

Variables affecting interpretation of skin prick test results

Suhan Günaştı Topal, Bilge Fettahloğlu Karaman, Varol L. Aksungur

Department of Dermatology, Faculty of Medicine, Çukurova University, Adana, Turkey

Abstract

Background: Both performer- and device-dependent variabilities have been reported in sizes of wheal responses to skin prick tests.

Objective: We aimed to evaluate whether or not variabilities in sizes of wheal responses influence the final interpretation of skin prick tests; in other words, the decision on whether or not there is an allergy to a given antigen.

Methods: Skin prick tests with positive and negative controls and extracts of *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus* were done for 69 patients by two different persons, using two different puncturing devices- disposable 22-gauge hypodermic needles and metal lancets.

Results: Among four different skin prick tests, the average coefficients of variation in sizes of wheal responses were near to or higher than 20% for all of them. On the other hand, in the final interpretation of results, kappa values indicated substantial or almost perfect agreements between these tests. However, the frequency of establishing allergy to the house dust mites widely ranged in these tests (20.8–35.8% for *D. farinae* and 20.8–28.3% for *D. pteronyssinus*).

Limitations: The conduction of the study in a single center and the comparisons of results of only two performers.

Conclusion: We feel that variabilities in sizes of wheal responses of skin prick test can influence its categorical results.

Correspondence:

Dr. Suhan Günaştı Topal,
Department of Dermatology,
Faculty of Medicine, Çukurova
University, Balcali, Saricam,
Adana 01330, Turkey.
E-mail: sugunasti@yahoo.com

Key words: Devices, intertester agreement, skin prick test, variability

Introduction

Skin prick test is a widely used test to detect immunoglobulin E mediated hypersensitivity to a given antigen. The size of the wheal which develops at the test site is influenced not only by the severity of the allergy, but also by other factors such as the type of the puncturing device and the skill of the tester. Numerous studies have been done to compare the effect of different puncturing devices on the size of the wheal.¹⁻⁸ Although fewer in number, there are also some previous studies demonstrating performer-dependent variability in skin prick test.^{9,10} On the other hand, the size of the wheal response to an antigen in skin prick test is not the sole determinant in deciding whether or not there is an allergy to this antigen. This is evaluated together with sizes of responses to both positive and negative controls.

In this study, we aimed to investigate whether or not the categorical result of skin prick test, which means the decision on whether or not there is an allergy to a given antigen, would be changed by the performer- or device-dependent variability in the size of the wheal

in the test. A doctor and a nurse, both experienced in doing skin prick tests, performed the test separately in the same patient. They used lancets and needles as puncturing devices, both of which are considered to be the tools of choice for skin prick test.⁶

Methods

Testers and patients

Two persons did skin prick tests in 69 patients with eczema. The first was a specialist in dermatology with 1½ years experience, while the second was a nurse working in our department of dermatology (Çukurova University, Medical School, Adana, Turkey) for 15 years. Both of them had had formal training, as well as practical experience of performing skin prick tests, for at least 3 years. Of the patients, 24 were men and 45 were women, their ages ranging from 12 to 76 years (mean 38.2 years). Their diagnoses were atopic dermatitis (19 patients), allergic contact dermatitis (37), irritant contact dermatitis (4), nummular dermatitis (5) and lichen

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Topal SG, Karaman BF, Aksungur VL. Variables affecting interpretation of skin prick test results. Indian J Dermatol Venereol Leprol 2017;83:200-4.

Received: September, 2015. **Accepted:** June, 2016.

Access this article online	
Quick Response Code:	Website: www.ijdvl.com
	DOI: 10.4103/0378-6323.192956

simplex chronicus (4). All patients were enrolled in the study after giving written informed consent.

Technique

The first tester performed the skin prick test on an area 3 cm distant from the antecubital fossa and 5 cm distant from the wrist, at the volar surface of the right forearm of each patient, while the second performed the same procedure on the left forearm. Before doing the test, test sites were wiped with ethanol. On each forearm, individual test sites were marked in two columns of four rows each with a 3 cm distance between them. Two drops of histamine dihydrochloride solution (1.7 mg/ml), used as the positive control, were placed on the first row, which was the nearest one to the antecubital fossa. The second row was used for the negative control (a mixture of sodium chloride, phenol and glycerol; 9, 4 and 563 mg/ml, respectively); the third, for an extract of *Dermatophagoides farinae* and the fourth, for an extract of *Dermatophagoides pteronyssinus*. Puncturing devices were disposable 22-gauge hypodermic needles and single use metal lancets with a 1 mm pointed tip and blunt shoulders. All solutions and lancets were obtained from Allergopharma (Reinbeck, Germany). On both forearms, individual test sites of the inner column were punctured with lancets at a right angle and those of the outer column with needles at an acute angle as described elsewhere.¹¹ Hence, a total of 16 pricks were done on each patient. Testing of each patient was completed in a single session. Antihistamines or any other drugs that might affect the skin prick test were stopped before the test as per durations suggested.¹²

Measurement and evaluation of skin prick test

Fifteen minutes after pricking, each investigator outlined wheals with a pen on their own application side. The size of a wheal was represented by its mean diameter. To calculate this mean diameter, the largest diameter of the wheal and a perpendicular one at the largest width of the former were measured with a transparent ruler and their sum was divided by two, as described elsewhere.¹¹

A test was considered to be valid (interpretable), if both of the following criteria were fulfilled: (1) The size of the wheal of the negative control should not exceed 3 mm and (2) the size of the wheal of the positive control should be at least 4 mm greater than the size of the wheal of the negative control.¹³ If the size of the wheal of a house dust mite extract was at least 3 mm greater than the size of the wheal of the negative control, this was accepted as positive.

Statistical analysis

Skin prick test done by the specialist and using the lancet was labeled as "specialist-lancet;" done by the specialist and using the needle, "specialist-needle;" done by the nurse and using the lancet, "nurse-lancet" and done by the nurse and using the needle, "nurse-needle." Paired *t*-test was used to calculate the mean of differences in sizes of wheals for negative and positive controls. To determine variability of measurement of sizes of wheals, the coefficient of variation for each patient was calculated as a ratio of the within-subject standard deviation to the mean.¹⁴ Then, the average of the individual coefficients of variations was taken. Inter-performer and inter-device agreements in the reading of tests for each house dust mite were evaluated with kappa tests in a subgroup of patients, in whom all of specialist-lancet, specialist-needle, nurse-lancet and nurse-needle tests were interpretable. Fleiss' kappa was calculated for overall comparisons of the four skin prick tests, using a calculator from statsToDo ([http://](http://www.statstodo.com)

www.statstodo.com). Cohen's kappa was calculated for pairwise comparisons, using Statistical Package for the Social Sciences (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp). Kappa values >0.80 indicated "almost perfect," 0.61–0.80 "substantial," 0.41–0.60 "moderate," 0.21–0.40 "fair" and 0.0–0.20 "poor" agreement.

Results

Minimum, maximum and mean sizes of wheals of the positive and negative controls in the specialist-lancet-, specialist-needle-, nurse-lancet-, and nurse-needle- labeled skin prick tests were given in Table 1. Differences in sizes of wheals between the paired tests were statistically significant for the positive control, but were not for the negative control [Table 2].

In measuring sizes of wheals to the positive controls, all of the four skin prick tests, labeled with specialist-lancet, specialist-needle, nurse-lancet and nurse-needle, resulted in equal values only in 2 of 69 (2.9%) patients. Three equal values were obtained in 5 (7.2%) patients and two equal values in 43 (62.3%) patients. In the remaining 19 (27.5%) patients, all of the four values were different. In contrast to the positive controls, getting four equal values was the norm for the negative controls, and in no patient were four different values observed. In 60 (87.0%) patients, all of the values obtained in specialist-lancet, specialist-needle, nurse-lancet and nurse-needle tests were zero, so they were equal. Three and two equal values were obtained in 4 (5.8%) patients and 5 (7.2%) patients, respectively. Getting four equal values was also the most common finding both for *D. farinae* and *D. pteronyssinus* and it was observed in 37 (53.6%) patients and 43 (62.3%) patients, respectively. Again, all of the values of the four skin prick tests in these patients were zero. Numbers of patients showing three, two and no equal values were 6 (8.7%), 18 (26.1%) and 8 (11.6%) for *D. farinae*, respectively. These numbers were 4 (5.8%), 14 (20.3%) and 8 (11.6%) for *D. pteronyssinus*, respectively. Thus, obtaining equal values in all of specialist-lancet, specialist-needle, nurse-lancet and nurse-needle tests was almost exclusively observed when there was no wheal response. In other words, if wheals occurred in these tests, there was a strong tendency to obtain different values. The average coefficient of variation was 29.1% for the positive control, 18.1% for the negative control, 34.2% for *D. farinae* and 19.8% for *D. pteronyssinus* [Table 3].

Comparison of the categorical results of the four skin prick tests, labeled with specialist-lancet, specialist-needle, nurse-lancet and nurse-needle, to house dust mites was done only in 53 patients, in whom all of these tests were interpretable. In this subgroup of the patients, the average coefficients of variation for the positive control, the negative control, *D. farinae* and *D. pteronyssinus* were 25.3%, 6.0%, 39.0% and 16.0%, respectively [Table 3]. A positive result to *D. farinae* was found in 12 (22.6%) patients with specialist-lancet test, in 16 (30.2%) patients with specialist-needle test, in 11 (20.8%) patients with nurse-lancet test and in 19 (35.8%) patients with nurse-needle test. All of these four tests resulted in same categorical results for *D. farinae* in 41 (77.4%) patients. Of these 41 patients, ten showed a positive result to *D. farinae*. A positive result to *D. pteronyssinus* was found in 15 (28.3%) patients with specialist-lancet test, in 14 (26.4%) patients with specialist-needle test, in 11 (20.8%) patients with nurse-lancet test and in 14 (26.4%) patients with nurse-needle test. All of these four tests resulted in same categorical results for *D. pteronyssinus* in 47 (88.7%) patients. Of these 47 patients, again ten showed a positive

result to *D. pteronyssinus*. The kappa value between the four skin prick tests was 0.699 (95% confidence interval = 0.589–0.809) for *D. farinae* and 0.834 (95% confidence interval = 0.724–0.944) for *D. pteronyssinus* [Table 3]. In pairwise comparisons, the lowest agreement was found for *D. farinae* between specialist-needle and nurse-needle tests (kappa = 0.617) and the highest agreement for *D. pteronyssinus* again between specialist-needle and nurse-needle tests (kappa = 0.903) [Tables 4 and 5].

Discussion

It is well known that there are both performer- and device-dependent variabilities in sizes of wheals in skin prick test.^{1-4,9,10} The aim of this study was that whether or not these variabilities could affect categorical results of the test. Before discussing our findings in this respect, we would like to clarify possible doubts about our materials and methods. Some of the queries could be: a) Why patients with some types of eczema, such as allergic contact dermatitis, were included in this study, since skin prick test has no role in the diagnosis or management of patients with these eczemas in routine daily practice. b) Why were the application sites of the investigators not randomly allocated despite a difference in reactivity reported between the right and the left forearms?^{15,16} c) Why were the wheal sizes measured by the testers instead of a third observer? d) Why was the categorical result based on comparison of the size of the wheals at the allergen vs negative control sites even though it has recently been noted that this method is no longer useful?¹²

To answer these doubts: a) Patients with some types of eczema were included since in this study, we had to take not only patients with a house dust mite allergy but also a substantial number of patients without such an allergy, to make comparisons of categorical results more objective between various skin prick tests.

b) Since our objective was to determine if size of wheals affects the final interpretation, and not the differences in wheal size *per se*, random allocation of test sites was not considered necessary

c) If one wishes to investigate whether or not differences between sizes of wheals are really due to performer's hands, the sizes of wheals should be measured by a third person. In this study, however, we aimed to investigate how the interpretation of results could be affected not only by the performers' hand skill, but also by his own mistakes in measurement, as in daily practice.

d) Despite the contrary view of a review,¹² comparison of wheal size utilising negative controls has been used in recent scientific studies.¹⁷⁻¹⁹

In this study, the coefficients of variation for the positive controls and *D. farinae* were markedly higher than the acceptable upper limit of 20% which has been recommended for skin prick test.²⁰ Those for the negative controls and *D. pteronyssinus* were near to this upper limit. Except than for the negative controls, these high values

Table 1: Minimum, maximum and mean sizes of wheals of the positive and negative controls in the specialist-lancet-, specialist-needle-, nurse-lancet- and nurse-needle-labeled skin prick tests

Labels of skin prick tests	Sizes of wheals (mm)			
	of the positive control		of the negative control	
	Maximum-minimum	Mean±standard deviation	Maximum-minimum	Mean±standard deviation
Specialist-lancet	0-10	5.9±1.50	0-3	0.12±0.56
Specialist-needle	3-16	7.4±2.05	0-3	0.19±0.69
Nurse-lancet	0-12	6.6±2.38	0-8	0.32±1.28
Nurse-needle	0-14	9.0±2.46	0-7	0.33±1.30

Table 2: Differences in sizes of wheals for the positive and negative controls between the paired tests

Pairs	Differences in sizes of wheals (mm)			
	for the positive control		for the negative control	
	Mean (95% CI)	P value	Mean (95% CI)	P value
Nurse-lancet versus specialist-lancet	0.74 (0.16-1.32)	<0.001	0.20 (0.02-0.42)	>0.05
Nurse-needle versus specialist-needle	1.61 (1.00-2.22)	0.013	0.14 (0.10-0.39)	>0.05

CI: Confidence interval

Table 3: Average coefficients of variation both in all patients and patients in whom categorical results were evaluated and agreements in categorical results

Solutions	All patients (n=69)	Patients in whom categorical results were evaluated (n=53)	
	Average coefficients of variation (%)	Average coefficients of variation (%)	Fleiss kappa values(95% CI)
Positive control	29.1	25.3	-
Negative control	18.1	6.0	-
<i>Dermatophagoides farinae</i>	34.2	39	0.699 (0.589-0.809)
<i>Dermatophagoides pteronyssinus</i>	19.8	16.0	0.834 (0.724-0.944)

CI: Confidence interval

Table 4: Pairwise comparisons for *Dermatophagoides farinae* between prick tests done by different performers and done with different devices

Pairwise comparisons	Frequencies of categorical results				Kappa values (95% CI)
	--	-+	+-	++	
Specialist-lancet versus specialist-needle	36	5	1	11	0.711 (0.493-0.929)
Specialist-lancet versus nurse-lancet	40	1	2	10	0.834 (0.651-1.0)
Specialist-lancet versus nurse-needle	34	7	0	12	0.687 (0.472-0.903)
Specialist-needle versus nurse-lancet	37	0	5	11	0.754 (0.550-0.959)
Specialist-needle versus nurse-needle	31	6	3	13	0.617 (0.390-0.845)
Nurse-lancet versus nurse-needle	34	8	0	11	0.638 (0.407-0.869)

CI: Confidence interval

Table 5: Pairwise comparisons for *Dermatophagoides pteronyssinus* between prick tests done by different performers and done with different devices

Pairwise comparisons	Frequencies of categorical results				Kappa values (95% CI)
	--	-+	+-	++	
Specialist-lancet versus specialist-needle	37	1	2	13	0.858 (0.701-1.0)
Specialist-lancet versus nurse-lancet	37	1	5	10	0.697 (0.468-0.925)
Specialist-lancet versus nurse-needle	37	1	2	13	0.858 (0.701-1.0)
Specialist-needle versus nurse-lancet	39	0	3	11	0.844 (0.672-1.0)
Specialist-needle versus nurse-needle	38	1	1	13	0.903 (0.771-1.0)
Nurse-lancet versus nurse-needle	39	3	0	11	0.844 (0.672-1.0)

CI: Confidence interval

were not markedly changed in the subgroup of patients, in whom comparisons of the categorical results were done. Hence, we showed that there was a marked difference in sizes of wheals between skin prick tests done by different performers and done using different devices, as shown previously in many other studies.¹⁻¹⁰

Stroking the skin results in a wheal and flare reaction. This phenomenon is known as the triple response of Lewis. The stronger the stroking, the more severe is the response. In skin prick test, there is trauma to the skin, and therefore, the triple response of Lewis may be expected in skin prick test. The strength of trauma in skin prick test may vary due to the performer's skill and to the form of the device. However, this response should affect all test sites. On the other hand, the categorical result of the test depends on the difference between the size of the wheals at the site of allergens and that at the site of the negative controls. If a performer or a device incites the triple response of Lewis, both of these sizes may be increased without changing the difference between them. Hence, it may be expected that performer- or device-dependent variabilities do not affect the categorical results.

At a first glance, this expectation was realized in our study since agreements in the categorical results of various types of skin prick test were either "substantial" or "almost perfect." Moreover, in most of the kappa values, the lower endpoint of the 95% confidence interval corresponded to a high agreement. On the other hand, differences in the frequency of patients showing a positive test result between various types of skin prick test were not within the clinically acceptable limit. While the kappa value between the four skin prick tests was 0.834 for *D. pteronyssinus* and corresponded to an almost perfect agreement, among 53 patients, a positive test result, in other words an allergy to this house dust mite, was found in 11 patients by one type of skin prick test, and in 15 patients by another type of skin prick test. Moreover, only ten of these 11 and 15 patients were common. In our opinion, skin prick test should be accepted to be an exception for the subject of inter-tester agreement for categorical variables and kappa values corresponding to a high agreement should be increased for skin prick test, as the situation of the coefficient of variation in skin prick test.

The main limitation of this study was that it was conducted in a single center and results of only two performers were compared. For more reliable conclusions, similar studies should be done in other centers and by comparing more than two performers.

Conclusion

We think that skin prick test shows marked performer- and device-dependent variabilities even in the categorical results. Therefore, studies aiming at standardization of the test should be done, since it is a valuable test because of its ease in application and its rapidity in giving results.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- Carr WW, Martin B, Howard RS, Cox L, Borish L. Comparison of test devices for skin prick testing. *J Allergy Clin Immunol* 2005;116:341-6.
- Yoon IK, Martin BL, Carr WW. A comparison of two single-headed and two multi-headed allergen skin test devices. *Allergy Asthma Proc* 2006;27:473-8.
- Dykewicz MS, Dooms KT, Chassaing DL. Comparison of the Multi-Test II and ComforTen allergy skin test devices. *Allergy Asthma Proc* 2011;32:198-202.
- Nelson HS, Rosloniec DM, McCall LI, Iklé D. Comparative performance of five commercial prick skin test devices. *J Allergy Clin Immunol* 1993;92:750-6.
- Nelson HS, Lahr J, Buchmeier A, McCormick D. Evaluation of devices for skin prick testing. *J Allergy Clin Immunol* 1998;101(2 Pt 1):153-6.
- Masse MS, Granger Vallée A, Chiriac A, Dhivert-Donnadieu H, Bousquet-Rouanet L, Bousquet PJ, et al. Comparison of five techniques of skin prick tests used routinely in Europe. *Allergy* 2011;66:1415-9.
- Caimmi D, Masse MS, Chiriac AM, Demoly P. Performances of an improved device for skin prick tests. *Int J Immunopathol Pharmacol* 2013;26:235-7.
- Buyuktiryaki B, Sahiner UM, Karabulut E, Cavkaytar O, Tuncer A, Sekerel BE. Optimizing the use of a skin prick test device on children. *Int Arch Allergy Immunol* 2013;162:65-70.
- Werther RL, Choo S, Lee KJ, Poole D, Allen KJ, Tang ML. Variability in skin prick test results performed by multiple operators depends on the device used. *World Allergy Organ J* 2012;5:200-4.

10. Fahrlander BC, Wüthrich B, Sennhauser FH, Vuille JC. Significant variations of skin prick test results between five fieldworkers in a multi-center study. In: Ring J, Behrendt H, Vieluf D, editors. *New Trends in Allergy IV*. 1st ed. Berlin: Springer; 1997. p. 13-7.
11. Bernstein IL, Li JT, Bernstein DI, Hamilton R, Spector SL, Tan R, *et al.* Allergy diagnostic testing: An updated practice parameter. *Ann Allergy Asthma Immunol* 2008;100 3 Suppl 3:S1-148.
12. Heinzerling L, Mari A, Bergmann KC, Bresciani M, Burbach G, Darsow U, *et al.* The skin prick test – European standards. *Clin Transl Allergy* 2013;3:2-10.
13. Australian Society of Clinical Allergy and Immunology. Skin Prick Testing for the Diagnosis of Allergic Disease. A Manual for Practitioners. Available from: <http://www.allergy.org.au/health-professionals/papers/skin-prick-testing>. [Last updated on 2013 Nov 15; Last cited on 2014 Nov 27. Newman BT, Kohn MA, editors. *Reliability and Measurement Error. Evidence Based Diagnosis*. 1st ed. Cambridge: Cambridge University Press; 2009. p. 10-39.
14. Antunes J, Borrego L, Romeira A, Pinto P. Skin prick tests and allergy diagnosis. *Allergol Immunopathol (Madr)* 2009;37:155-64.
15. Wise SL, Meador KJ, Thompson WO, Avery SS, Loring DW, Wray BB. Cerebral lateralization and histamine skin test asymmetries in humans. *Ann Allergy* 1993;70:328-32.
16. Shu SA, Chang C, Leung PS. Common methodologies in the evaluation of food allergy: Pitfalls and prospects of food allergy prevalence studies. *Clin Rev Allergy Immunol* 2014;46:198-210.
17. Wood RA, Sicherer SH, Vickery BP, Jones SM, Liu AH, Fleischer DM, *et al.* The natural history of milk allergy in an observational cohort. *J Allergy Clin Immunol* 2013;131:805-12.
18. Baker A, Empson M, The R, Fitzharris P. Skin testing for immediate hypersensitivity to corticosteroids: A case series and literature review. *Clin Exp Allergy* 2015;45:669-76.
19. Fatteh S, Rekkerth DJ, Hadley JA. Skin prick/puncture testing in North America: A call for standards and consistency. *Allergy Asthma Clin Immunol* 2014;10:2-9.