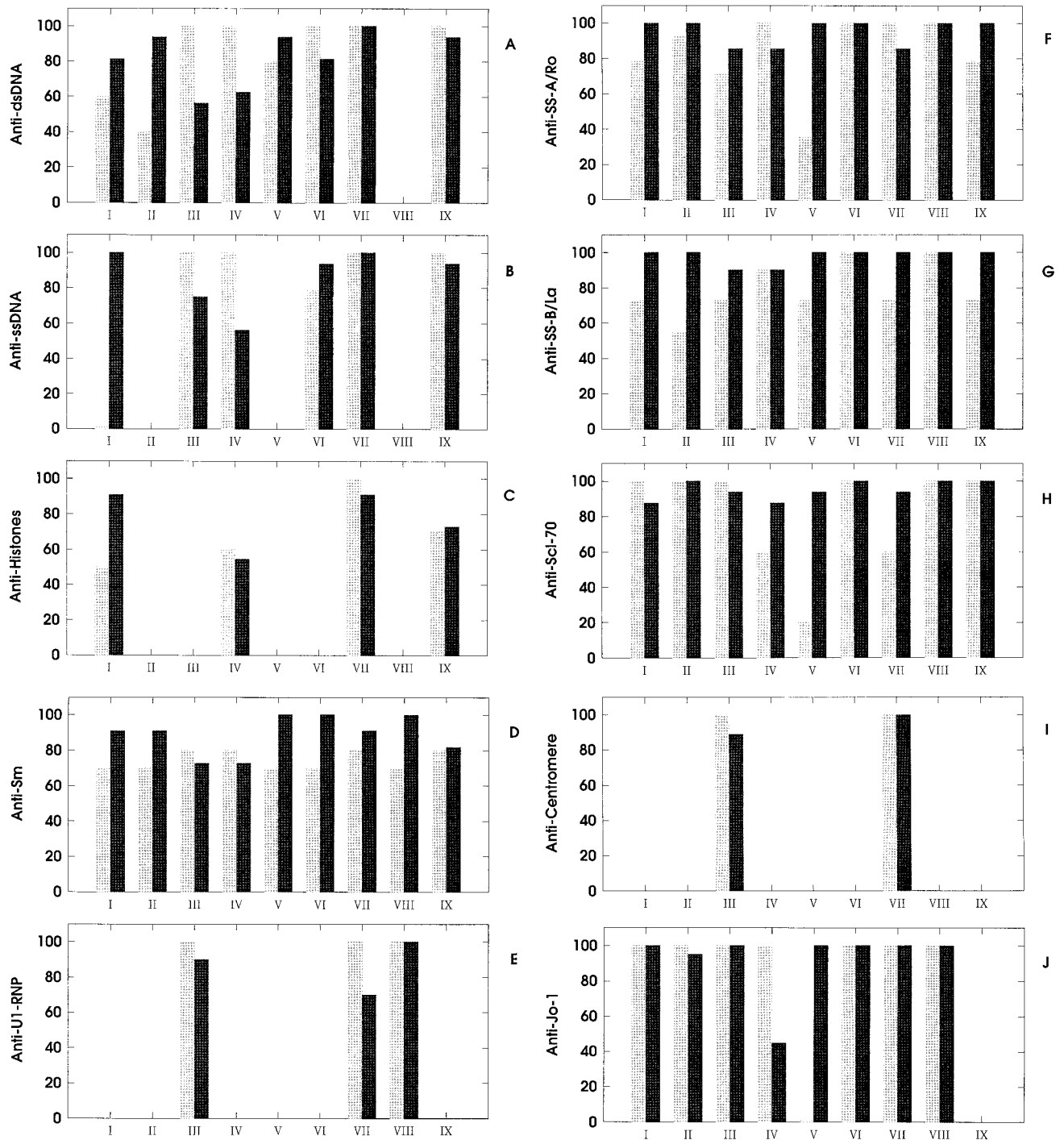


Figure 2. Individual sensitivity and specificity findings for the 9 manufacturers, designated I–IX as in Figure 1, for anti–double-stranded DNA (anti-dsDNA) (A), anti–single-stranded DNA (anti-ssDNA) (B), anti-histone (C), anti-Sm (D), anti-U1-RNP (E), anti-SSA/Ro (F), anti-SSB/La (G), anti-Scl-70 (H), anti-centromere (I), and anti-Jo-1 (J). Designation of assay positivity or negativity was made by each manufacturer, using the recommended dilution level for that kit. Shaded bars denote percent sensitivity and solid black bars percent specificity.

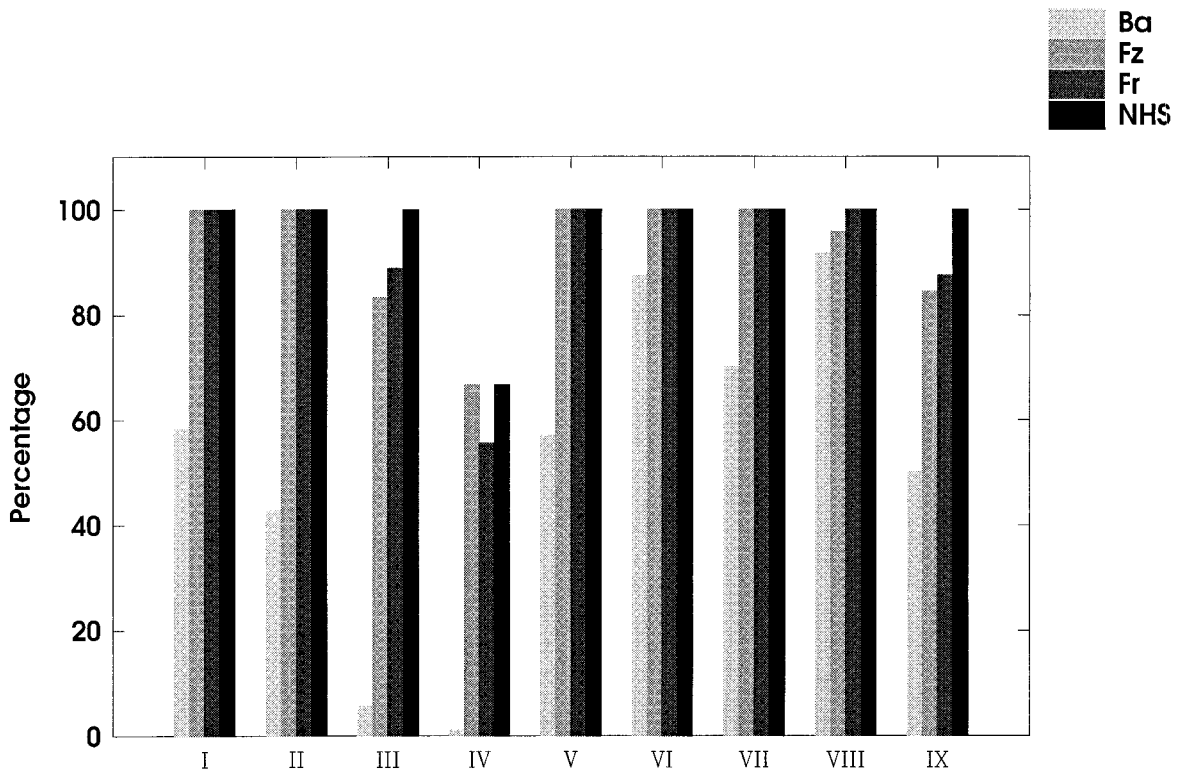


Figure 3. Specificity findings for the 9 manufacturers for multiple myeloma sera, Ba, Fz, and Fr, and normal human serum (NHS). False-positive results were determined from the manufacturer’s designation of positivity for any antinuclear antibody, at the recommended dilution level.

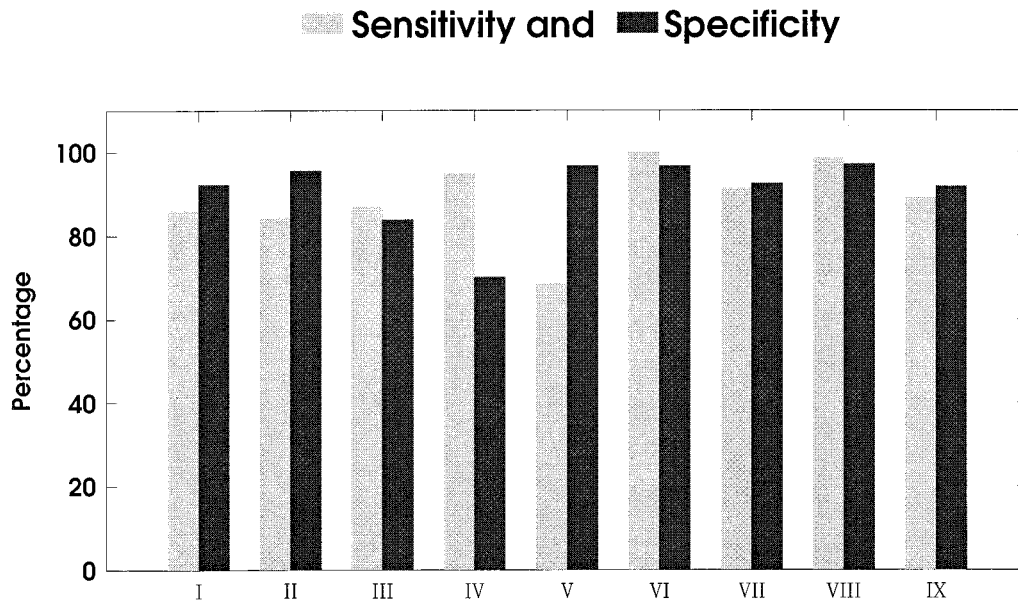


Figure 4. Overall sensitivity and specificity findings for the 9 manufacturers. Overall values were calculated by pooling each manufacturer’s results for assay positivity or negativity, for the antinuclear antibodies tested by at least 8 of the manufacturers (see Figure 2).

the manufacturers in terms of precision, sensitivity, and specificity.

RESULTS

Figure 1 depicts box plots of the CV calculated from the replicate results with the 5 duplicate test samples. The 9 manufacturers could be partitioned into 3 subgroups: an initial grouping of 2 manufacturers, henceforth denoted I and II, with excellent precision (reproducibility); a second grouping of 5 manufacturers, denoted III–VII, with good precision; and a third grouping of 2 manufacturers, denoted VIII and IX, with relatively poor precision compared with the others (Figure 1). The precision ordering of these test kits depends on several factors, including the quality of the test kits, the experience of the technicians performing the assays, and the number of different antigen–antibody specificities analyzed by the manufacturer.

The above reproducibility data were assessed solely from the replicate samples. In contrast, sensitivity and specificity assessments were based on different data, obtained from test sera containing anti-dsDNA, anti-single-stranded DNA (anti-ssDNA), antihistone, anti-Sm, anti-U1 RNP, anti-SSA/Ro, anti-SSB/La, anti-Scl-70, anticentromere, and anti-Jo-1 (Figures 2A–J). Not all of the manufacturers had kits designed to assay all ANAs: all manufacturers had kits capable of detecting anti-Sm, 1 of 9 kits did not test for anti-dsDNA (Figure 2A), 6 of 9 did not test for anti-U1 RNP (Figure 2E), and only 2 kits tested for anticentromere (Figure 2I).

The ideal kit would have 100% sensitivity and 100% specificity for the antigen–antibody reaction being assessed. Poor performance of the kit would be manifested by both low sensitivity and low specificity or by substantial differences between sensitivity and specificity, such as high sensitivity and low specificity or vice versa.

Some test kits had both high sensitivity and high specificity, e.g., manufacturer VII's kit for anti-dsDNA (100% sensitivity and 100% specificity; Figure 2A), but this manufacturer's test kits for anti-SSB/La and anti-Scl-70 (Figures 2G and H) showed lower sensitivity compared with test kits of several other manufacturers. In general, many of the test kits performed exceptionally well, such as those for anti-SSB/La, anti-Scl-70, anticentromere, and anti-Jo-1 (Figures 2G–J). However, every manufacturer showed deficiencies, to a greater or lesser extent, in the performance of their test kits for some of the antigen–antibody specificities.

Figure 3 shows the specificities of the kits relative

to multiple myeloma sera Ba, Fz, and Fr, and normal human serum (test samples R, S, T, and U, respectively). No corresponding sensitivity results are depicted, since these serum samples do not contain ANA. The data in this figure demonstrate that multiple myeloma sera with high concentrations of IgG generally do not give false-positive results in EIA systems except when a serum contains cryoprecipitates (serum Ba). In this situation, some manufacturers' test kits (III and IV) gave false-positive readings in the majority of their assay systems, whereas other manufacturers (VI and VIII) registered false-positive results in only 1 or 2 of their assay systems.

Overall sensitivities and specificities of the manufacturers' kits are plotted in Figure 4. Interpretation of the data should focus on 2 important factors: 1) both sensitivity and specificity should be high, and 2) the differential between sensitivity and specificity should be slight. The majority of manufacturers' kits performed relatively well in fulfilling these requirements, except for kit IV (high sensitivity but low specificity) and kit V (low sensitivity and high specificity). The former performance characteristic would cause many false-positive results, and the latter would cause many false-negative results.

Figure 5 depicts the relative orderings of the 9 manufacturers separately in terms of overall sensitivities and overall specificities and the groupings established with the Cox-Spiotvoll procedure. It is clear that the patterns of overall sensitivity and specificity are remarkably different. The overall sensitivity of 1 subgroup (VI, VIII, IV) was quite high, and that of the other kits was reasonably high with the notable exception of kit V. The manufacturers are grouped quite differently with regard to overall specificity: 7 of the 9 kits were clustered together, with overall specificity substantially higher than that of kits III and IV. It should be noted that the orderings bear little relationship to each other or to the precision ordering shown in Figure 1. Spearman's rank correlation between the sensitivity and specificity orderings of the manufacturers was only 0.30; the rank correlations between the precision ordering and the sensitivity and specificity orderings, respectively, were -0.53 and -0.49 .

The manufacturers provided information on the manner in which the autoantigens in their kits were prepared, since it was possible that this might be a factor in the performance characteristics. Seven of the 9 manufacturers used animal tissue antigens, which were affinity purified with autoantibodies. Only 2 manufacturers (III and VII) used recombinant antigens obtained as expression products from complementary DNA clones.

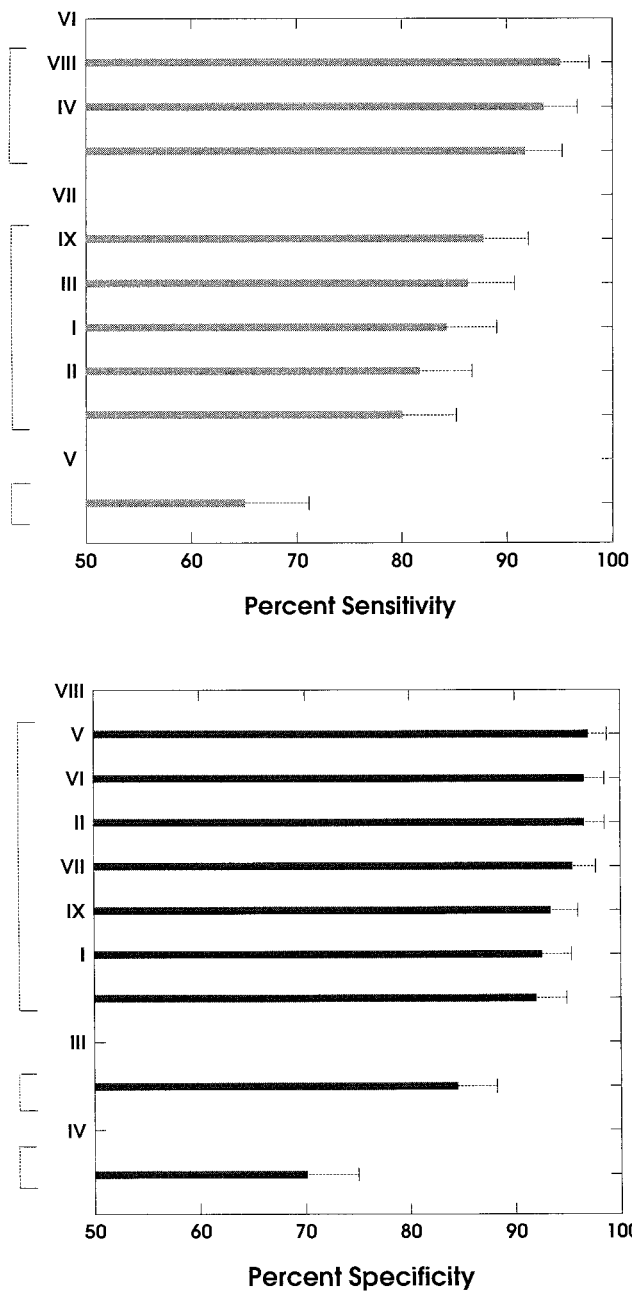


Figure 5. Orderings of the manufacturers into 3 subgroups with regard to overall sensitivity and overall specificity. Orderings were effected by means of the method of Cox and Spjøtvoll (11) (see Materials and Methods). The horizontal line extending from the right of each bar depicts 1 SD for the observed sensitivity or specificity.

As can be seen from Figure 4, the overall sensitivity and specificity of the kits from manufacturers III and VII did not differ significantly from the performance of the larger group that used affinity-purified antigens.

Analysis of the data showed that mixtures of antibodies of different specificities (Table 1) did not interfere with each other. This conclusion could be reached by examining the OD readings for a particular antibody such as anti-SSB/La at a fixed dilution, but admixed with antibodies of other specificities (for example, samples A, E, and J). The OD readings for anti-SSB/La were similar in all 3 samples, and similar findings were observed for other samples of mixed antibodies.

DISCUSSION

The present study had the full cooperation of 9 manufacturers of enzyme-based ANA test kits, and this unprecedented international effort has allowed a critical evaluation of EIA ANA test systems. With the information gained, it is possible to answer many of the questions raised above and to address the specific aims of the study. With regard to sensitivity and specificity of the test kits, the performance data show that the kits can be divided into 2 groups. For the first group of kits, designed to measure antibodies to dsDNA, ssDNA, histones, and Sm, the majority of manufacturers need improvement; for the second group, designed to measure antibodies to U1 RNP, SSA/Ro, SSB/La, Scl-70, centromere, and Jo-1, the majority of the manufacturers show good performance. It is of interest that the kits for anti-dsDNA show great variations in sensitivity and specificity among manufacturers I–VI (Figure 2A), but the sensitivity and specificity data for kits VII and IX reveal very good performance characteristics. These latter 2 manufacturers reported that they used plasmid DNA as the source of their antigens, but this clearly is not the only factor influencing accuracy, because kit III, showing 100% sensitivity but poor specificity, also had plasmid DNA as the antigen. The other manufacturers used cellular DNA (e.g., from calf thymus) as the source of antigen. Nevertheless, the data in Figure 2A show that it is possible to achieve acceptable sensitivity and specificity for anti-dsDNA, as demonstrated with kits VII and IX.

Although only 4 kits were capable of testing for antihistone antibodies, 1 of these (kit VII), using histones purified from tissue antigens, showed high sensitivity and specificity (Figure 2C). The greatest problem encountered with all test kits was measurement of anti-Sm antibody (Figure 2D). All test kits were deficient in terms of sensitivity, but the specificity of some kits was good. The problem associated with this antigen–antibody system could not be attributed to the use of either affinity-purified tissue antigens or recombinant

antigens, since both were being used in different kits. Since the Sm antigen system is a complex of many proteins, it is possible that the sensitivity could be enhanced by using recombinants of several antigens, including the E-F-G complex of the Sm system (14). As shown in Figure 2E, only 3 test kits were designed to measure anti-U1 RNP antibodies. Kit III, using recombinant protein, and kit VIII, using tissue antigens, performed well in terms of both sensitivity and specificity.

In contrast to the variable performance of the test kits described above, the performance of the kits designed to detect anti-SSA/Ro, anti-SSB/La, anti-Scl-70, anticentromere, and anti-Jo-1 was generally good in that some kits in each category achieved 100% sensitivity and specificity (Figures 2F-2H). There were a few exceptions, as can be seen in these figures.

Figure 3 demonstrates that the presence of high levels of IgG in multiple myeloma sera does not necessarily cause false-positive test results; several kits (I, II, V, VI, VII, and VIII) showed almost perfect specificity. However, cryoprecipitates caused problems: false-positive results were recorded with every test kit when specimen Ba, containing cryoprecipitates, was used. Some kits (e.g., IV) registered false-positive results for all antibody specificities, whereas others (e.g., VIII) registered false positivity for only 1 antibody specificity.

Figure 1 shows the rank ordering of intralaboratory reproducibility when the data were evaluated with regard to the replicate samples sent to each manufacturer. It can be seen that the range of coefficients of variation from kit I to kit IX was quite considerable. Many factors could be involved, including the expertise of the technicians performing the tests, the quality of the kits themselves, and the instrumentation used by each manufacturer. More insight into the reasons for intralaboratory variability might be afforded by subsequent analysis comparing the precision obtained by various International Union of Immunological Societies laboratories, using the same test kits as those used by the manufacturers in the present study. Analysis of these data is under way.

It should be emphasized that the rank orderings regarding precision (intralaboratory variability; Figure 1) and the rank orderings for sensitivity and specificity (Figure 5) do not demonstrate any correlation that might imply a definite overall superiority of test kits from one manufacturer versus another. As noted above, there was no direct correlation between the precision ordering and the rank orderings of sensitivity and specificity.

In general, this study demonstrates that EIAs

need improvement for certain antigen-antibody systems, especially anti-dsDNA and anti-Sm. However, the data also show that for every one of the antigen-antibody systems analyzed, some test kits achieved 100% sensitivity as well as 100% specificity. For several antigen-antibody systems, the majority of the test kits performed remarkably well. These data can be interpreted to mean that the EIA system for detection of autoantibody has the potential of being accurate and reliable. If intralaboratory variability (i.e., precision) can be improved, perhaps with better design of the test kits or advances in instrumentation, future versions of this assay system would be welcome. These improved versions would help to standardize assays for antibodies for which other detection systems introduce too much inaccuracy and variability.

The manufacturers that participated in this study have been informed of their own identities in the numbering system so they can identify the deficiencies in their own test kits. It is hoped that they will be able to make use of this information to improve the performance of their kits. It might be expected that investigators who use the immunoassay kits of a manufacturer participating in this study would be interested in the performance of the kits. We hope that when a manufacturer is approached for such information, the company will be willing to volunteer the information and perhaps be able to show that deficiencies have been corrected. If this happens, a collaboration such as this between academia and industry would have achieved the purpose of improving a group of important diagnostic assays.

ACKNOWLEDGMENTS

The authors are grateful to Martha Byrd and Lloyd Slivka for their invaluable assistance.

REFERENCES

1. Rose NR, Mackay IR. The autoimmune diseases II. San Diego: Academic Press, Inc.; 1992.
2. Von Muhlen CA, Tan EM. Autoantibodies in the diagnosis of systemic rheumatic diseases. *Semin Arthritis Rheum* 1995;24:323-58.
3. Van Venrooij WJ, Maini RN, editors. Manual of biological markers of disease. Dordrecht (The Netherlands): Kluwer Academic Publishers; 1996.
4. Tan EM, Feltkamp TEW, Smolen JS, Butcher B, Dawkins R, Fritzler MJ, et al. Range of antinuclear antibodies in "healthy" individuals. *Arthritis Rheum* 1997;40:1601-11.
5. Engvall E, Perlmann P. Enzyme-linked immunosorbent assay, ELISA. III. Quantitation of specific antibodies by enzyme-linked anti-immunoglobulin in antigen-coated tubes. *J Immunol* 1972; 109:129-35.
6. Jaskowski TD, Schroder C, Martins TB, Mouritsen CL, Litwin

- CM, Hill HR. Screening for antinuclear antibodies by enzyme immunoassay. *Am J Clin Pathol* 1996;105:468-73.
7. Emlen W, O'Neill L. Clinical significance of antinuclear antibodies: comparison of detection with immunofluorescence and enzyme-linked immunosorbent assays. *Arthritis Rheum* 1997;40:1612-8.
 8. Tan EM, Fritzler MJ, McDougal JS, McDuffie FC, Nakamura RM, Reichlin M, et al. Reference sera for antinuclear antibodies. I. Antibodies to native DNA, Sm, nuclear RNP and SS-B/La. *Arthritis Rheum* 1982;25:1003-5.
 9. Tan EM, Feltkamp TEW, Alarcon-Segovia D, Dawkins RL, Homma H, Kalden JR, et al. Reference reagents for antinuclear antibodies [letter]. *Arthritis Rheum* 1988;31:1331.
 10. Smolen JS, Butcher B, Fritzler MJ, Gordon T, Hardin J, Kalden JR, et al. Reference sera for antinuclear antibodies. II. Further definition of antibody specificities in international antinuclear antibody reference sera by immunofluorescence and Western blotting. *Arthritis Rheum* 1997;40:413-8.
 11. Cox DR, Spjotvoll E. On partitioning groups of means. *Scand J Stat* 1992;9:147-52.
 12. Kruskal WH, Wallis WA. Use of ranks in one-criterion analysis of variance. *J Am Stat Assoc* 1952;47:583-621.
 13. Kendall MG, Gibbons GD. Rank correlations methods. 5th ed. New York: Oxford University Press; 1990.
 14. Brahms H, Raker VA, van Venrooij WJ, Lührmann R. A major, novel systemic lupus erythematosus autoantibody class recognizes the E, F, and G Sm snRNP proteins as an E-F-G complex but not in their denatured states. *Arthritis Rheum* 1997;40:672-82.