

Short communication

Receiver operating characteristic curve analysis of cutaneous skin test reactions to predict hay fever and asthma symptoms in the Normative Aging Study

Background: We used receiver operating characteristic (ROC) curves to determine optimal cutoff values and to evaluate the ability of ragweed, mixed grasses, mixed trees, and house-dust antigens to predict symptoms of hay fever and asthma.

Methods: Between 1984 and 1993, 1386 members of the Normative Aging Study cohort had at least one examination that included evaluation of cutaneous hypersensitivity to common aeroallergens by skin prick testing. Standard questionnaires based on the American Thoracic Society DLD-78 questionnaire obtained information on respiratory symptoms, illness, and smoking habits.

Results: Ragweed was the best univariate predictor for both hay fever and asthma symptoms: the area under the ROC curve for hay fever was 0.71 (95% CI 0.67–0.74); for asthma, it was 0.65 (95% CI 0.55–0.75). For hay fever, dichotomizing the average reaction to four antigens at 2 mm gave a sensitivity of 56% and a specificity of 83%; at 3 mm, the sensitivity was 38% and specificity 92%. For asthma, dichotomizing the skin test reaction at 2 mm gave a sensitivity ranging from 30% for reaction to mixed grasses to 56% for the ragweed antigen reaction.

Conclusions: Although in this older population skin test reactivities did not predict current asthma symptoms with high sensitivity, our results support the use of cutoff values of 2 or 3 mm as commonly used in studies of hay fever symptoms.

**J. M. Weintraub^{1,3}, D. Sparrow²,
S. T. Weiss¹**

¹Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; ²Normative Aging Study, VA Outpatient Clinic, Boston, MA;

³Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Key words: allergens; asthma; hay fever; ROC curve; sensitivity and specificity; skin tests.

June M. Weintraub
Channing Laboratory
Brigham and Women's Hospital
181 Longwood Avenue
Boston
MA 02115
USA

Accepted for publication 2 November 2000

Skin tests have been used for over 100 years to characterize atopy and assess the individual response to particular allergens. A number of methodological issues arise in the use of skin tests for assessment of atopy, including the question of how the wheal and flare reactions are to be interpreted. The size of the wheal or flare reaction may range up to 45 mm, but usually does not exceed 11 or 12 mm (1). In a normal population, the distribution of reaction sizes typically follows a bimodal distribution. Because many individuals do not react to the antigens, the primary peak value is put at 0 mm; the range at which secondary peaks are seen differs depending on the potency of the antigen being tested as well as on individual responsiveness; for example, the peak response to house-dust antigen is 5–6 mm, while the peak response to Bermuda grass is 9–10 mm (1, 2).

Selection of a reaction size at which to declare a skin test reaction as positive requires a compromise between acceptable levels of sensitivity and the specificity of the test. Many studies use a diameter of 2 (3) or 3 (4, 5) mm as the cutoff level over which the response to the skin

test is considered positive; however, other definitions of positivity exist, including the erythema response (6, 7), the ratio of histamine wheal size to the allergen wheal size (8), a summary index combining the reactions to several antigens (9), and the area of the wheal or erythema reaction (10).

Previous studies have compared the clinical relevance and optimal cutoff values of skin test results for predicting clinical symptoms of allergy. Brand et al. evaluated several definitions of skin test positivity in relation to symptoms of house-dust allergy, concluding that a ratio of wheal to histamine size of 0.7 and 7 mm absolute wheal size were both optimal cutoffs (11). In a study of 499 Italian patients aged 8–59 years who had skin test wheal sizes greater than 3 mm, Pastorello et al. determined the optimal cutoff wheal area of 32 mm² for predicting hay fever or asthma symptoms (10). However, few studies have systematically evaluated optimal cutoff values for single antigen reactions or arithmetic combinations of antigen reactions in population samples that include symptomatic and asymptomatic

matic individuals. We used receiver operating characteristic (ROC) curves to determine optimal cutoff values and to evaluate the ability of several different skin test antigens to predict symptoms of hay fever and asthma in a relatively homogeneous group of healthy older men.

Material and methods

Population

The Normative Aging Study (NAS) is a longitudinal study of aging located at the Veterans Administration Outpatient Clinic in Boston (USA). The NAS cohort has been described in detail elsewhere (12). Recruited in the 1960s, the cohort initially consisted of 2280 male volunteers who were health-screened at entry; those with chronic medical conditions including asthma, chronic bronchitis, and chronic sinusitis were excluded. Members of this cohort have reported every 3–5 years for a comprehensive evaluation, including medical history, physical examination, blood and urine test, and spirometry. Beginning in 1984, allergy skin testing has also been performed.

Questionnaire data

Standard questionnaires based on the American Thoracic Society DLD-78 questionnaire (13) were used to obtain information on respiratory symptoms and illness as well as smoking habits. Current smokers were defined as those men who had smoked at least one cigarette/day for at least the past year, and who were still smoking at least 1 month before the examination. Former smokers were defined as those men who had previously smoked at least one cigarette/day for at least 1 year, but who had ceased smoking more than 1 month before the examination. Never smokers were defined as those men who had never smoked cigarettes or had smoked fewer than a total of 20 packs during their lifetime. Hay fever was defined as a positive response to one or both of the questions, “Have you ever had hay fever?” and “Was it confirmed by a doctor?” Subjects who reported that they had a doctor’s diagnosis of asthma were considered current asthmatic subjects.

Allergy skin testing

Cutaneous hypersensitivity to common aeroallergens was determined by skin testing by the prick method of Pepys (14). The volar surface of each arm was cleaned and dried, and glycerin-preserved allergens and a glycerin control were placed at 5-cm intervals. The allergens were supplied by Hollister-Stier (Spokane, WA, USA) and included house dust (1:10), ragweed (1:20), mixed trees (1:20), and mixed grasses (1:20). The allergens and glycerin control were in number-coded bottles in order to blind the research assistant performing the test. The skin was pricked with a Greer needle (Greer Laboratories, Winston-Salem, NC, USA). The reaction was measured at 20 min by adding the largest diameter of the wheal to its perpendicular, and then halving the sum. When a wheal reaction occurred to the glycerin control, the wheal diameter of this reaction was subtracted from the wheal diameter of each allergen reaction, before analysis.

If the wheal reaction to the glycerin control was greater than the wheal diameter of the allergen reaction, the wheal size for that allergen was set to zero. The mean wheal size was the arithmetic mean of the reactions to the four allergens (ragweed, mixed trees, mixed grasses, and house dust). Other arithmetic means were calculated similarly (e.g., mean of ragweed and mixed grasses reaction equals sum of the size of the two reactions divided by two).

Study sample

Between 1984 and 1993, 1386 cohort members had at least one examination that included skin testing. The individuals were between the ages of 43 and 93 at the time of testing, and for those individuals for whom multiple test results were available, the most recent results were used. One subject lacked information on asthma status, giving a total of 1385 men in the analyses of current asthma symptoms.

Statistical analysis

We evaluated by logistic regression the univariate associations between reaction to each of the four antigens, smoking, and age to prevalence of current asthma and hay fever symptoms. Multiple logistic regression models with different combinations of the four antigens and interaction terms were also considered. Model significance was assessed by the likelihood ratio test.

The predictive values of univariate skin test results and arithmetic combinations of wheal sizes were assessed and compared by ROC (15, 16). ROC curves plotting sensitivity against 1-specificity were generated, and the areas under the curves (AUC) calculated. The AUCs were compared by the nonparametric method for comparing the areas under correlated sample ROC curves described by DeLong et al., in which a covariance matrix is estimated and the resulting test statistic has a chi-square distribution (17).

Results

Age and smoking effects

The 1386 study subjects ranged in age from 43 to 93 years, with 277 subjects reporting hay fever symptoms and 27 current asthma at the time of their skin prick test. Overall, 51% of the population had a skin test reaction greater than or equal to 2 mm to at least one antigen. Mean age was similar regardless of the presence of asthma or hay fever symptoms (Table 1). The prevalence of any reaction greater than 2 mm decreased with age, from 61% among men under age 50 to 38% among men age 80 or older (*P* for trend under 0.001 in a logistic regression model adjusted for cigarette smoking).

The prevalence of any reaction greater than 2 mm was slightly higher in current smokers than nonsmokers

Table 1. Summary statistics

	Hay fever		Asthma	
	Symptoms	No symptoms	Symptoms	No symptoms
Number of subjects	277 (20%)	1109 (80%)	27 (2%)	1358 (98%)
Hay fever symptoms			17 (63%)	259 (19%)
Asthma symptoms	17 (6%)	10 (1%)		
Age (mean) years	64.0	65.3	64.1	65.1
Smoking (%)				
Never	90 (33%)	325 (29%)	5 (19%)	410 (30%)
Former	162 (58%)	657 (59%)	21 (78%)	797 (59%)
Current	25 (9%)	127 (12%)	1 (4%)	151 (11%)
Mean antigen reactions (mm)				
Ragweed	3.4	0.9	2.9	1.4
Mixed grasses	3.5	1.0	2.1	1.5
Mixed trees	2.3	0.8	2.1	1.1
House dust	1.5	0.8	1.9	0.9

(59% vs 50%, $P=0.047$). Neither smoking nor age was a significant univariate predictor of hay fever or asthma symptoms, and inclusion of smoking and age as covariates in the models with antigen reactions was not significant and did not substantially change the univariate estimates.

ROC analysis

Of the four antigens considered alone, ragweed was the best univariate predictor for both hay fever and asthma symptoms. The area under the ROC curve (AUC) for hay fever was 0.71 (95% CI 0.67–0.74); for asthma, the AUC was 0.65 (95% CI 0.55–0.75).

For hay fever, the AUC for ragweed was not significantly greater than the AUC for mixed grasses antigen (AUC=0.67, $\chi^2=2.86$, $P=0.091$), although it was significantly larger than that for either mixed trees (AUC=0.64, $\chi^2=10.03$, $P=0.002$) or house dust (AUC=0.60, $\chi^2=25.93$, $P<0.001$) (Table 2). Arithmetic combinations of the antigen reactions provided better performance than any of the single antigens alone. Of the combinations, the AUC for the arithmetic average of the wheal reaction to the four antigens was significantly higher than that for ragweed alone (AUC=0.77, $\chi^2=18.07$, $P<0.001$). The model using the highest reaction to any of the four antigens as a predictor also had high predictive value (AUC=0.78), and this was significantly better than the model based on the mean of the four aeroallergen responses ($\chi^2=3.83$, $P=0.05$). Composite scores and weighted averages were also investigated, but these did not perform significantly better than the models using the highest of four or three antigens. (Results not shown.)

For current asthma, the AUC for ragweed antigen was the highest, but this did not differ significantly from the AUCs for any of the other antigens or antigen combinations that we considered (Table 2).

For univariate models, each point on the ROC curve represents the sensitivity and 1-specificity of a model that dichotomizes the reaction score at a particular value of the wheal size or summary value (Fig. 1). In this way, the continuous values can be assessed, and a predictive value for a dichotomized reaction deter-

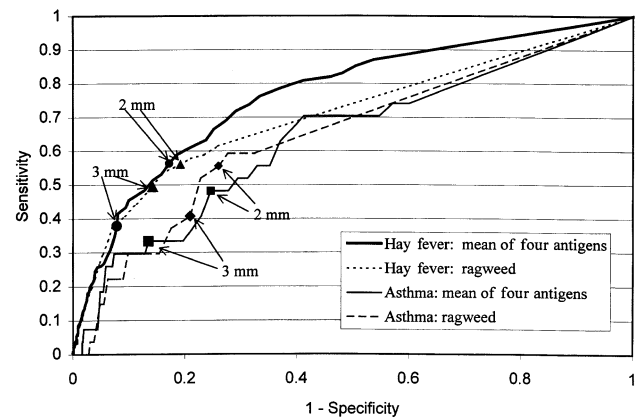


Figure 1. ROC curves for skin test reactions as predictors of hay fever and asthma.

mined. For hay fever, dichotomizing the mean antigen reaction at 2 mm gave a sensitivity of 56% and a specificity of 83%; at 3 mm, the sensitivity was 38% and the specificity 92%; at 5 mm, the sensitivity was 16% and the specificity 98%. As the dichotomization level is raised further, individuals with smaller reactions are missed, but at the same time the true negative rate increases.

Skin test reactions did not identify current asthma as well as they did hay fever. Dichotomizing the skin test reaction at 2 mm gave a sensitivity ranging from 30% for reaction to mixed grasses to 56% for the ragweed antigen reaction. Dichotomizing at higher cutoff values produced highly specific results, but the sensitivity levels are not likely to be acceptable for any application.

Discussion

In these analyses, the univariate models based on arithmetic means of two, three, or all four of the antigen reaction measurements performed as well as or better than the models including each antigen reaction as a separate covariate along with interaction terms (results not shown). For hay fever, the model based on the highest measurement of all four antigens provided the best fit and the largest predictive value based on the area under the ROC curve. However, the difference between the area under the ROC curve for the “highest” reaction to all four antigens compared to the mean of the reactions was not significant for either hay fever or asthma, and for simplicity and relevance, the latter model is preferable.

The members of the Normative Aging Study cohort are unique in that those with a history of childhood asthma were excluded at baseline. This exclusion could explain why the relationships between skin test reactivity and asthma were not as strong as those for hay fever. Additionally, because the cohort is older, their reactions to cutaneous skin tests were probably

Table 2. Areas under ROC curves for predictors of hay fever and asthma

Predictor	Hay fever				Asthma			
	AUC [†]	95% CI	χ^2 [‡]	P value	AUC [†]	95% CI	χ^2 [‡]	P value
Ragweed	0.71	(0.67, 0.74)	REF		0.65	(0.55, 0.75)	REF	
Mixed grasses	0.67	(0.64, 0.71)	2.86	0.091	0.52	(0.41, 0.62)	3.65	0.056
Mixed trees	0.64	(0.61, 0.68)	10.03	0.002	0.59	(0.49, 0.70)	0.95	0.331
House dust	0.60	(0.57, 0.64)	25.93	<0.001	0.62	(0.52, 0.73)	0.28	0.596
Highest	0.78	(0.74, 0.81)	20.06	<0.001	0.63	(0.52, 0.75)	0.17	0.682
Mean	0.77	(0.74, 0.80)	18.07	<0.001	0.64	(0.53, 0.76)	0.05	0.830

[†] Area under receiver operating characteristic curve.

[‡] χ^2 for difference between area under curve for each predictor, compared to area under curve for ragweed antigen alone.

smaller than those that would be seen in younger subjects, and this could also have attenuated the strength of the relationships. A recent review found that among adults the proportion of asthma cases attributable to atopy as measured by skin test positivity was, on average, 37%, but, as measured by total serum IgE, the proportion attributable to atopy could range from 13% to 67%, depending on the cutoff value used (18).

Our results are similar to those found in a general population sample of 9651 Swiss adults aged 18–60 years, where a mean wheal reaction to any of eight aeroallergens greater than 3 mm predicted current allergic rhinitis with a specificity of 85.9% and a sensitivity of 68.4% (5). It should be noted that as additional antigens are added to the battery of tests, the likelihood of having a positive reaction to at least one of the antigens, regardless of the cutoff level used, will increase.

Since smoking and age were not significant confounders in this population, we did not include these

variables in our models. This allowed us to interpret directly the points on the ROC curves to examine the different cutoff values. These single variable models are appealing for their easy interpretability. Although skin test reactivities to outdoor and indoor aeroallergens do not predict current asthma symptoms with high sensitivity, our results support the use of the 2- or 3-mm cutoff values commonly used in studies of hay fever symptoms.

Acknowledgments

The Normative Aging Study is supported by the Cooperative Studies Program/ERIC of the US Department of Veterans Affairs, and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). Dr Sparrow is an Associate Research Career Scientist of the Medical Research Service of the Department of Veterans Affairs. Grants for this study included HL34645 (National Heart, Lung, and Blood Institute), T32 ES07069 (National Institutes of Health), and CA40356 (National Institutes of Health).

References

1. BARBEE RA. Methodological issues in the assessment of atopy. In: WEISS ST, SPARROW D, editors. *Airway responsiveness and atopy in the development of chronic lung disease*. New York: Raven Press, 1989:121–156.
2. TOLLERUD D, O'CONNOR G, SPARROW D, WEISS S. Asthma, hay fever, and phlegm production associated with distinct patterns of allergy skin test reactivity, eosinophilia, and serum IgE levels. *Am Rev Respir Dis* 1991;**144**:776–781.
3. GOTTLIEB D, SPARROW D, O'CONNOR G, WEISS S. Skin test reactivity to common aeroallergens and decline of lung function: the Normative Aging Study. *Am J Respir Crit Care Med* 1996;**153**:561–566.
4. SIBBALD B, RINK E. Labeling of rhinitis and hay fever by doctors. *Thorax* 1991;**46**:378–381.
5. TSCHOPP JM, SISTEK D, SCHINDLER C, et al. Current allergic asthma and rhinitis: diagnostic efficiency of three commonly used atopic markers (IgE, skin prick tests, and Phadiatop). *Allergy* 1998;**53**:608–613.
6. GERGEN PJ, TURKELTAUB PC. The association of individual allergen reactivity with respiratory disease in a national sample: data from the second National Health and Nutrition Examination Survey 1976–80 (NHANES II). *J Allergy Clin Immunol* 1992;**90**:579–588.
7. GERGEN PJ, TURKELTAUB PC. The association of allergen skin test reactivity and respiratory disease among whites in the US population. *Arch Intern Med* 1991;**151**:487–492.
8. NIEMEIJER NR, FLUKS AF, DE MONCHY JGR. Optimization of skin testing. II. Evaluation of concentration and cutoff values, as compared with RAST and clinical history, in a multicenter study. *Allergy* 1993;**48**:498–503.
9. BURROWS B, MARTINEZ FD, HALONEN M, BARBEE RA, CLINE MG. Association of asthma with serum IgE levels and skin-test reactivity to allergens. *N Engl J Med* 1989;**320**:271–277.
10. 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 allergy from patients with asymptomatic allergy to common aeroallergens. *J Allergy Clin Immunol* 1995;**96**:580–587.
11. BRAND PLP, KERSTJENS HAM, JANSEN HM, KAUFFMAN HF, DE MONCHY JGR. Interpretation of skin tests to house dust mite and relationship to other allergy parameters in patients with asthma and chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 1993;**91**:560–570.
12. BELL B, ROSE CL, DAMON H. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* 1972;**2**:5–17.
13. FERRIS BG Jr. Epidemiology standardization project. *Am Rev Respir Dis* 1978;**118**(6, Part 2):1–88.
14. PEPYS J. Skin tests in diagnosis. In: GELL PGH, COOMBS RRA, LACHMAN PJ, editors. *Clinical aspects of immunology*. Oxford: Blackwell Scientific, 1985:5–80.
15. HANLEY JA, MCNEIL BJ. The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
16. HANLEY JA, MCNEIL BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;**148**:839–843.
17. DELONG ER, DELONG DM, CLARKE-PEARSON DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–845.
18. PEARCE N, PEKKANEN J, BEASLEY R. How much asthma is really attributable to atopy? *Thorax* 1999;**55**:268–272.