

Reproducibility of the skin prick test

I. VOHLONEN^{1,2}, E. O. TERHO³, A. KOIVIKKO⁴, T. VANTO⁴,
A. HOLMÉN² & O. P. HEINONEN¹

¹National Public Health Institute, Helsinki, ²Research Institute for Social Security, Social Insurance Institution of Finland, Helsinki, ³Dept. of Clinical Allergology, University of Turku, and ⁴Department of Pediatrics, Turku University Central Hospital, Finland

Reproducibility of the skin prick method of testing for allergy was studied in 20 subjects examined by four nurses. Hypodermic needles were used for pricking and the test panel included a histamine control, a diluent control, and nine allergens. The reproducibility of the method was best when the size of the weal reaction caused by an allergen was expressed as the geometric area of the weal. When the weal reaction was expressed as the ratio of the weal reaction caused by an allergen to that caused by histamine, the reproducibility of the method was decreased considerably. When the ratios were further classified into three class ratings, reproducibility was very low. The reduction in reproducibility was due to the low reproducibility of histamine reactions. According to these results, at least in epidemiological studies the weal reactions should be expressed as geometric areas. In clinical practice it might also be preferable to express prick test results as the diameters of the weals without adjusting them by histamine reactions.

Key words: allergens; epidemiology; histamine; reproducibility; skin prick test.

Accepted for publication 18 May 1989

The prick test is a common skin test method used for investigating the aetiology of immediate (Type I) allergies (18). This method has also been used in epidemiological studies to identify atopic subjects (2). Usually such tests have been performed with hypodermic needles (18), but recently the needles have often been replaced by prick test lancets (17). In clinical use in Scandinavia, the result of prick test is usually expressed as a ratio of the weal area produced by an allergen and the weal area produced by a histamine control (1). Outside Scandinavia and in many epidemiological studies the results have often been expressed, without reference to histamine control, as dichotomous (positive/negative) variables. Reactions with weal diameters of 3 mm or more are considered to be positive (4, 5-10). Despite the wide use of prick tests, studies of the reproducibility of this method are rare (3, 15, 20). The purpose of this study,

which was carried out in conjunction with a large population study, was to investigate the reproducibility of the prick test method.

SUBJECTS AND METHODS

A random sample of 2584 subjects from two geographical areas in Finland were selected in 1980 for an epidemiological study designed to investigate the occurrence of allergy to fodder yeast (16). Of this group, 10 symptomatic subjects (asthma or periodic wheeze, atopic dermatitis, allergic conjunctivitis, or hay fever or other allergic rhinitis) and 10 asymptomatic subjects were randomly selected and invited to participate in this study of the reproducibility of the skin prick test method.

The prick tests were made by four specially trained nurses. On the same occasion each nurse tested the same subject on different sites

of the forearms. The order of subjects, the order of testing nurses, the order of test solutions in the test panel, the order of testing sites of subjects (right or left forearm, radial or ulnar side), and the order of reading were randomized. Allergic state of the subject and the contents of the test solutions were unknown to the nurses. Each nurse made a total of 220 pricks and read 880 weals.

The antigen panel included 1 mg/ml histamine, glycerol-buffer diluent of the allergens, *Candida albicans* (5%), non-dialysed preparation of fodder yeast (20%), dialysed preparation of fodder yeast (20%), *Candida utilis* (10%), Bencard's fodder yeast preparation (10%), the house dust mite *Dermatophagoides pteronyssinus* (1 HEP), dog dander and hair (1 HEP), mugwort (*Artemisia vulgaris*) pollen (1 HEP), and timothy (*Phleum pratense*) pollen (1 HEP). The fodder yeast preparations (except for that prepared by Bencard) and *Candida utilis* (pure culture of fodder yeast) were prepared at the Department of Medical Microbiology, University of Turku. The principles of the preparation procedure have been described elsewhere (16, 21). *Candida albicans*, house dust mite, dog dander and hair, and pollen allergens as well as histamine and diluent control were commercial preparations (Soluprick®) made by Allergologisk Laboratorium A/S, Copenhagen. The same batches of test solutions were used throughout the study.

Hypodermic needles (25 G) were used for skin pricking. The needle was introduced through the drop of the test solution at 45° angle to the skin. The skin was pricked by elevating the skin slightly with the sharp point of the needle. The weal reactions were recorded 15–20 min later in the same order as the pricking had occurred, by drawing the outlines of the weals with an ink drawing pen. Each nurse recorded only the weal reactions resulting from her own pricking. The outlines of the weals were transferred to the study forms with transparent tape. One month later the sizes of all weals were read from the forms by the same nurses without knowledge of the pricker. The measurements of the weal diameters were done according to the Nordic recommendations. Skin test planimeter was not available for the study.

The size of the weal was expressed in four different ways depending on the type of analysis: 1) as the geometric area of a circle defined by $\pi \times r^2$, where r was estimated by dividing the sum of the greatest diameter of the weal and the diameter perpendicular to it by 4, 2) as the geometric area of a square defined by the product of the two above mentioned diameters, 3) as corrected area by dividing the area of the weal produced by an allergen with that produced by histamine, both reduced by the area of the weal produced by a diluent control ($A_{\text{allergen}} - A_{\text{diluent}} / A_{\text{histamine}} - A_{\text{diluent}}$), and 4) as a three-class

Table 1

Mean areas* ($\pm$ SD) of weal reactions to test solutions according to the type of test solution and the nurse who made and read the test ($n = 20$ tested subjects, 80 reactions per test solution)

Test solution	Nurse 1	Nurse 2	Nurse 3	Nurse 4
Histamine	11.7 (9.6)	9.7 (9.4)	12.9 (8.1)	14.5 (4.9)
Control solution	0.1 (0.4)	0.4 (0.6)	0.2 (0.4)	0.2 (0.5)
Timothy pollen	2.1 (6.2)	2.5 (6.2)	4.7 (13.4)	5.2 (13.5)
Mugwort pollen	7.7 (11.4)	4.8 (7.3)	7.4 (12.5)	9.1 (13.9)
Dog dander and hair	1.5 (4.4)	1.1 (2.9)	3.0 (7.9)	2.5 (7.3)
<i>Dermatophagoides pteronyssinus</i>	1.7 (3.2)	2.2 (3.5)	1.6 (2.0)	2.8 (5.1)
<i>Candida albicans</i>	2.6 (4.6)	3.0 (4.5)	3.4 (5.6)	3.4 (4.2)
<i>Candida utilis</i>	3.0 (4.3)	3.5 (4.2)	3.0 (6.5)	7.3 (14.3)
Bencard's fodder yeast	2.4 (4.4)	4.0 (7.2)	5.1 (8.4)	5.1 (7.8)
Dialysed fodder yeast	2.8 (5.5)	3.2 (6.4)	3.0 (4.4)	10.8 (26.2)
Non-dialysed fodder yeast	7.0 (12.8)	4.0 (6.4)	3.8 (6.8)	6.2 (11.1)
All solutions	3.8 (6.1)	3.5 (5.3)	4.4 (6.9)	6.1 (9.9)

*Area: $\pi \times r^2$, where $r = (\text{diameter 1} + \text{diameter 2})/4$.

Table 2

Mean areas* of weal reactions to 11 test solutions according to the nurse who made the test and the nurse who read the test

Pricker ¹	Reader ²			
	Nurse 1 (880)	Nurse 2 (880)	Nurse 3 (880)	Nurse 4 (880)
Nurse 1 (220)	4.9	4.5	5.3	5.6
Nurse 2 (220)	4.2	4.2	4.7	6.6
Nurse 3 (220)	4.3	5.3	5.4	5.8
Nurse 4 (220)	6.8	6.8	7.3	7.5

*Area: Diameter 1 $\times$ Diameter 2. ¹(No. of tests made).

²(No. of tests read).

(-/+ / + +) corrected area rating, where - = corrected area ≤ 0.24 (no reaction), + = corrected area 0.25–0.74 (mild reaction), and + + = corrected area ≥ 0.75 (strong reaction).

The differences between the four nurses' measurements of the areas of the weal reactions produced by the 11 test solutions were analysed by one-way ANOVA (11). The variation of weal reactions to various test solutions ($n = 11$), the variation of weal reactions to various test solutions according to the nurse ($n = 4$), and the variation of the weal reactions according to ways ($n = 3$) of expressing the size of the weal reaction were examined by variation coefficients (17). Estimations of standard errors of variation coefficients were not computed since it was not considered necessary to test the statistical significance of differences between various coefficients of variation.

The relationships between weal reactions to various test solutions were analysed by correlation analysis (12). The categorized ratings given by different nurses were compared by non-parametric tests of equal ratings (19). The reproducibility of weal reactions expressed as continuous variables (geometric areas) was analysed between the nurses by intraclass correlations (9, 22), and the reproducibility of the weal reactions expressed as categorized ratings was analysed by kappa coefficients (6, 8). In addition, confidence intervals (7) were estimated for the areas and the mean diameters of the weal reactions with respect to concordant (positive and negative) and discordant (positive or negative) categorized ratings.

RESULTS

When the same nurse both made the prick and read the reaction, and the size of the weal reaction was expressed as the geometric area of a circle, the mean sizes of the weal reactions did not differ statistically significantly between nurses for any of the test solutions. The mean area (all nurses) of the weal reactions was largest for histamine and smallest for the diluent control. Of the allergen solutions, mugwort pollen caused the largest mean weal reaction (Table 1).

The sizes of the weal reactions varied slightly according to which nurse read the diameters and clearly differed according to which nurse made the pricks (Table 2). In this analysis, for practical reasons, the size of a weal reaction was expressed as the product of the two perpendicular diameters of the weal (geometric area of a square).

In variation coefficients of weal reactions to different test solutions, the differences between nurses were rather small. Some of the most representative variation profiles are shown in Table 3. The variation coefficient of the weal area for histamine was less than half that for individual allergens; but of all active test solutions, the relative difference between nurses was largest for the variation coefficients for histamine. Apart from reactions to control solution, the largest variation between nurses was for

Table 3

Variation coefficients* of the areas of weal reactions to test solutions according to the type of test solution and the nurse who made and read the test ($n = 20$ tested subjects, 80 reactions per test solution)

Test solution	Nurse 1	Nurse 2	Nurse 3	Nurse 4
Histamine	0.82	0.97	0.63	0.34
Control solution	4.00	1.50	2.00	2.50
Timothy pollen	2.95	2.48	2.85	2.60
Dog dander and hair	2.93	2.64	2.63	2.92
<i>Dermatophagoides pteronyssinus</i>	1.88	1.59	1.25	1.82
Dialysed fodder yeast	1.96	2.00	1.47	2.43

* Standard deviation of the area/Mean of the area; where area: Diameter 1 $\times$ Diameter 2.

Table 4

Variation coefficients* of the areas of weal reactions to test solutions according to the type of test solution and the type of grading ($n = 20$ tested subjects, 80 reactions per test solution)

Test solution	Grading	
	Uncorrected ¹	Corrected ²
Histamine	0.67	-
Control solution	2.50	-
Timothy pollen	2.89	1.46
Mugwort pollen	1.56	1.03
Dog dander and hair	2.95	0.72
<i>Dermatophagoides pteronyssinus</i>	1.71	0.96
<i>Candida albicans</i>	1.52	0.86
<i>Candida utilis</i>	2.00	1.00
Bencard's fodder yeast	1.69	1.24
Dialysed fodder yeast	2.88	1.23
Non-dialysed fodder yeast	1.81	1.09

¹area: $\pi \times r^2$, where $r = (\text{diameter } 1 + \text{diameter } 2)/4$.

²area: $(\text{allergen diameter } 1 \times \text{allergen diameter } 2) - (\text{control diameter } 1 \times \text{control diameter } 2) / (\text{histamine diameter } 1 \times \text{histamine diameter } 2) - (\text{control diameter } 1 \times \text{control diameter } 2)$.

*Standard deviation of the area/Mean of the area.

correlation coefficients of weal reactions to dialysed fodder yeast.

According to correlation analysis, the weal reactions were highly significantly correlated to the various allergens, especially to different preparations of fodder yeast. Correlations between weal reactions to histamine and those to other solutions were weak.

Next, the size of the weal reaction to allergen was expressed as the geometric area of a square and divided by that to histamine, both being reduced by the area of the weal reaction to the diluent control (corrected area). Now the variation coefficients for areas of the weals produced by nine allergens were reduced to half of those computed based only on the size of the weal reaction expressed as the geometric area of a circle (Table 4). This reduction was due to the difference between the variation coefficients of the areas of histamine and allergens.

When the weal reactions to allergens expressed as corrected areas were categorized into three class ratings, the distribution of subjects according to reactions to different allergens differed considerably, depending on which nurse

had made the prick test. The differences were greatest for differentiation between negative and mild positive reactions. Concordant negative ratings between the four nurses were found in 5%, concordant any positive ratings in 29%, and discordant ratings in 66% of the allergen tests. The 95% confidence interval for the diameter of the weal reaction in concordant negative ratings was less than 2 mm, in discordant ratings 2–3 mm, and in concordant positive ratings more than 3 mm.

Reproducibility of the prick test method was greatest when weal reactions were expressed as geometric areas on a continuous scale (Table 5). Among the lowest reproducibility coefficients were those for histamine and the diluent control. Reproducibility decreased when weal area was compared with that for histamine and corrected for the weal area of the diluent control (corrected area). When the corrected weal reactions were expressed as categorized ratings, the reproducibility of the corrected ratings was further decreased to less than half that for geometric area.

DISCUSSION

As expected, histamine solution produced the largest mean weal reactions, since half of the subjects were asymptomatic. These reactions did not correlate with those produced by different allergens. Furthermore, although among the 20 subjects the variation coefficient for the weal reaction to histamine was small, the relative difference between nurses in the variation coefficient for histamine, as well as that for diluent control, was great. The reproducibility of pricking with histamine was lower than that for pricking with allergens. Consequently, if the size of the weal reaction is expressed according to current Scandinavian practice (1) by comparing the weal reaction produced by an allergen with that produced by histamine, the variation between prickers in the numerator is smaller than that in the denominator. Thus, the method is very sensitive, but low in specificity (13).

The current Scandinavian practice tends to give false negative results for allergic reactivity.

Table 5

Reproducibility coefficients of the prick test method for the four nurses according to the type of test solution and the type of grading the area of the weal reaction ($n = 20$ tested subjects, 80 reactions per test solution)

Test solution	Uncorrected ¹	Corrected ²	Categorized ³
	Intraclass-correlation		Kappa
Histamine	0.48	-	-
Control solution	0.17	-	-
Timothy pollen	0.77	0.68	0.40
Mugwort pollen	0.84	0.47	0.34
Dog dander and hair	0.81	0.49	0.26
<i>Dermatophagoides pteronyssinus</i>	0.73	0.46	0.14
<i>Candida albicans</i>	0.61	0.44	0.22
<i>Candida utilis</i>	0.57	0.59	0.26
Bencard's fodder yeast	0.82	0.68	0.35
Dialysed fodder yeast	0.42	0.56	0.25
Non-dialysed fodder yeast	0.81	0.53	0.25

¹area: $\pi \times r^2$, where $r = (\text{diameter 1} + \text{diameter 2})/4$.

²area: (allergen diameter 1 $\times$ allergen diameter 2) -
(control diameter 1 $\times$ control diameter 2)/
(histamine diameter 1 $\times$ histamine diameter 2) -
(control diameter 1 $\times$ control diameter 2).

³grade: - = corrected area ≤ 0.24 ; + = corrected area 0.25-0.74; ++ = corrected area ≥ 0.75 .

In clinical work this low specificity is not necessarily a problem since the results of skin tests are usually interpreted in view of the subject's clinical history, and additional investigations can be performed. In addition, reproducibility of the histamine reactions could be improved by using concentrations for histamine that are 10 times stronger than those used in this study as reported later by others (14, 20). In epidemiological studies, however, the situation is different. The low specificity leads to misclassification of the study population in terms of the exposure or the disease and therefore weakens the basis for estimating risk in aetiological studies.

The suitability of histamine-corrected readings of weal reactions for epidemiological studies was further questioned by comparing variation coefficients for weal areas expressed as uncorrected geometric areas and corrected areas. Reduction of variation coefficients to less than half those for uncorrected geometric areas indicates that less than half of the variance of allergic weal reactions is due to patient's individual immunological characteristics. In an epidemiological study this would mean that if esti-

mations of the histamine-corrected areas are used, the power of the study is reduced considerably and an unnecessarily large sample is needed. In addition, since the mean area of histamine reactions was considerably larger than that of the reactions to allergens, the effects of the differences in nurses' performances are probably reflected unpredictably in the corrected area estimations as low reproducibility of prick testing between different nurses.

The reproducibility of the prick test method was further decreased when histamine-corrected weal reactions were categorized into three class ratings. Concordant ratings (positive or negative) were obtained only in about one-third of the tests. The low reproducibility was clearly shown when coefficients of reproducibility for the three ways of defining the size of the weal reaction were compared. The accepted level of reproducibility (kappa or intraclass correlation) usually is 0.80, which would mean here that in epidemiological studies the estimation of weal reaction as geometric area of the reaction, without any correction for reactions to histamine or control solution, would be preferable. The area of a circle forms about 80% of the area of a

square and the variances of the corresponding areas are the same. Therefore, the area of the weal reaction could be estimated well enough as the product of the two perpendicular diameters, the area of a square. This study showed that when the areas of weal reactions were measured in this way, the differences between the readings of different nurses were small.

The reproducibility of the prick test method also depends, of course, on technical factors. Our analysis showed that the areas of the weal reactions differed more according to which nurse made the prick test than according to who read the corresponding weal reactions. Thus the main source of methodological error for prick testing in epidemiological studies seems to be the use of several prickers rather than the use of several readers. The same nurse should be the pricker throughout a study. If this is not feasible, the prickers should be evenly distributed within the comparison groups, so that each nurse will prick the same proportion in each group. Better standardization of pricking may be achieved by using prick-test lancets (3, 17). The areas of weal reactions can be read after testing by using the record drawn on transparent tape. Thus the reproducibility of the readings can easily be checked.

In conclusion, in epidemiological studies the results of skin tests using the prick technique evidently should be expressed as geometric areas on a continuous scale. These areas are estimated as the product of two perpendicular diameters of the area of each weal reaction. Both histamine and diluent controls are unnecessary. In clinical practice the inclusion of histamine control, to detect hyporeactive subjects, and of diluent control, to detect subjects who yield non-specific reactions, is justified. Use of these solutions as controls in grading the weal reactions, however, causes more misclassification than does their exclusion from the grading. Thus, even in clinical practice, it might be preferable to express the results of skin prick tests as geometric areas or mean diameters of the weals. By using the prick test method and the concentrations of test solutions used in the present study, weal reactions with a diameter exceeding 3 mm apparently are true positive

reactions. This agrees with previous findings (4).

REFERENCES

1. Aas, K. & Belin, L.: Standardization of diagnostic work in allergy. *Acta Allergol.* 27, 439-469, 1972.
2. Barbee, R. A., Lebowitz, M. D., Thompson, H. G. & Burrows, B.: Immediate skin-test reactivity in a general population sample. *Ann. Intern. Med.* 84, 129-133, 1976.
3. Basomba, A., Sastre, A., Pelaez, A., Romar, A., Campos, A. & Carcia-Villalmanzo, A.: Standardization of the prick test. A comparative study of three methods. *Allergy* 40, 395-399, 1985.
4. Bryant, D. H., Burns, M. W. & Lazarus, L.: The correlation between skin tests, bronchial provocation tests and serum level on IgE specific for common allergens in patients with asthma. *Clin. Allergy* 5, 145-157, 1975.
5. Chan-Yeung, M., Schulzer, M., MacLean, L., Dorken, E. & Grzybowski, S.: Epidemiologic health survey of grain elevator workers in British Columbia. *Am. Rev. Respir. Dis.* 121, 329-338, 1980.
6. Cohen, J.: A coefficient of agreement for nominal scales. *Educ. Psych. Meas.* 20, 37-46, 1960.
7. Dixon, W. J. & Massey, F. J.: Introduction to statistical analysis. New York: McGraw-Hill, 1973.
8. Fleiss, J. L.: Measuring nominal scale agreement among many raters. *Psych. Bull.* 76, 378-382, 1971.
9. Fleiss, J. L. & Cohen, J.: The equivalence of weighted kappa and intraclass correlation coefficient as measures of reliability. *Educ. Psych. Meas.* 33, 613-619, 1973.
10. Haahtela, T. & Jokela, H.: Asthma and allergy in Finnish conscripts. *Allergy* 34, 413-420, 1979.
11. Hays, W. L.: Statistics for the social sciences. New York: Holt, Rinehart and Winston, 1973.
12. Kerlinger, F. N. & Pedhazur, E. J.: Multiple regression in behavioral sciences. New York: Holt, Rinehart and Winston, 1973.
13. MacMahon, B. & Pugh, T. F.: Epidemiology. Principles and methods. Boston: Little, Brown and Company, 1970.
14. Malling, H. J.: Skin prick testing and the use of histamine references. *Allergy* 39, 596-601, 1984.
15. Malling, H. J.: Reproducibility of skin sensitivity using a quantitative skin prick test. *Allergy* 40, 400-404, 1985.
16. National Board of Health: Fodder yeast allergy in Äänekoski in 1980. Epidemiological, clinical and allergological study of health hazards caused by a fodder yeast plant. (In Finnish). Statement No. 8. Helsinki: National Board of Health, 1981.
17. Østerballe, O. & Weeke, B.: A new lancet for skin prick testing. *Allergy* 34, 209-212, 1979.
18. Pepys, J.: Skin tests in diagnosis. In: Gell, P. G. H., Coombs, R. R. A. & Lachman, P. J., eds. Clinical aspects of immunology. 3rd ed. Oxford: Blackwell Scientific Publications, 1975:55-80.

19. Siegel, S.: Nonparametric statistics for behavioral sciences. New York: McGraw-Hill, 1956.
20. Taudorf, E., Malling, H. J., Laursen, L. C., Lanner, Å. & Weeke, B.: Reproducibility of histamine skin prick test. Inter- and intravariation using histamine dihydrochloride 1, 5, and 10 mg/ml. *Allergy* 40, 344-349, 1985.
21. Terho, E. O., Vohlonen, I., Husman, K., Rautalahti, M., Tukiainen, H. & Viander, M.: Sensitization to storage mites and other work-related and common allergens among Finnish dairy farmers. *Eur. J. Respir. Dis.* 71 (Suppl. 152), 165-174, 1987.
22. Winer, B. J.: Statistical principles in experimental design. Tokyo: McGraw-Hill, 1971.

Address:

Prof. *Erkki O. Terho*, M.D.

University of Turku

Department of Clinical Allergology

Hämeenkatu 4

SF-20500 Turku

Finland