
Precision and accuracy of commercial laboratories' ability to classify positive and/or negative allergen-specific IgE results

Sheryl L. Szeinbach, PhD*; James H. Barnes, PhD†; Timothy J. Sullivan, MD‡; and P. Brock Williams, PhD§

Background: Accurate and reliable evaluation of the presence or absence of allergen-specific IgE is important in the differential diagnosis of allergic disease. A variety of different commercial tests are available for this purpose. There are few data available to judge how the results of these different tests compare with one another in everyday use.

Objective: To examine prospectively the extent of comparability among specific IgE results from different laboratories.

Methods: Six diagnostic laboratories employing five different methods to assay specific IgE were selected. Aliquots from 26 serum samples that contained variable levels of IgE specific to 17 common aeroallergens were sent in triplicate to each study laboratory during a 6-week time period. Results were reported numerically and by class scores and then compared by examining their concordance using Kendall's W nonparametric statistical test. In addition, cut-off values were compared by a best agreement analysis using reported results. Reproducibility was determined using precision profiles based upon the coefficient of variation among triplicates for each allergen across the range of reported results.

Results: In all, 7,813 tests were analyzed. Concordance among different assays in commercial use with one exception was not good. This was particularly true around the cut-off region where most assays demonstrated high imprecision. The Pharmacia CAP System used by two different laboratories demonstrated highly comparable results with good precision. Some assays were reproducible but not accurate. Others were neither reproducible nor accurate.

Conclusions: The results of this study indicate that not all commercial laboratories/assays for specific IgE provide reproducible and accurate data. Significant potential for misdiagnosis was detected for some reported results. Methods were identified that do give sensitive, accurate, and reproducible results.

Ann Allergy Asthma Immunol 2001;86:373–381.

INTRODUCTION

Accurate and reliable evaluation of the presence or absence of allergen-specific IgE is an important element in the differential diagnosis of allergic disease. These measurements are used to

identify causative agents, determine the degree of sensitization, and when combined with the clinical history and physical examination, are used to make important treatment decisions.

A variety of different commercial tests for measuring specific IgE antibodies are available to physicians. Many physicians are under the impression that

serologic test results for specific IgE antibodies from different laboratories are reliable and interchangeable. Contributing to this belief is the use of a common classification system for reporting results. This classification system was originally introduced to make it easier for physicians to use these laboratory results but has obscured the fact that different procedures are used to obtain these seemingly identical scores. A number of studies have addressed this point and indicate that it is unlikely that these test results are equivalent.^{1–3}

In a recent study, Williams et al⁴ reported on the analysis of 12,708 specific IgE test results on the same serum samples sent through normal channels to six different major commercial laboratories using five different assay procedures. Their study found that specific IgE assays from different laboratories sometimes gave significantly different results. They suggested that although some procedures are capable of reporting accurate and reproducible results, others do poorly with respect to both quantification and reproducibility.

Commercial laboratory reports include results reported as log-related classes or actual mass units where increasing values represent a higher probability for IgE-mediated disease. Cut-off values for the different classes may be defined in terms of multiples of the level of background binding or by reference to serial dilutions of a positive control, but are highly dependent on the precision of the assay around the cut-off point.⁵ These cut-off values are determined differently for the various procedures but all basically relate to whether specific IgE antibodies can

* College of Pharmacy, Ohio State University, Columbus, OH.

† School of Business Administration, University of Mississippi, University, MS.

‡ Emory University School of Medicine, Atlanta, GA.

§ University of Missouri Medical School, Kansas City, MO and IBT Reference Laboratory, Lenexa, KS.

This study was supported by a grant from Pharmacia & Upjohn Diagnostics Division, Kalamazoo, Michigan.

Received for publication May 17, 2000.

Accepted for publication in revised form November 12, 2000.

or can not be determined with certainty.

In this study, we extend the analysis reported in our previous paper by using commercial laboratory recommended cut-off values to examine the concordance of reported positive and negative values on identical serum samples. We also compare the consequences of varying the cut-offs with respect to each assay method. In addition, we report each laboratory's precision around the cut-off point and for the entire range of analysis.

METHODS

Serum Samples

Twenty-six random anonymous serum samples from consented individuals containing different levels of specific IgE antibodies to a variety of different allergens were obtained from the Pharmacia Diagnostics AB serum bank (Uppsala, Sweden). All samples were aliquoted in serum vials containing 2.2 mL, given a coded identification number, randomized for shipment, and then stored at -70°C . The undiluted serum samples were assayed by each laboratory on three separate occasions.

Serum samples were submitted to each of six laboratories as routine clinical samples. These included shipping a weekly supply of samples to physicians to be picked up by the relevant laboratory on a biweekly basis. Five to seven samples were included in each

shipment. As a result of this process, study samples were sent to each laboratory over a span of 6 weeks.

Allergens

Tests for specific IgE antibodies were requested on each submitted sample for the following 17 allergens commonly evaluated during diagnosis of allergic problems: *Dermatophagoides farinae* (d2); cat dander (e1); dog dander (e5); Bermuda grass, *Cynodon dactylon* (g2); timothy grass, *Phleum pratense* (g6); white oak, *Quercus alba* (t7); short ragweed, *Ambrosia elatior* (w1); giant ragweed, *Ambrosia trifida* (w3); mugwort sage, *Artemisia vulgaris* (w6); Russian thistle, *Salsola kali* (w11); rough pigweed, *Amaranthus retroflexus* (w14); *Penicillium notatum* (m1); *Cladosporium herbarum* (m2); *Aspergillus fumigatus* (m3); *Mucor racemosus* (m4); *Candida albicans* (m5); and *Alternaria tenuis/alternata* (m6).

Assay Procedures

Assay procedures and laboratories in the present study included: Allergy Testing Laboratories (Gainesville, FL) using an in-house-developed modified RAST with license from Hycor Biomedical (Garden Grove, CA) to produce allergen paper discs in their laboratory (A-RAST); Allercare (Rome, GA) using Diagnostic Products Corporation-AlaSTAT (Los Angeles, CA; A-STAT) who reported results using both a standard and an extended classification system; Commonwealth Medical

Laboratory (Fairfax, VA) using a modified RAST manufactured by Hycor Biomedical (C-RAST); IBT Reference Laboratory (Lenexa, KS) using the Pharmacia CAP System-Standard scoring (Pharmacia Diagnostics AB, Uppsala, Sweden; I-CAP); Laboratory Corporation of America (Burlington, NC) using an in-house-developed modified RAST with a license from Hycor Biomedical to produce allergen paper discs in their laboratory (L-RAST); and SmithKline Beecham Clinical Laboratories (St. Louis, MO) using the Pharmacia CAP System-Standard scoring (S-CAP). The class score values for the laboratories considered in this study are shown in Table 1.

Data Analysis

Individual laboratory reports for each of the serum samples were collected and forwarded to an independent statistical team who broke the code and performed the statistical analysis. In this study, 7,813 useable assay results were available for analysis.

As an initial step in our analysis, we performed a graphical analysis of the different assay procedures by plotting the reported values of one assay against the reported values of another. The focus of this comparison was on values falling into classes 0 through II, representing 2,882 useable observations from the 7,813 reported results. The resulting scatter plots of values

Table 1. Assay Classifications and Scale Values

	I-CAP, S-CAP & A-STAT ¹	A-STAT ²	A-RAST & C-RAST	L-RAST
Class	kU _A /L	IU/mL	Count	Count
0	<0.35 ³	0.00–0.10	0–500	0–500
0/I		0.11–0.24	501–750	501–750
I	0.35–0.69	0.25–0.39	751–1600	751–1600
II	0.70–3.40	0.40–1.29	1601–3600	1601–3600
III	3.50–17.40	1.30–3.89	3601–8000	3601–8000
IV	17.50–49.9	3.90–14.99	8001–18000	8001–18000
V	50.0–99.9	15.00–24.99	18001–40000	>18000
VI	>100	>25.00	>40000	

¹ AlaSTAT standard classification system.

² AlaSTAT extended classification system.

³ I-CAP reported kU_A/L values down to 0.10 and classified values between 0.10 and 0.35 as Class 0/I.

were then divided into sectors based on the cut-off values used by the specific laboratories for the cut-off between Class 0/I and I (for this analysis, class 0 was included with class 0/I) and the cut-off between class I and class II. This resulted in creating nine separate sectors of each graph as depicted in Fig 1.

Sectors numbered 1, 5, and 9 were those where each assay agreed on the classification or, more specifically, where the observed values fell within the same class for the two assays examined in the graph. Sectors numbered 2, 4, 6, and 8 were those where there was a one-class difference between the assays, but this difference did not result in classifying a positive into a negative or vice versa (eg, a class 0, 0/I by one assay reported as a class I by the other). Sectors numbered 3 and 7 represented the most discordant results. In sector 3, the assay on the abscissa reported the class as II whereas the assay on the ordinate reported the same sample as a class 0 or 0/I. This difference was considered to represent a discordant negative result. Observations falling into sector 7, on the other hand, were those samples reported as class 0,0/I by the assay on the abscissa, and reported as class II by the assay on the ordinate. This difference was considered to represent a discordant positive result.

To examine statistically the magnitude of deviation across the different assay procedures on their classification of the different samples according to their degree of concordance, Kendall's coefficient of concordance (Kendall's W) was used as a second step in our analysis. Kendall's W⁶ is a nonparametric statistical test that is interpretable as a measure of agreement among raters or in the present context assay procedures. Each assay procedure was itself a reference and each serum sample was the object being evaluated. The test statistic associated with Kendall's W is the χ^2 .⁶

The third step in the analysis consisted of varying the cut-off values to examine the impact on classification of

assay values. Each assay procedure pair was examined three times. In the first case, the cut-off value between a positive and a negative was set for both assays at the lower cut-off value. In the second case, the cut-off value for the assay on the ordinate was raised to the higher cut-off value. In the third case, both assays were raised to the higher cut-off values. The purpose of varying cut-off values was to examine the analytical sensitivity of the different assay procedures based on the pattern of reported readings. This three-step procedure resulted in a four-quadrant plot with one positive and one negative cut-off for each assay. The number of concordant negative and positive results (lower left and upper right quadrant), the number of discordant negative results (lower right quadrant), and the number of discordant positive results (upper left quadrant) were then computed. Viewed differently, the result of this procedure was to treat the class I values as positive in the first case, positive for the abscissa assay procedure but negative for the ordinate assay procedure

in the second case, and negative for both assay procedures in the third case.

To examine the reproducibility of the reported results, precision profiles based on the coefficient of variation (CV) for all samples within a class were determined. The coefficient of variation was calculated by determining the standard deviation for each sample across all three replicates, dividing this value by the mean of the replicates, and multiplying by 100 to produce a percentage value. The higher the CV value, the greater the variation among the three assay results of the three identical serum samples. Because the CV calculation is computed from the three replicate samples, the measure is independent of the particular scaling system used by the laboratory.

RESULTS

Variability was seen across the reported results of different laboratory/assay procedures in reporting the analysis of identical samples. It was found that Kendall's W across all assay pro-

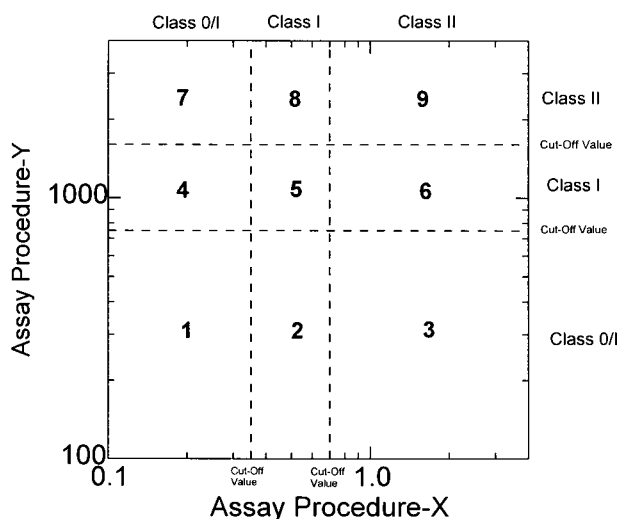


Figure 1. Allergen class cut-off sectors. Sectors numbered 1, 5, and 9 are those sectors where assay procedure-X and assay procedure-Y agree on the classification. Sectors numbered 2, 4, 6, and 8 are sectors where there was a one-class difference between the assays, but this difference did not result in classifying a positive into a negative or vice versa (eg, a class 0/I by one assay reported as a class I by the other). In sector 3, the assay on the abscissa reported the class as II whereas the assay on the ordinate reported the same sample as a class 0/I (discordant negative). Observations falling into sector 7, on the other hand, were those samples reported as class 0/I by the assay on the abscissa and reported as class II by the assay on the ordinate (discordant positive).

cedures was 0.120 ($\chi^2_5 = 49.41$, $P < 0.01$), demonstrating significantly different results from the different assays. Considering only the three laboratories that used a modified RAST procedure, Kendall's W was 0.037 ($\chi^2_2 = 96.50$, $P < 0.01$) indicating that these three laboratories were statistically different in their ability to classify the sample sera. For the two laboratories using the Pharmacia CAP system, Kendall's W was 0.001 ($\chi^2_1 = 1.82$, $P = 0.17$) indicating no significant difference in the classification of results by these two lab-

oratories. Comparison of the A-STAT reported results revealed that they differed statistically ($P \leq 0.01$) from the modified RAST procedure group (A-RAST, C-RAST, and L-RAST) and from the Pharmacia CAP (I-CAP and S-CAP) group.

The variability was particularly notable in the lower end of the assay scales as shown in Fig 2. A summary of these observations is shown in Table 2 where it can be noted that the agreement and discordant sectors varied greatly across the different compari-

sons. Table 2 also demonstrates that the results of A-RAST, A-STAT, C-RAST, and L-RAST compared with the I-CAP were biased toward discordant negatives (sector 3 > sector 7).

Varying the Class Cut-Off Value

We next examined whether varying the cut-off values of the different assay procedures could lead to more uniform results in these data. The results of this analysis are shown in Table 3 (most favorable comparisons for each assay pair are highlighted in Table 3).

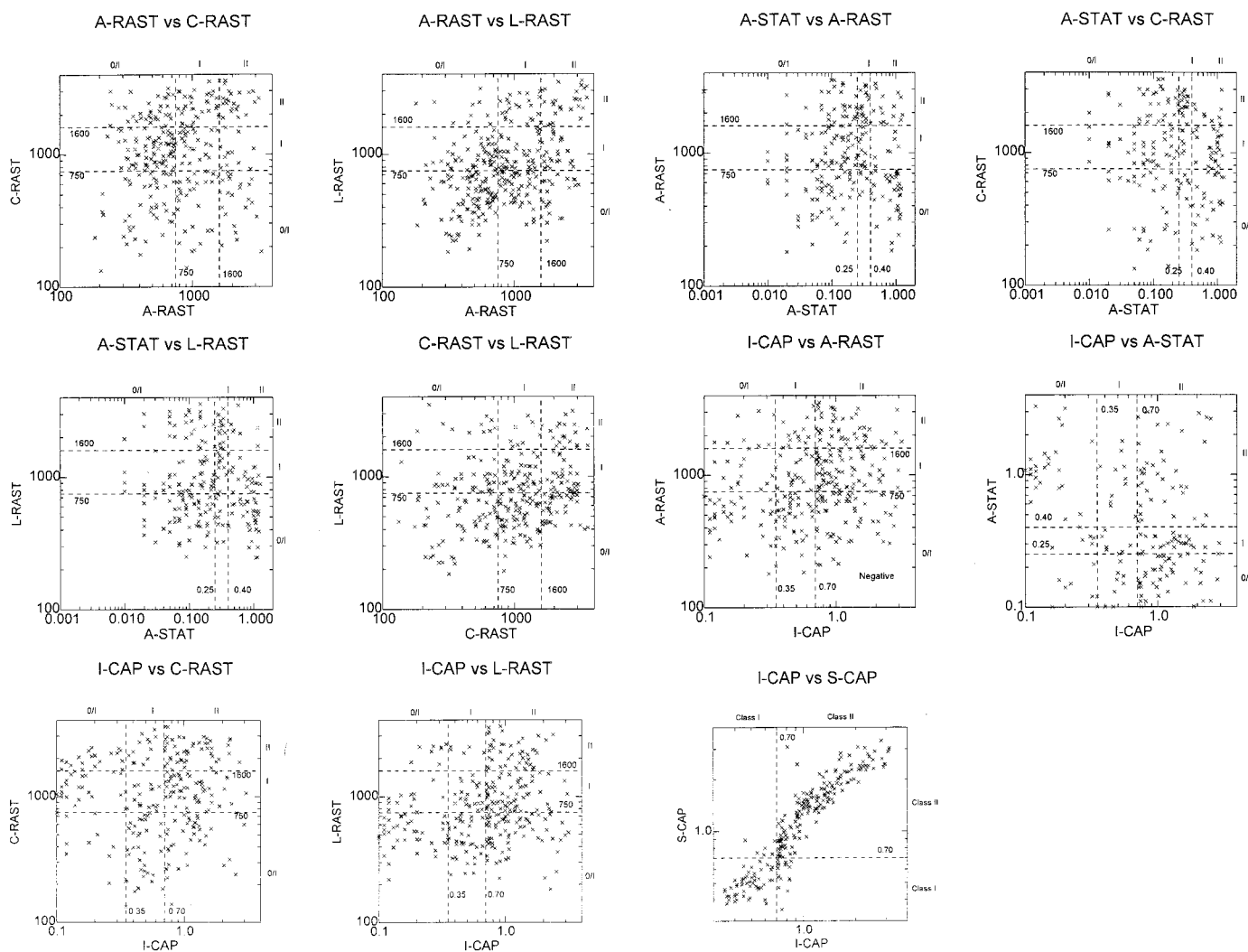


Figure 2. Comparison plots of assay procedures in classes 0-II. A scatter plot for each nonredundant pair of laboratories is shown. The pattern or lack thereof reflects the degree of concordance between the two plotted laboratory results. Because there was such strong agreement between I-CAP and S-CAP, it was determined that plotting S-CAP against each of the other assay procedures would provide little additional information. For A-STAT, we used their extended classification system ranges for plotting the cut-off boundaries since the emphasis of this paper is at the lower cut-off values. Also, we did develop plots using A-STAT's standard system and the extended system produced overall more concordant plots.

Table 2. Comparison of Assay Procedures in Classes 0-II

Assay Procedure ¹		Cut-off values between classes ²				Number of observations by sector number arranged in order of agreement (N = 2882 useable observations in classes 0-II)									
x-axis	y-axis	0/I-I	I-II	0/I-I	I-II	1	5	9	6	8	2	4	3	7	
A-RAST	C-RAST	750	1600	750	1600	47	25	23	11	32	31	74	23	33	
A-RAST	L-RAST	750	1600	750	1600	92	40	29	25	23	49	47	23	10	
A-STAT	A-RAST	0.25	0.40	750	1600	9	14	13	18	6	25	5	20	19	
A-STAT	C-RAST	0.25	0.40	750	1600	5	7	9	19	26	16	5	17	7	
A-STAT	L-RAST	0.25	0.40	750	1600	12	10	7	16	18	20	4	17	16	
C-RAST	L-RAST	750	1600	750	1600	60	40	21	29	8	62	35	35	11	
I-CAP	A-RAST	0.35	0.70	750	1600	36	19	47	67	12	37	13	58	8	
I-CAP	A-STAT	0.35	0.70	0.25	0.40	8	3	4	14	4	12	2	43	7	
I-CAP	C-RAST	0.35	0.70	750	1600	13	19	48	54	13	33	21	41	20	
I-CAP	L-RAST	0.35	0.70	750	1600	40	24	42	62	5	39	11	67	8	
I-CAP	S-CAP	0.35	0.70	0.35	0.70	74	47	158	17	5	10	4	1	6	

¹ Only nonredundant comparisons are shown. That is, placing I-CAP on the y-axis to compare against A-RAST would essentially provide the same result as I-CAP on the x-axis.

Since there was such strong agreement between I-CAP and S-CAP, it was determined that plotting S-CAP against each of the other assay procedures would provide little additional information.

² Class 0/I includes class 0 reported values.

When comparing one modified RAST procedure to another, concordant results were more prevalent when the cut-off for both assay pairs was set at the higher (1600 counts, class II) level. When I-CAP was compared against any of the modified RAST procedures, the optimal approach was to use the recommended cut-off (treat class I as a positive reading) value for I-CAP or S-CAP. Comparing A-STAT to the three modified RAST procedures produced mixed results. In two cases (A-RAST and L-RAST), the better cut-off value for both assays was when both assays were at the higher level. However, for the C-RAST comparison, the mixed cut-off values outperformed the others. This was surprising because the mixed cut-off values (one assay treating class I as positive and the other treating it as negative) would be expected to produce the worst results.

For the comparison of the two laboratories using the Pharmacia CAP system, the recommended cut-off produced the optimal solution. Because results of these two laboratories agreed so closely, even raising the cut-off value for both to the higher level did not produce detrimental results. As expected, a mixed cut-off value resulted in degradation in the number of agreements.

Possible Cause of the Discordant Results: Laboratory/Assay Reproducibility

Because this was a blinded study, it was difficult to determine the precise cause of the discordant results. However, one likely cause could be imprecision of the different assays, particularly in the lower range. Therefore, we constructed precision profiles across the range of reported values for each laboratory/assay.

These precision profiles consisted of a series of box and whisker plots which provide a convenient way to visually inspect the characteristics of the entire distributions of the CVs (Fig 3). The charts enable simultaneous reporting of central tendency (the median), sample variability (interquartile range), and the detection of outliers (unusually small or large values for individual cases in the sample).

As one can see from Fig 3, for all the laboratories in the current study, the range for the CV varied across the different reported classes. Although no single pattern exists across all the laboratories, it does appear that higher CVs generally occur in the lower classes and become smaller as the class score increases. The one major excep-

tion is A-STAT where the CV first decreased and then increased in the upper class scores. A-STAT had an average coefficient of variation of 66% across all classes with the top 50% of their CV values ranging from a low of 4% to a high of 140% (box edges). Of the three modified RAST systems, C-RAST had an average CV of 15% with values ranging from a low of 1% to a high of 30% across all classes. C-RAST was least precise in classes 0, I, and II. A-RAST was least precise in class 0/I where it had the highest variability of the modified RAST procedures. Overall, A-RAST had an average CV of 26%, ranging from a low of 5% to a high of 50% across all classes. L-RAST had an average CV of 16% and a range of 2% to 27%. I-CAP had an average CV of 14% with a range of 1% to 25%. S-CAP had an average CV of 11% with a range of 6% to 20%. I-CAP's least precise area was in classes 0/I and V, which were approximately the same as S-CAP's most precise classes. S-CAP was least precise in class II.

I-CAP reported 1,002 positive (defined as class I and higher) and 324 negative results (defined as classes 0 and 0/I) with an overall CV of 10% (ranging from 7% to 13% for the 1,002

Table 3. Effect of Changing Class Cut-Off Values

Assay Procedure ¹		Cut-off value between positive and negative		Number of observations and percentage of total (N = 2882)					
x-axis	y-axis	x-axis	y-axis	Agree		Discordant Positive		Discordant Negative	
A-RAST	C-RAST	750	750	138	46%	107	36%	54	18%
A-RAST	C-RAST	750	1600	176	59%	33	11%	90	30%
A-RAST	C-RAST	1600	1600	200	67%	65	22%	34	11%
A-RAST	L-RAST	750	750	209	54%	72	19%	107	28%
A-RAST	L-RAST	750	1600	191	57%	137	41%	10	3%
A-RAST	L-RAST	1600	1600	257	76%	48	14%	33	10%
A-STAT	A-RAST	0.25	750	101	47%	77	36%	37	17%
A-STAT	A-RAST	0.25	1600	130	60%	26	12%	59	27%
A-STAT	A-RAST	0.40	1600	131	61%	45	21%	39	18%
A-STAT	C-RAST	0.25	750	90	48%	66	35%	30	16%
A-STAT	C-RAST	0.25	1600	95	51%	29	16%	62	33%
A-STAT	C-RAST	0.40	1600	92	49%	46	25%	48	26%
A-STAT	L-RAST	0.25	750	112	53%	64	30%	36	17%
A-STAT	L-RAST	0.25	1600	115	54%	27	13%	70	33%
A-STAT	L-RAST	0.40	1600	122	58%	44	21%	46	22%
C-RAST	L-RAST	750	750	158	52%	46	15%	97	32%
C-RAST	L-RAST	750	1600	124	41%	11	4%	166	55%
C-RAST	L-RAST	1600	1600	218	72%	19	6%	64	21%
I-CAP	A-RAST	0.35	750	181	61%	21	7%	95	32%
I-CAP	A-RAST	0.35	1600	108	36%	8	3%	181	61%
I-CAP	A-RAST	0.70	1600	152	51%	20	7%	125	42%
I-CAP	A-STAT	0.35	0.25	78	51%	23	15%	51	34%
I-CAP	A-STAT	0.35	0.40	48	32%	17	11%	87	57%
I-CAP	A-STAT	0.70	0.40	55	36%	25	16%	72	47%
I-CAP	C-RAST	0.35	750	147	56%	41	16%	74	28%
I-CAP	C-RAST	0.35	1600	95	36%	20	8%	147	56%
I-CAP	C-RAST	0.70	1600	134	51%	33	13%	95	36%
I-CAP	L-RAST	0.35	750	173	58%	19	6%	106	36%
I-CAP	L-RAST	0.35	1600	98	33%	8	3%	192	64%
I-CAP	L-RAST	0.70	1600	156	52%	13	4%	129	43%
I-CAP	S-CAP	0.35	0.35	301	93%	10	3%	11	3%
I-CAP	S-CAP	0.35	0.70	241	75%	6	2%	75	23%
I-CAP	S-CAP	0.70	0.70	293	91%	11	3%	18	6%

Shaded values represent the best cut-off combination in terms of fewest discordant results.

positive results). S-CAP, also using the Pharmacia CAP System, reported 994 positive and 332 negative results with an overall CV of 11% (ranging from 7% to 14% for the 994 positive results). This is an agreement of 98% between the two laboratories performing blinded samples analyzed according to their routine procedure for clinical samples. The other laboratories had the following CVs for their positive results: A-RAST, 20% (range 16% to 26%); A-STAT, 28% (range 17% to 49%); C-RAST, 13% (range 5% to 28%); and L-RAST, 15% (range 9% to 31%). Further, we found that not only did the CV

vary across classes but also varied across allergens. This variability across allergens can be seen in Table 4.

The molds and weeds showed the highest CVs. The CVs were also high for mite with A-STAT, A-RAST, and C-RAST.

DISCUSSION

Ideally, the results from all laboratories, regardless of procedure used, should be accurate, reliable, and interchangeable. Unfortunately, results observed in the present study revealed that this ideal state does not presently exist for some currently available com-

mercial assays. Discordant positive and negative results were seen with all allergens. Although the majority of these occurred with mold and weed allergens, they were also found throughout the reported test results, even for common allergens such as cat, dog, and mite. Further, a substantial portion of this variation remained even when the low-level results were eliminated from consideration. This indicated that there were some major differences in the results reported by some of these laboratories.

The practical value of any laboratory test resides in its ability to give accurate, consistent, and thus reliable

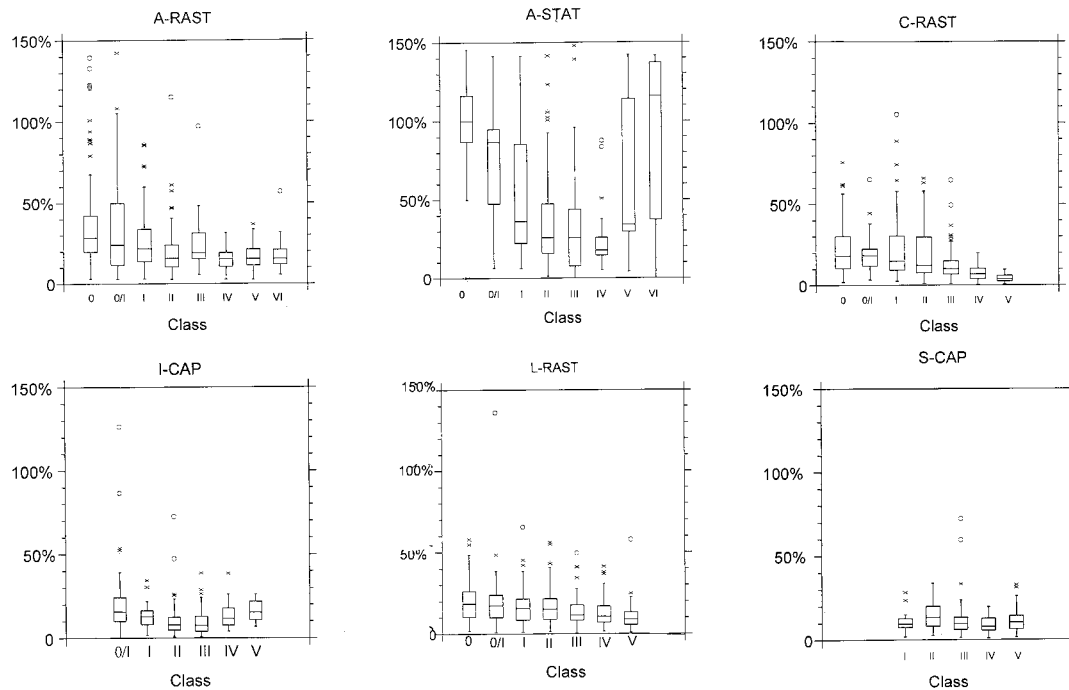


Figure 3. Precision profiles of assay procedures. The length of each box shows the range within which the central 50% of the CV value fell, with the box edges at the first and third quartiles. The whiskers show the range of CV values that fell within 1.5 times the second quartile range on the lower whisker and 1.5 times the third quartile range on the upper whisker. Values between 1.5 and 3 times the respective interquartile range were plotted with asterisks. Values beyond 3 times the respective interquartile range, called far outside values, were plotted with unfilled circles. Not all laboratories reported values in every class. I-CAP reported values in class 0 as <0.10 and in class VI as >100. S-CAP reported both class 0 and 0/I as <0.35 and in class VI as >100. C-RAST reported no readings in class VI. As a result of these differences in reporting, it was not possible to compute the CV for those affected classes

Table 4. Lab Precision Summary for Positive Readings

Allergen	A-RAST	A-STAT	C-RAST	I-CAP	L-RAST	S-CAP
d2 <i>Dermatophagoides farinae</i>	17.7%	17.7%	23.7%	8.8%	9.9%	11.2%
e1 Cat dander	16.8%	18.8%	5.3%	11.3%	9.7%	14.1%
e5 Dog dander	17.6%	19.7%	11.0%	10.3%	15.4%	7.4%
g2 Bermuda grass (<i>Cynodon dactylon</i>)	19.7%	28.7%	8.2%	11.6%	14.2%	11.1%
g6 Timothy grass (<i>Phleum pratense</i>)	16.9%	22.8%	5.4%	9.4%	14.0%	13.9%
t7 White oak (<i>Quercus alba</i>)	17.4%	16.5%	10.5%	7.0%	12.9%	11.6%
w1 Short ragweed (<i>Ambrosia elatior</i>)	18.2%	24.0%	7.5%	8.7%	14.1%	6.5%
w3 Giant ragweed (<i>Ambrosia trifida</i>)	24.6%	28.2%	10.0%	11.8%	9.9%	12.0%
w6 Mugwort sage (<i>Artemisia vulgaris</i>)	22.0%	49.1%	7.8%	12.6%	12.9%	10.5%
w11 Russian thistle (<i>Salsola kali</i>)	16.3%	27.0%	16.0%	8.1%	14.8%	8.5%
w14 Rough pigweed (<i>Amaranthus retroflexus</i>)	16.3%	17.8%	23.5%	8.1%	15.0%	13.1%
m1 <i>Penicillium notatum</i>	20.8%	41.4%	15.1%	8.5%	17.2%	11.0%
m2 <i>Cladosporium herbarum</i>	26.1%	32.1%	15.3%	10.4%	11.9%	10.3%
m3 <i>Aspergillus fumigatus</i>	20.7%	26.8%	14.9%	8.2%	11.5%	10.7%
m4 <i>Mucor racemosus</i>	21.7%	27.2%	28.4%	11.2%	31.3%	9.4%
m5 <i>Candida albicans</i>	20.4%	25.7%	15.7%	11.2%	17.0%	12.5%
m6 <i>Alternaria tenuis/alternata</i>	23.4%	18.3%	9.6%	10.8%	8.5%	10.6%
Average for Lab	20.3%	27.5%	13.4%	9.8%	14.7%	10.8%

Values less than 15% are shaded as recommended acceptable by NCCLS (NCCLS 1997).

results.^{4,5} Clinically, the misclassification of patients by serologic tests for

specific IgE antibodies could result in the implementation of unnecessary

treatments or the omission of effective clinical interventions. Inconsistent re-

sults in some assay systems could lead to lack of confidence in all in vitro assays for specific IgE. Scientific studies involving in vitro measurement for specific IgE could lead to inaccurate conclusions if faulty systems were used.

On the positive side, this study did find that two laboratories (I-CAP, S-CAP) with blinded samples sent through normal channels produced precise and consistent results. These findings are consistent with other research that has found the Pharmacia CAP System to be reproducible, quantitative, and accurate.^{1,2,4,7-14}

Three varieties of the modified RAST test were used in this study. Although the modified RAST test has been suggested to have increased sensitivity when compared with skin testing, several studies indicate that this method is not entirely quantitative and may provide discordant positive results.^{1,3,15} Our study indicated that the three laboratories using modified RAST were statistically different from one another ($P < 0.01$) in their ability to classify the sample sera. One laboratory (C-RAST, using modified RAST provided by Hycor Biomedical) did demonstrate a significant proportion of discordant positive results. A similar finding was noted recently,¹⁶ when using Hycor's test for latex. The other two laboratories that used the modified RAST procedure exhibited more discordant negative results. Because all three laboratories used the same cut-off value for determining positive and negative results and presumably similar methods, the discrepancies probably reflect differences in allergen source and coupling efficiency. The results from C-RAST, which had the opposite problem of too many discordant positives, may possibly be attributed to problems with non-specific binding of label resulting from reagents or procedures. If these disagreements were largely because of low-end variability, raising the cut-off should have made a larger difference. Because these differences were not observed, there is no evidence supporting the suggestion that the modified RAST

procedure possesses increased sensitivity over other assay formats. In fact, evidence in Table 4 would suggest that all the modified RAST formats were less sensitive than the Pharmacia CAP System.

It is difficult to understand the poor performance of the AlaSTAT system used by Allercare, where a larger number of discordant negative results were observed with the AlaSTAT system (FDA-approved) than with two in-house manufactured tests. These results imply that either the test itself did not work well in this laboratory or that the calibration of the assay was less than optimal. Variability of results for this procedure has been reported in other studies.^{2,17}

Reported CVs for most of these assays were much lower than we observed. These are usually determined in one laboratory by repeated assay runs. Our data were derived from what physicians receive as everyday results. This suggests that the reported precision of most of these assays is overestimated. One could also argue that the reported CVs of negative or very low values are not useful because a small difference between replicates may result in a proportionally larger CV in this range. This is why we presented the precision profiles across the entire range of reported values for each assay. The higher precision reported in the upper classes is likely to be related to the fact that classes are log-related. Further, because the modified RAST test is likely to be saturated in this area, less variability would be expected.³

The strengths of this study included the large number of tests compared, the fact that they were sent through normal, everyday channels, and the fact that these samples were all run and analyzed in a blinded fashion. Thus, any observed deviations resulted from the laboratories themselves and/or the procedure they have selected to use. Weaknesses arise from the fact that with the exception of the cut-off data, the results must be compared on a class score basis. This is because the classes used by different laboratories are not likely to represent equal amounts of

specific IgE as they are not based upon a quantitative unit assignment.

Because the samples for this study were aliquoted, frozen, and sent to physicians who then sent them to the respective laboratories, one could argue that the variability seen in this study was a result of the procedures used to blind this study. However, this is unlikely for three reasons. First, serum IgE antibodies have been shown to be stable for determination by assays such as those used in the present study.¹⁸ Second, the fact that I-CAP and S-CAP results were so close in both precision and accuracy suggests that the sample distribution procedure used in this study did not bias the results. Third, the fact that the different laboratories were consistent in their discordant results and that they were different for different allergens would strongly suggest that factors other than the blinding procedure were the basis for the discordant results.

CONCLUSION

Although tests for specific IgE have progressed, some laboratories continue to use methods that lack precision and reliability. This is unfortunate because laboratory services may be contracted under a managed care plan, thus leaving few alternatives for both patients and physicians. Misclassification of a patient's allergy status could lead to significant unnecessary morbidity and expense. A study of the costs and clinical consequences of misclassification of patients from the perspective of managed care organizations and patients will be reported separately.

Our data combined with those of Williams et al⁴ indicate that some commercial assays for specific IgE are precise and accurate, whereas others are precise but not accurate, and still others are neither precise nor accurate. These conclusions are also allergen-dependent.

Well-performing serologic tests for specific IgE antibodies are available to physicians. These tests exhibit outstanding performance characteristics, are robust, and can be used with confidence in identifying causative agents

in individuals with allergic diatheses. The clinical application of accurate and reliable IgE antibody test results interpreted within the context of the clinical history and physical examination can add significant objectivity to the evaluation of allergic patients.

REFERENCES

1. Kelso JM, Sodhi N, Gosselin VA, Yunginger JW. Diagnostic performance characteristics of the standard Phadebas RAST, modified RAST, and Pharmacia CAP system versus skin testing. *Ann Allergy* 1991;67:511-514.
2. Costongs GM, Janson PC, Hermans WJ, et al. Evaluation of performance characteristics of automated measurement systems for allergy testing. *Eur J Clin Chem Clin Biochem* 1995;33:295-305.
3. Williams PB, Dolen WK, Koepke JW, Selner JC. Immunoassay of specific IgE: use of a single point calibration curve in the modified radioallergosorbent test. *Ann Allergy* 1992;69:48-52.
4. Williams PB, Barnes JH, Szeinbach SL, Sullivan TJ. Analytic precision and accuracy of commercial immunoassays for specific IgE: establishing a standard. *J Allergy Clin Immunol* 2000;105:1221-1230.
5. National Committee for Clinical Laboratory Standards (NCCLS). Evaluation methods and analytical performance characteristics of immunological assays for human immunoglobulin (IgE) antibodies of defined specificities: Approved Guideline. Vol 16, No. 5; 1996. NCCLS document 1/LA-20-P.
6. Siegel S, ed. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill, 1956.
7. Yman L. Die neue Generation der Allergie-Testung: Pharmacia CAP System. In-vitro Diagnostica Special. *Das wissenschaftliche Magazin zum Thema* 1990;1:18-22.
8. Paganelli R, Ansotegui IJ, Sastre J, et al. Specific IgE antibodies in the diagnosis of atopic disease. Clinical evaluation of a new in vitro test system, UniCAP, in six European allergy clinics. *Allergy* 1998;53:763-768.
9. Liappis N, Starke A. Evaluation des Immuno-CAP. System in der In-vitro-Allergiediagnostik. *Allergologie* 1997;20:291-295.
10. Williams PB, Dolen WK, Koepke JW, Selner JC. Comparison of skin testing and three in vitro assays for specific IgE in the clinical evaluation of immediate hypersensitivity. *Ann Allergy* 1992;68:35-45.
11. Pastorello EA, Incorvaia C, Pravettoni V, et al. A multicentric study on sensitivity and specificity of a new in vitro test for measurement of IgE antibodies. *Ann Allergy* 1991;67:365-370.
12. Pastorello EA, Incorvaia C, Ortolani C, et al. Studies on the relationship between the level of specific IgE antibodies and the clinical expression of allergy. I. Definition of levels distinguishing patients with symptomatic from patients with asymptomatic allergy to common aeroallergens. *J Allergy Clin Immunol* 1995;96:580-587.
13. Liappis N, Lantto O, Nilsson G, Rydén B. The role of inter-laboratory comparisons in quality improvement of specific IgE measurements with the Pharmacia CAP System RAST RIA and FEIA. *Clin Lab* 1996;42:707-713.
14. Leimgruber A, Mosimann B, Claeys M, et al. Clinical evaluation of a new in-vitro assay for specific IgE, the immuno CAP System. *Clin Exp Allergy* 1991;21:127-131.
15. Dolen WK, Williams PB, Koepke JW, Selner JC. Immunoassay of specific IgE: low level assays require measurement of allergen specific assay background. *Ann Allergy* 1992;69:151-156.
16. Hamilton R, Biagini R, Krieg E. Multi-Center Latex Skin Testing Study Task Force. Diagnostic performance of Food and Drug Administration-cleared serologic assays for natural rubber latex-specific IgE antibody. *J Allergy Clin Immunol* 1999;103:925-930.
17. Lundberg M, Wrangsjo K, van Hage-Hamsten M. Diagnosis of latex allergy. *Allergy* 1997;52:1042-1043.
18. Johansson SG, Yman L. In vitro assays for immunoglobulin E. Methodology, indications, and interpretation. *Clin Rev Allergy* 1988;6:93-139.

Requests for reprints should be addressed to:
P. Brock Williams, PhD
11840 Woodland Road
Olathe, KS 66061
E-mail:pbrockwil@aol.com