
The reproducibility of the allergy skin test scoring and interpretation by board-certified/board-eligible allergists

William A. McCann, MD and Dennis R. Ownby, MD

Background: Allergy skin testing is a cornerstone in the evaluation of the allergic patient. This seemingly simple test is subject to multiple variables that can affect the result.

Objective: To evaluate the degree of variability among board-certified/board-eligible allergists in the scoring and interpretation of allergen skin tests.

Materials and Methods: A series of allergen prick skin tests were digitally photographed and a questionnaire generated. Approximately 70 board-certified/board-eligible allergists were asked to grade each test item and to interpret them as positive, negative, or indeterminate or if they desired a followup intradermal test.

Results: Thirty-three interpretable responses were obtained. The majority of respondents (24) used a grading scale of 0 to 4. Agreement among physicians using a 0 to 4+ scale ranged from a standard deviation of 0.26 to 1.35, with greatest agreement on items with median/mode scores of 4+. The largest standard deviations were found on test items with median/mode scores of 1+ to 2+. Interpretation of the test items also showed greatest variation for those items with median/mode scores of 1+ to 2+. The number of intradermal tests requested ranged from 0 to 11 (of 22 test items).

Conclusions: The results demonstrate interphysician variation in the scoring and interpretation of epicutaneous skin tests. A questionnaire such as the one used here may serve a useful quality control instrument to ensure reproducible scoring of skin tests. In addition, the results highlight the need for greater study on the clinical utility of intradermal skin testing when epicutaneous tests are negative or equivocal.

Ann Allergy Asthma Immunol 2002;89:368-371.

INTRODUCTION

Allergy skin testing for immediate hypersensitivity is a cornerstone in the evaluation of the patient with allergic disease. This seemingly simple test, however, is subject to multiple variables that can affect the result. Some of these are patient dependent, such as the patient's age, underlying skin condition, or use of medications that can interfere with the test results (eg, antihistamines, tricyclic antidepressants). Testing dependent variables include the quality of the extracts used, the testing technique and testing device used, the location on the body to which the tests are applied, and the distance between individual test sites.¹ Finally, individual physician scoring and interpretation of allergen skin tests may add further variability.

Multiple studies have been performed to investigate many of these variables. Meinert et al² evaluated the changes in size of allergen wheal reactions over a 2-year period in 1,135 primary school children. This analysis demonstrated a year-to-year increase in the average wheal size related to the natural growth of the children. This change, however, affected the interpretation of the test if the observer based the reading on wheal size solely. Interpretation base on a ratio of

wheel size between the allergen skin test and histamine control was unaffected. Grabbe et al³ examined the effect of patient skin disorders on skin test results by examining tests on patients with atopic eczema, allergic rhinitis, and nonatopic controls. As might be expected, skin test results were less reproducible in patients with severe eczema. Imber⁴ analyzed multiple test variables including the testing method, extract concentration, diluent, extract manufacturer, and mixtures on skin test reactivity using 40 patients and a total of 2,395 skin tests. He demonstrated that each variable did indeed affect skin test results. Multiple studies have evaluated the validity and reproducibility of various skin test devices.⁵⁻⁷ One of these studies⁶ also noted a significant gradient of reaction size on the back. The use of different methods of scoring the test results can also add to variability.^{8,9}

Given these multiple variables, one might expect to see variations in skin testing procedures and results among clinical practices. Rodriguez et al¹⁰ analyzed skin test data from four allergy practices in Virginia. The study demonstrated significant differences between practices in the number of skin test performed, the number of positive tests per patient, and the grading of a positive test. This highlighted the need for greater standardization of skin testing procedures and quality control. To that end, both the American College of Allergy, Asthma and Immunology and the American Academy of Allergy, Asthma and Immunology have adopted po-

Allergy-Immunology Section, Department of Pediatrics, Medical College of Georgia, Augusta, Georgia.

Received for publication February 5, 2002.

Accepted for publication in revised form March 25, 2002.

sition statements that detail standardized skin testing procedures.^{11,12}

In our review of the literature we did not identify any study that directly examined the role of interphysician variation in the scoring and interpretation of epicutaneous allergen skin test results.

OBJECTIVE

To evaluate the degree of variability among board-certified and board-eligible allergist-immunologists in the scoring and interpretation of allergen skin tests.

MATERIALS AND METHODS

The study protocol was reviewed and approved by the institution's human assurance committee. Adult patients undergoing routine allergen skin prick testing in the allergy and immunology clinic were asked to participate. After informed consent was obtained, a series of digital photographs were taken of individual skin tests. Each photo included a millimeter ruler to provide scale. For each patient, photographs of positive (histamine and/or compound 48/80) and negative (glycerosaline) controls were taken in addition to several extract skin tests. From the photographs taken, three series (A, B, and C) were selected based upon photographic quality and differing degrees of wheal and flare. Each series contained a negative control, at least one positive control, and a series of unknown, or test, items (A1-A9, B1-B7, and C1-C6; Fig 1). An accompanying questionnaire asked board-certified or board-eligible allergists to score or grade each test item as they would in their office with a brief explanation of the grading system he/she used. In addition, they were asked to interpret each test item, given an appropriate history and physical examination, as positive, negative, or indeterminate, or whether they desired to follow up with an intradermal skin test. Approximately 70 allergists were asked to participate; the majority (45) were members of a large, multisite private allergy practice. Interpretable results were tabulated. Skin

tests scored on a 0 to 4+ semiquantitative scale were grouped, and average grade, median and mode grade, and standard deviations were calculated.

RESULTS

Thirty-six responses were obtained, two of which were incomplete and one was not interpretable because of multiple responses to individual items. Thirty-three complete and interpretable responses were available for analysis, with 21 responses from the large, multisite practice. The majority of respondents (24) used a grading scale of 0 to 4+, with 17 members of the multisite practice using such a scale. With some minor variations, most of these physicians scored tests using the modified University of Michigan grading system (Table 1).¹³

Two respondents recorded the wheal and flare in millimeters. The remainder (7) provided no grading and only provided interpretations of the test items.

Agreement among physicians using a 0 to 4+ scale ranged from a standard deviation of 0.26 for items A8 and C3 to 1.35 for item C5. The lowest standard deviations, and thus greatest agreement, were found for those items with median and mode scores of 4+. The largest standard deviations (more than one) were found on test items with a median and mode scores of 1+ to 2+ (Table 2).

Interpretation of the test items also revealed a range of agreement. On several test items (A8, B4, B6, C1, and C3), all 33 respondents interpreted the test as positive. Twenty-five or more physicians, or 75%, interpreted a total of 10 items as positive. At the other end of the spectrum, four items were interpreted as negative by 20 or more physicians. As seven respondents preferred to perform intradermal tests on all items not interpreted as positive, this represents 75% (20 of 26) of the respondents interpreting tests as negative in agreement for a given item. The greatest variation in interpretation was seen in those test items with median/mode

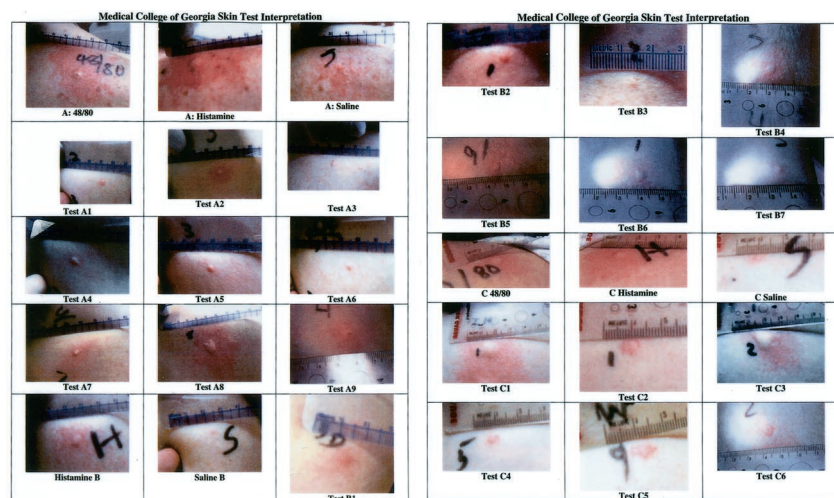


Figure 1. Questionnaire generated from digital photographs of epicutaneous skin tests. Each of three series of photographs contains appropriate positive and negative control tests followed by several test items (A1-A9, B1-B7, and C1-C6). Each photograph contains a millimeter ruler for scale.

Table 1. Grading System for Skin Testing

Grade	Skin appearance
0	No reaction or reaction no different than negative control
1+	Erythema less than 21 mm
2+	Wheal less than 3 mm and erythema larger than 21 mm
3+	Wheal greater than 3 mm with surrounding erythema
4+	Wheal with pseudopods and surrounding erythema

scores of 1+ to 2+ with standard deviations greater than 1.0 (Table 3).

The number of intradermal tests requested by an individual physician ranged from 0 to 11 (of 22 test items) with an average of 3.5 and a standard deviation of 3.75.

Analysis of the responses from physicians from the multisite practice versus other respondents did not reveal any significant differences in the grading or interpretation of individual test items.

DISCUSSION

First introduced by Blackley in 1873,¹⁴ allergen skin testing has been an essential part of the evaluation of the allergic patient for more than 100 years. Although appearing to be a rather simple test, it is one influenced by multiple intrinsic and extrinsic factors. Our results indicate that interphysician differences in the grading and interpretation of allergen skin

Table 2. Skin Test Grading

Test item	Median	Mode	Mean	StDev
A1	0	0	0.33	0.56
A2	2	2	1.94	0.81
A3	0	0	0.29	0.62
A4	3	3	2.81	1.00
A5	1	1	1.48	0.79
A6	0	0	0.30	0.70
A7	2.5	3	2.35	0.78
A8	4	4	3.93	0.26
A9	3	3	3.33	0.55
B1	3	3	2.83	0.54
B2	1	1	1.55	1.01
B3	0	0	0.37	0.93
B4	3	3	2.80	0.52
B5	2	2	1.98	0.88
B6	3	3	3.23	0.63
B7	2	2	1.52	0.98
C1	4	4	3.68	0.55
C2	2	2	1.75	1.18
C3	4	4	3.93	0.26
C4	0.25	0	0.67	0.86
C5	1.5	1	1.89	1.35
C6	2	3	2.33	0.87

Results from 24 respondents using 0 to 4+ grading system.

Table 3. Skin Test Interpretation

Test item	Positive	Negative	Indeterminate	Intradermal
A1	0	22	2	9
A2	18	4	2	8
A3	2	21	1	9
A4	27	1	1	4
A5	6	6	9	11
A6	1	21	2	8
A7	27	2	0	4
A8	33	0	0	0
A9	30	0	0	1
B1	28	1	1	1
B2	10	7	7	9
B3	1	23	3	6
B4	33	0	0	0
B5	19	3	3	7
B6	33	0	0	0
B7	9	9	7	7
C1	33	0	0	0
C2	12	5	7	9
C3	33	0	0	0
C4	1	19	4	9
C5	10	4	7	11
C6	26	2	1	4

Number corresponds to number of respondent interpreting a given test item as positive, negative, indeterminate, or requiring intradermal test. Not all 33 respondents interpreted all test items.

tests may add yet another variable to the performance of this integral test.

The results indicate that most of the respondents use a similar 0 to 4+ grading system. This may, however, reflect the fact that many of the respondents are part of a large multisite group. Grading of individual test items demonstrates fairly consistent scoring of tests at either end of the spectrum, ie, those with strong reactivity or little to no reactivity. Test items that fell in between the extremes generated the most interphysician scoring variability. Although this may reflect inherent limitations of the survey itself (eg, photographic quality, lack of three-dimensional relief), it also demonstrates interphysician differences in the scoring of skin prick tests and suggests that not all physicians strictly follow the scoring criteria used in a semiquantitative scoring system. Although proficiency test methods for evaluating accuracy, precision, and reproducibility of skin testing are encouraged,¹¹ one should also ensure reproducible, accurate scoring of skin tests. Our results suggests that a survey such as this one could be used to ensure that all personnel charged with scoring skin tests do so in a uniform, consistent fashion. It also suggests that measuring wheal and flare in millimeters alone, without a semiquantitative scale, may avoid this pitfall and thus be more reproducible.

Interpretation of the skin tests demonstrates a pattern similar to that seen in the scoring of the tests. Again, there is marked agreement on strongly positive and negative tests and more disagreement on test items in between. Although these

intermediate tests may indicate weaknesses within individual test items, they also highlight the need to ensure consistent scoring of the test items and agreement as to what constitutes a significant, positive reaction.

In addition, the results show a wide variation in the use of intradermal skin tests as adjuncts to prick skin testing. Although intradermal skin tests are essential in the diagnosis of Hymenoptera and drug allergy, their use in the evaluation of inhalant allergy is a matter of some debate. The customary concentration used for intracutaneous testing is as much as 50 to 100 times more diluted than that used for prick/puncture testing, thus providing increased sensitivity. This increased sensitivity may, however, come at the loss of clinical specificity. One study¹⁵ comparing the clinical utility of epicutaneous with intradermal skin testing in timothy grass-allergic patients found that a positive intradermal skin test in light of a negative prick test did not indicate clinically significant sensitivity to timothy grass. A similar study¹⁶ evaluating cat allergy also found that intradermal testing added little to the diagnostic evaluation. There are, however, fewer data on the use of intradermal skin testing with weaker or poorer quality extracts, such as dog or mold.¹⁷ This ambiguity is reflected in the Practice Parameters for Allergy Diagnostic Testing,¹¹ which states that

"Intracutaneous testing may be useful and should be pursued if the prick/puncture test is negative or equivocal to allergens strongly suggested by the patient's history or exposure."

Certainly, the physicians in our sample appear to reflect a wide range of opinion regarding the use of intradermal skin tests after a negative or equivocal epicutaneous skin test.

CONCLUSION

The use and interpretation of allergen skin testing can be viewed as a distinctive hallmark that separates the allergist from other physicians. Although appearing quite simple, the test is fraught with multiple testing variables and its interpretation requires a thorough history and clinical judgment. As such, some interphysician variation in the scoring and interpretation of epicutaneous skin testing is to be expected and our results reflect this variation. We feel that the use of a questionnaire such as the one used here may serve a useful quality control instrument to ensure reproducible scoring of skin tests, especially by physicians employing a 0 to 4+ semiquantitative scale. Larger-scale studies with this modality are warranted to validate this approach. In addition, the results highlight the need for greater study on the clinical utility of intradermal skin testing when epicutaneous tests are negative or equivocal.

ACKNOWLEDGMENTS

The authors thank the patient volunteers and physicians who participated in this study.

REFERENCES

1. Nelson H. Variables in allergy skin testing. *Allergy Proc* 1994; 15:265-268.
2. Meinert R, Frischer T, Karmaus W, Kuehr J. Influence of skin prick test criteria on estimation of prevalence and incidence of allergic sensitization in children. *Allergy* 1994;49:526-532.
3. Grabbe J, Zuberbier T, Wagenpfeil S, Czarnetzki BM. Skin prick tests to common allergens in adult atopic eczema and rhinitis patients: reproducibility on duplicate and repeated testing. *Dermatology* 1993;186:113-117.
4. Imber W. Allergic skin testing: a clinical investigation. *J Allergy Clin Immunol* 1977;60:47-55.
5. Mahan C, Spector S, Siegel S, et al. Validity and reproducibility of multi-test skin test device. *Ann Allergy* 1993;71:1.
6. Nelson H, Rosloniec D, McCall LI, Ikke D. Comparative performance of five commercial prick skin test devices. *J Allergy Clin Immunol* 1993;92:750-756.
7. Basomba A, Sastre A, Pelaez A, et al. Standardization of the prick test. A comparative study of three methods. *Allergy* 1985; 40:395-399.
8. Vohlonen I, Terho E, Koivikko A, et al. Reproducibility of the skin prick test. *Allergy* 1989;44:525-531.
9. Eigenmann P, Sampson H. Interpreting skin prick tests in the evaluation of food allergy in children. *Pediatr Allergy Immunol* 1998;9:186-191.
10. Rodriguez G, Dyson M, Mohagheghi H. The art and science of allergy skin testing. *Ann Allergy* 1988;61:428-432.
11. Bernstein IL, Storms WW. Practice parameters for allergy diagnostic testing. Joint Task Force on Practice Parameters for the Diagnosis and Treatment of Asthma. The American Academy of Allergy, Asthma and Immunology and American College of Allergy, Asthma and Immunology. *Ann Allergy Asthma Immunol* 1995;75:543-625.
12. Board of Directors, AAAAI. Position statement: Allergen skin testing. *J Allergy Clin Immunol* 1993;92:636-637.
13. Sheldon J, Lovell R, Mathews K. *A Manual of Clinical Allergy*. 2nd ed. Philadelphia, PA: W. B. Saunders Company, 1967.
14. Blackley C. *Experimental Researches on the Causes and Nature of Catarrhus Aestivus (Hay Fever or Hay-Asthma)*. London: Balliere, Tindall and Cox, 1873.
15. Nelson H, Oppenheimer J, Buchmeier A, et al. An assessment of the role of intradermal skin testing in the diagnosis of clinically relevant allergy to timothy grass. *J Allergy Clin Immunol* 1996;97:1193-1201.
16. Wood RA, Phipatanakul W, Hamilton RG, Eggleston PA. A comparison of skin prick tests, intradermal skin tests, and RASTs in the diagnosis of cat allergy. *J Allergy Clin Immunol* 1999;103:773-779.
17. Nelson H. Variables in skin testing. *Immunol Allergy Clin North Am* 2001;21:281-290.

Requests for reprints should be addressed to:

William A. McCann, MD

Allergy-Immunology Section

Medical College of Georgia

Augusta, GA 30912

E-mail: wmccann@mail.mcg.edu