

Receiver-operating characteristic (ROC) curves: a fundamental tool for improving the clinical usefulness of *in vitro* IgE tests

Plebani M, Borghesan F, Basso D, Faggian D. Receiver-operating characteristic (ROC) curves: a fundamental tool for improving the clinical usefulness of *in vitro* IgE tests.

Allergy 1996; 51: 407–411. © Munksgaard 1996.

In order to establish the most efficient thresholds for serum allergen-specific IgE measured by the Pharmacia CAP System, a “second-generation” *in vitro* method, we evaluated results from 89 subjects with suspected inhalant allergies, using receiver-operating characteristic (ROC) curve analysis. Sera samples were analyzed by the Pharmacia CAP System for specific IgE, the choice of allergens to be tested being based upon the symptoms and clinical history of each patient. Results were analyzed by ROC analysis for the five most representative allergens, cat dander (e_1), *Dermatophagoides pteronyssinus* (d_1), *Lolium perenne* (g_5), wall pellitory (w_{10}), and wormwood (w_5). The areas under the ROC curves were found to be satisfactory, ranging from 0.931 (e_1) to 0.974 (g_5) when we excluded w_5 , which had a significantly smaller area (0.81). To establish the most efficient threshold for each allergen, we calculated the clinical sensitivity, specificity, efficiency, and negative and positive predictive values. The thresholds giving a higher diagnostic efficiency were as follows: 0.40 kUa/l for d_1 , 0.55 kUa/l for g_5 , 0.50 for e_1 , 0.65 kUa/l for w_{10} , and 1.00 kUa/l for w_5 . It is concluded that quantitative reporting of specific IgE measurement has numerous advantages, but the choice of the positive threshold seems to be a prerequisite for obtaining the optimal clinical efficiency. It is also suggested that a specific threshold should be adopted for each allergen.

In vitro tests for IgE are technically valid for the measurement of allergen-specific IgE antibody in the serum of patients with inhalant allergy. In particular, it has been demonstrated that “second-generation” techniques provide reliable results for IgE measurement when they are performed well and interpreted appropriately in the light of the patient’s clinical history and physical examination (1–3).

The first available, and probably the most widely used “second-generation” technique for serum IgE measurement, the so-called Pharmacia CAP System, introduced important analytic improvements, including the capacity for quantitative reporting, the standard curve of the system being calibrated against the World Health Organization standard 75/502 for total IgE antibody. Quantitative reporting has numerous advantages but lacks a correspondence between arbitrary class scores and

IgE antibody units. Therefore, the positive threshold for the CAP System is usually chosen arbitrarily, according to the manufacturer’s recommendations, without the opportunity to compare the sensitivities and specificities of the different cutoffs.

The aim of this study was to establish the positive threshold for the major aeroallergens measured by the immunoCAP System using receiver-operating characteristic (ROC) curve analysis (4, 5), which gives exhaustive information on the accuracy of a test, including all its possible decision levels. This information may be visualized in a graph (ROC plot) where sensitivity (true-positive ratio) is plotted against 1-specificity (false-negative ratio) for various cutoff levels of the test.

Material and methods

An analysis was made of 89 serum samples from patients – who had given informed consent – with

**M. Plebani^{1,2}, F. Borghesan¹,
D. Basso¹, D. Faggian¹**

¹Department of Laboratory Medicine, Laboratorio Centrale, University Hospital, Padua; ²Centre for Biomedical Research, Castelfranco Veneto, Italy

Key words: allergen-specific IgE, inhalant allergies; receiver-operating characteristic (ROC) curves; skin prick test.

Dr Mario Plebani
Laboratorio Centrale
Azienda Ospedale di Padova
Via Giustiniani, 2
35128 Padova
Italy

Accepted for publication 4 March 1996

suspected inhalant allergies ($n=89$; 49 females, 40 males; mean age = 28.5 years, range 10–61 years). All patients had respiratory symptoms, with rhinitis and/or asthma, and had been referred to our allergy outpatient clinic by their primary care physician. A careful clinical history was taken. Inclusion criteria were as follows: age between 10 and 61 years, no medication which could interfere with the results of skin testing and no immunotherapy in the past, and no signs of autoimmune or other immunologic disorders or neoplasias. The choice of allergens to be tested both *in vivo* and *in vitro* was based upon the symptoms and clinical history of each patient. The final diagnosis was based upon full concordance between case history, clinical data, and *in vivo* test results. Data from some patients were discordant, and were therefore not included in the ROC analysis, although the specimens were analyzed by ImmunoCAP and by other sophisticated *in vitro* techniques (immunoblot and CAP inhibition). In 20 healthy subjects (10 males and 10 females; mean age 27 years, range 16–58) with negative *in vivo* tests, specific IgE against all major allergens were assayed by ImmunoCAP. Five allergens were considered, as they were numerous enough for a reliable statistical analysis (Table 1).

In vivo tests

Skin testing was performed by the prick method and following the recommendations of the Italian Society for Allergy and Clinical Immunology (6). Allergen extracts in 50% glycerin of mites (5000 PNU/ml), pollens (10 000 UP/ml), and *Alternaria* (10 000 PNU/ml) from Bayropharm (Milan, Italy) were used. These extracts have long been used in our clinical practice. The same lots were used throughout the study.

The skin was pricked with prick lancets (Hollister Stier, Milan, Italy). Positive and negative controls were placed, histamine phosphate (5 mg/ml) and

50% glycerin alone being used, respectively. Test reactions were read after 30 min, and wheal diameters, measured at right angles, were recorded. A wheal and flare reaction corresponding to a quarter of the diameter produced by histamine was graded as one plus (+), half the diameter as two (++), the same diameter as three (+++), and twice the diameter as four (++++).

In vitro tests (Pharmacia CAP System)

Undiluted sera were assayed separately, according to a previously described method (7, 8), the manufacturer's instructions being closely followed. The fluoroimmunoassay version was used. Test sera (50 μ l) and standards were incubated at room temperature with the ImmunoCAP for 30 min before being washed. For the measurement of allergen-bound IgE antibody, we added specific combinations of polyclonal and monoclonal anti-IgE (Fc specific) antibodies labeled with β -galactosidase. After a second incubation of 150 min, the ImmunoCAP was washed and the substrate, 4-methylumbelliferyl- β -D-galactoside, added. After 10 min, the reaction was stopped, and the fluorescence was detected in a fluorometer (FluoroCount Pharmacia Diagnostics, Uppsala, Sweden).

All steps (except for the sera and solid-phase distribution) and the final reading were carried out by an automated instrument (AutoCAP, Pharmacia Diagnostics, Uppsala, Sweden). Results were expressed quantitatively in kilo-units per liter (kUa/l) using the suggested cutoff of 0.35 kUa/l. The reproducibility of the method was tested both within assay (20 replicates) and between assay (10 replicates in 10 different analytic sessions) experiments. Within-assay coefficients of variation (CV%) ranged from 1.19 (e_1) to 5.24 (d_1), respectively. Between assays, CVs ranged from 4.45 (e_1) to 7.74 (g_5), respectively.

Table 1 Number of positives and negatives for each allergen

Allergen	Positives (n)	Negatives (n)	Patients not considered*
Cat dander (e_1)	16 (11 M, 5 F)	65 (25 M, 40 F)	8 (4 F, 4 M)
<i>Dermatophagoides pteromyssinus</i> (d-1)	30 (16 M, 14 F)	53 (20 M, 33 F)	6 (4 M, 2 F)
<i>Lolium perenne</i> (g_5)	38 (23 M, 15 F)	46 (13 M, 33 F)	5 (4 M, 1 F)
Wall pellettizer (w_{10})	26 (13 M, 13 F)	57 (22 M, 35 F)	6 (5 M, 1 F)
Wormwood (w_5)	15 (6 M, 9 F)	68 (30 M, 38 F)	6 (4 M, 2 F)

M = males; F = females

* Patients not considered for ROC analysis because of discordance between clinical data and *in vivo* results.

ROC curves and predictive values

Sensitivity and specificity were defined according to Galen & Gambino (9). A false-positive ratio corresponds to 1-specificity. With the ROC curve, by plotting the false-positive ratio (i.e., 1-specificity) against the true-positive fraction (i.e., sensitivity) and varying the threshold of specific IgE levels, we define the positive threshold that offers the best possible compromise between the lowest false-positive fraction and the highest true-positive fraction. The ideal point, obtained with an ideal test, is the value of specific IgE giving a sensitivity equal to 1 and a specificity equal to 1, corresponding to 1-specificity equal to zero. For each allergen, we also considered the areas under the ROC curve. By using the positive threshold defined by ROC analysis, we calculated sensitivity, specificity, efficiency, predictive positive, and negative values with Test Evaluation (10) and GraphROC (11) software.

Results

Patients were classified as positive or negative for the single allergen on the basis of the correspondence between case history, clinical data and *in vivo* results. In healthy subjects, the ImmunoCAP always gave results below the detection limit. These data were not considered in constructing the ROC curves. Table 1 shows the number of results considered positive or negative for each allergen used for constructing ROC plots. In six out of eight patients with discordant data, the results of CAP inhibition and immunoblot confirmed the clinical suspicion in agreement with CAP results.

Figs. 1–5 illustrate the ROC curves for d_1 , g_5 , e_1 , w_{19} , and w_5 , respectively. The corresponding areas under the curves were 0.958 for d_1 , 0.974 for g_5 , 0.931 for e_1 , 0.960 for w_{19} , and 0.812 for w_5 , respectively.

For d_1 , with 0.40 kUa/l as the threshold value, the sensitivity was 0.96 and the specificity 0.94, corresponding to a clinical efficiency of 0.95. When we chose a higher threshold (1.20 kUa/l) in the search for a higher specificity (1.00), the sensitivity decreased to 0.80 and the clinical efficiency to 0.92.

For g_5 , the same clinical efficiency was found with threshold values ranging from 0.36 and 0.55 kUa/l. At the lowest value (0.36 kUa/l), the clinical sensitivity was 1.00 and the specificity 0.97, while at the highest value (0.55 kUa/l), specificity increased to 1.00 and sensitivity was 0.97.

For e_1 , the threshold with the highest efficiency (0.98) was at 0.50 kUa/l, sensitivity being 1.00 and specificity 0.98.

For w_{19} , optimal clinical efficiency (0.96) was found at a threshold of 0.65 kU/l, sensitivity being

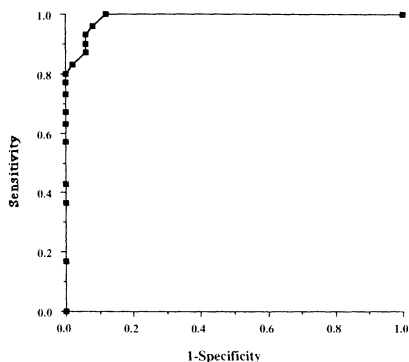


Fig. 1. ROC curve obtained with Pharmacia CAP System for *Dermatophagoides pteromyssinus* allergen (d_1). Sensitivity = true-positive fraction; 1-specificity = false-positive fraction.

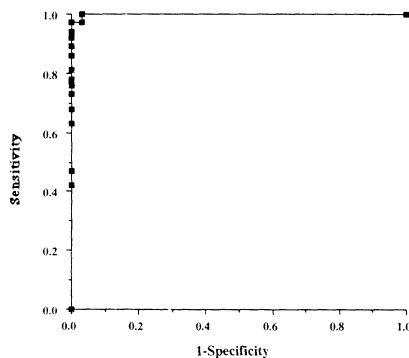


Fig. 2. ROC curve obtained with Pharmacia CAP System for *Lolium perenne* (g_5) allergen.

0.92 and specificity 0.98. At lower thresholds (0.35–0.37 kUa/l), the sensitivity was 1.00 although specificity was unsatisfactory (0.877). On the other hand, at higher thresholds (from 1.40), the specificity rose to 1.00, resulting in a low sensitivity (0.84).

Finally, w_5 was found to be the least efficient allergen. Optimal efficiency (0.85) was found at a threshold of 1.00 kUa/l, sensitivity being 0.60 and the corresponding specificity 0.91. Nor at lower thresholds did we find a sensitivity higher than 0.80; in the same way, a specificity of 1.00 was reached at a very high value (7.0 kUa/l), the corresponding sensitivity (0.20) being unacceptable. Table 2 illustrates these results.

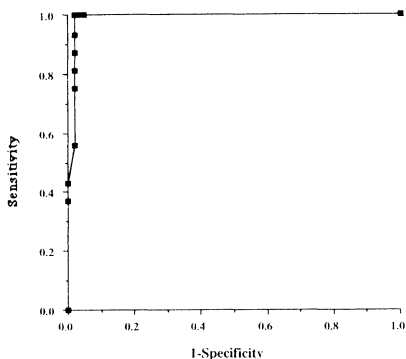


Fig. 3. ROC curve obtained with Pharmacia CAP System for cat dander (c_1) allergen.

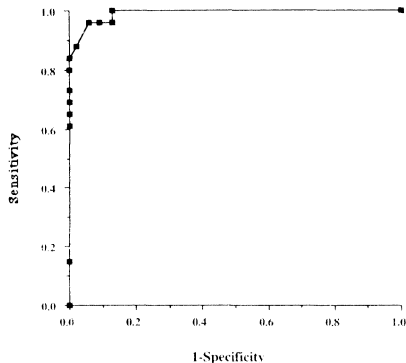


Fig. 4. ROC curve obtained with Pharmacia CAP System for wall pellicular (w_{10}) allergen.

Discussion

According to the Executive Committee of the American Academy of Allergy and Immunology,

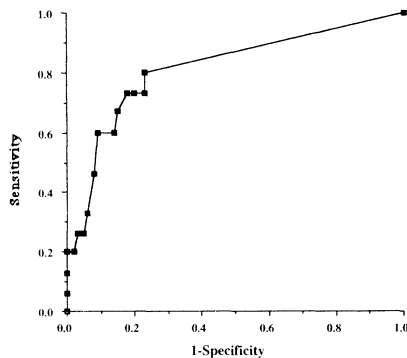


Fig. 5. ROC curve obtained with Pharmacia CAP System for wormwood (w_2) allergen.

"quantitative reporting offers numerous benefits, including facilitation of the usual statistical criteria for threshold selection and determination of precision profiles" (12). Previous papers described the analytic and clinical efficiency of the Pharmacia CAP System for IgE measurement (2, 3, 7), including the measurement of the imprecision profiles (13). The aim of this paper was to establish the most efficient possible threshold values for five major aeroallergens by ROC curve analysis. To our knowledge, this method has been used for the evaluation of *in vitro* IgE tests only by De Blay et al., who analyzed two different techniques (Pharmacia Phazedym and Abbott Matrix) (14), and by Williams et al. (15), who used ROC analysis to evaluate the efficiency of *in vitro* and *in vivo* methods.

To evaluate the specificity of each allergen, we used a reference population made up of symptomatic subjects for allergens different from those considered. This choice was made in order to evaluate any possible cross-reactivity between allergens. In a group of 20 healthy controls, no positive results for specific IgE were found. They were therefore not considered for ROC analysis since specificity is

Table 2. Clinical sensitivity, specificity, and efficiency at the thresholds selected

Allergen	Threshold (kU/L $\cdot$ l)	Sensitivity	Specificity	PPV*	NPV**	Efficiency
d_1	0.40	0.96	0.94	0.90	0.98	0.95
g_6	0.55	0.97	1.00	1.00	0.97	0.98
e_1	0.50	1.00	0.98	0.94	1.00	0.98
w_{-4}	0.85	0.92	0.98	0.96	0.96	0.96
w_2	1.00	0.60	0.91	0.60	0.91	0.85

* PPV = positive predictive value, ** NPV = negative predictive value.

always 100% and, if included, they decrease the prevalence of positives. Moreover, some patients whose specimens yielded discordant clinical and *in vivo* data were discharged and considered at a later date with sophisticated *in vitro* techniques. Confirming previous data (3), in six out of eight patients, *in vitro* tests gave results more concordant with the clinical findings than *in vivo* tests; however, they were not considered for the ROC analysis since our study aim was to identify threshold levels in a well-defined population.

The clinical performance of a laboratory test can be described in terms of "diagnostic efficiency". The latter refers to the quality of the information provided by the classification device and should be distinguished from the usefulness or practical value of the information itself. ROC plots provide a pure index of accuracy by demonstrating the limitation in a test's ability to discriminate between alternative states of health over the complete spectrum of operating conditions. Furthermore, once the plot is generated, a user can readily go on to other activities, such as selecting decision thresholds. The most common overall measure is the area under the ROC curve. Traditionally, this area has always been >0.5 , values ranging from 1.0 (perfect separation of the test values of the two groups) to 0.5 (no apparent distributional difference between the two groups of test values). We found that the areas under the ROC curves were satisfactory, ranging from 0.931 (e_1) to 0.974 (g_5) when w_4 was excluded. For this allergen, in fact, the area under the ROC plot was significantly different (0.8). This finding confirms the high clinical efficiency of the CAP System for the IgE measurement of the most common aeroallergens but also indicates the limitations of this technique for some particular allergens. The evidence that there are fewer "positives" for this allergen ($n = 15$) than "negatives" ($n = 65$) gives rise to an alternative hypothesis.

Different thresholds for each allergen were found, thus demonstrating the importance of threshold selection for an optimal clinical use of IgE measurement in serum. In some cases, we found two different thresholds for the same allergen, one resulting in a better specificity, the other in a better sensitivity. This may suggest a different practical utilization of the results of *in vitro* measurement of the allergen-specific IgE antibody depending on the specific clinical situation, thus stressing the need to confirm a clinical suspicion (specificity) or the choice of underlining the sensitivity of the test. From a clinical viewpoint, because more than 300 allergen specificities exist, the definition of specific thresholds should be limited to the major aeroallergens. Other allergens, such as foods or Hymenoptera venoms, require further investigation.

Acknowledgment

We gratefully acknowledge the expert secretarial assistance of Ms Paola Sauguin.

References

- Leimgruber A, Mosimann B, Claeys M, et al. Clinical evaluation of a new *in vitro* assay for specific IgE, the ImmunoCAP System. *Clin Exp Allergy* 1991;21:127-31.
- Bousquet J, Chanet P, Chantal I, Michel F-B. Comparison between RAST and Pharmacia CAP System: a new automated specific IgE assay. *J Allergy Clin Immunol* 1990;85:1039-43.
- Plebani M, Borghesan F, Faggian D. Clinical efficiency of *in vitro* and *in vivo* tests for allergic diseases. *Ann Allergy* 1995;74:1:23-8.
- Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39(4):561-77.
- Hanley JA. Receiver-operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging* 1989;29:307-35.
- Società Italiana di Allergologia e Immunologia Clinica. "Memorandum" sulla diagnostica delle allergopatie. *Folia Allergol Immunol Clin* 1987;34:387 (Fed Med 1987;40:861-70).
- Ewan PW, Coote D. Evaluation of a capsulated hydrophilic carrier polymer (the ImmunoCAP) for measurement of specific IgE antibodies. *Allergy* 1990;45:22-9.
- Kleine-Tebbe J, Eickholt M, Gätjen M, et al. Comparison between MAGIC-LITE and CAP-System: two automated specific IgE antibody assays. *Clin Exp Allergy* 1992;22:475-84.
- Galen RS, Gambino SR. Beyond normality: the predictive value and efficiency of medical diagnosis. New York: John Wiley, 1975.
- Gerhardt W, Keller H. Evaluation of test data from clinical studies. I. Terminology, graphic interpretation, diagnostic strategies, and selection of sample groups. II. Clinical review of the concepts of efficiency, receiver-operating characteristic (ROC), and likelihood ratios. *Scand J Clin Lab Invest* 1986; Suppl 181:1-74.
- Kairisto V, Poola A. Software for illustrative presentation of basic clinical characteristics of laboratory tests GraphROC for Windows. *Scand J Clin Lab Invest* 1995;55 Suppl. 222:43-60.
- American Academy of Allergy and Immunology. Executive Committee. Position Statement. The use of *in vitro* tests for IgE antibody in the specific diagnosis of IgE-mediated disorders and in the formulation of allergen immunotherapies. *J Allergy Clin Immunol* 1992;90:263-7.
- Romano L, Zelazski D. Study of interassay, interlab imprecision profile of specific IgE measurement against two allergens with CAP System. *Allergy Clin Immunol News* 1994; Suppl. 2:4.
- De Blay F, Zana H, Offner M, Verot A, Velten M, Pauli G. Receiver-operating characteristic analysis: a useful method for a comparison of the clinical relevance of two *in vitro* IgE tests. *J Allergy Clin Immunol* 1993;92:255-63.
- 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.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.