

Reproducibility of skin prick test results in epidemiologic studies: a comparison of two devices

Illi S, Garcia-Marcos L, Hernando V, Guillen JJ, Liese A, von Mutius E. Reproducibility of skin prick test results in epidemiologic studies: a comparison of two devices. Allergy 1998; 53: 353–358. © Munksgaard 1998.

The reproducibility of skin prick tests under field conditions is essential for comparing prevalences between centers in epidemiologic multicenter studies. This study aimed to evaluate and compare the reproducibility of two widely used skin prick test devices: the Multi-Test and the ALK lancet. The subjects were 28 children, aged 6–14 years, with known sensitivities to *Dermatophagoides pteronyssinus* (*D. pter.*). Both devices were applied to each subject on two occasions, 1 week apart, by different, randomly assigned fieldworkers, using histamine, negative control, and the *D. pter.* allergen extract. For all three tested solutions, mean wheal sizes were larger for the Multi-Test than for the ALK lancet. The coefficient of variation for histamine was 21.8% for the Multi-Test and 17.3% for the ALK lancet. The coefficients of variation for the allergen *D. pter.* amounted to 47.4% for the Multi-Test and to 24.6% for the ALK lancet. The percentage of concordant test results was 92.6% for the Multi-Test and 100.0% for the ALK lancet for a cutoff point of wheal size equal to or greater than 1 mm. The results of this study suggest that the single ALK lancet performs slightly better than the Multi-Test device with respect to reproducibility under conditions of epidemiologic field studies.

**S. Illi¹, L. Garcia-Marcos²,
V. Hernando², J. J. Guillen²,
A. Liese³, E. von Mutius¹**

¹University Children's Hospital Munich, Germany; ²Unidad de Investigacion, Cartagena, Spain; ³Institute of Epidemiology, Münster, Germany

Key words: reproducibility; skin prick test.

Sabina Illi
Universitätskinderklinik
Lindwurmstr. 4
D-80337 München
Germany

Accepted for publication 2 December 1997

In contrast to clinical diagnostic applications of allergy tests, one aim of epidemiologic research is the comparison of allergy prevalences between different regions and countries. However, differences in the prevalence of atopy between different countries, as assessed by skin test reading, may not only depict actual differences in prevalences, but may also be attributable to differences in the skin prick test methods used and the cutoff point chosen for a positive result. In epidemiologic research, particularly in multicenter studies, reproducibility of skin prick test results among fieldworkers is one of the foremost points of interest. Hence, standardized procedures and uniform testing methods are of major importance to ensure comparable results.

Most previous studies of the reproducibility of skin prick tests did not consider the aspects of epidemiologic fieldwork; i.e., prick tests were applied by physicians or nurses and not by field-

workers (1–7). Furthermore, in most studies, the prick tests were applied by a single investigator (1, 3–5, 7, 8). However, studies that did take the effects of fieldwork conditions into account (9, 10) have shown that, in spite of careful training and standardized test protocols, systematic differences in skin prick test results among fieldworkers may occur, introducing substantial bias into comparisons of allergen sensitization rates. Several methods for skin prick testing are available, each with characteristic advantages and disadvantages. The European Academy of Allergology and Clinical Immunology (EAACI) (11) recommends skin prick test methods based on single lancets, as opposed to multitesting devices. However, multitests have the advantage of quick and easy application. This study aimed to evaluate and compare two widely used skin prick testing devices, a multitest device and a single lancet, with respect to their reproducibility under field conditions.

Material and methods

Skin prick test devices and allergen extracts

Skin prick tests were performed with the single ALK lancet (ALK, Denmark) and the Multi-Test device (Center Laboratories, USA). The ALK lancet is a disposable individual needle, and the Multi-Test applicator is a disposable plastic device for the simultaneous prick puncture administration of eight different antigens. The solutions used for testing were histamine phosphate (10 mg/ml histamine base) in 50% glycerosaline and *Dermatophagoides pteronyssinus* (*D. pter.*). The negative control was 50% glycerosaline. The solutions used for the ALK lancet and for Multi-Test were identical, all provided by ALK.

Study population

The study population was recruited from the Out-patient Allergy Clinic in Cartagena, Spain. All patients aged 6–14 years whose parents had given informed consent and who had asthma and known sensitivities to *D. pter.*, defined as previous 2+ or 3+ skin test reactions to mites, diagnosed with the use of various skin prick devices and allergen extracts, were included in the study. Exclusion criteria were the use of Astemizol in the last 4 weeks before testing and the use of other antihistamines or oral steroids in the week before testing, as well as eczematous skin lesions at the site of testing. This resulted in a sample size of 28 children.

Study design

Four fieldworkers, without previous working experience in an allergy laboratory, were carefully trained according to a standardized test protocol. According to this protocol, each fieldworker had to perform at least three series of 16 skin prick tests with histamine until the coefficient of variation of a series dropped below 20%, a limit recommended by the EAACI (11). All children were tested on two occasions, 1 week apart. The tests on these two occasions were applied in an exactly identical way, but by a different, randomly assigned fieldworker for each occasion. Skin prick tests were applied in the afternoon (1400–1600) on the volar surface of both forearms, 3 cm distal from the elbow crease (11). The Multi-Test device was used on one arm, and the ALK lancet on the other. On the first occasion, the two skin test methods were randomly assigned to the right and left arms, respectively. The allocation of the skin prick devices to the left or right arm remained unchanged on the second occasion. Three different extracts were tested: histamine, negative control, and *D. pter.* To ensure

comparability, the distances between application sites of extracts for the ALK lancet were chosen in accordance with the respective distances of the Multi-Test device, hence measuring 2.5 cm in the transversal and 1.5 cm in the longitudinal axis. Extracts were applied on each arm in the following order (proximal to distal):

Histamine	Negative control
Negative control	Histamine
<i>D. pter.</i>	<i>D. pter.</i>

The tests were read after 10 min for histamine and after 15 min for the allergen extract and negative control (2, 5). All wheal sizes were measured as the mean of the longest diameter (a) (excluding pseudopodia) and the midpoint perpendicular diameter (b); i.e., $(a+b)/2$, measured to the nearest 1 mm. This measurement of the wheal sizes was done in two ways. For the first measurement (reading 1), the wheal size of each extract was measured directly on the site by a transparent, flexible ruler. For the second measurement (reading 2), the wheal was first circled with a thin, felt-tip pen. This circle was then transferred to a transparent tape attached to the arm. The wheal size was then measured on this transparent tape.

Statistical analysis

In a first step, the mean, the standard deviation, and, as a measure of reproducibility, the coefficient of variation ($CV = \text{standard deviation}/\text{mean} \times 100$) of wheal reactions to negative control, histamine, and the allergen extract *D. pter.* were computed for each skin prick test method, and this was combined for the two testing occasions. This was done separately for the two different ways of measuring the wheal size. The wheal sizes resulting from the two prick methods were compared for each of the three extracts with the nonparametric Wilcoxon rank-sum test. Additionally, the results of the four fieldworkers were compared with the nonparametric Kruskal-Wallis test in order to detect a possible fieldworker effect.

In a second step, the concordance of test results for the allergen *D. pter.* between occasions was analyzed. First, the mean wheal size of the two pricks per child, needle, occasion, and test substance was calculated. Three dichotomous variables were then introduced, by defining a positive reaction to the allergen extract as a mean wheal size of ≥ 1 , ≥ 2 , and ≥ 3 mm, the last usually being applied in clinical as well as epidemiologic practice. Then, in order to adjust the wheal size intraindividually, the mean reaction per child, needle, and occasion for the negative control was subtracted from the corresponding

Table 1. Study population: number, age, and sex of children and number of applied skin pricks

	Total	Prick test method	
		Multi-Test	ALK lancet
No. of children	28	27	17
Age (years): mean $\pm$ standard deviation	10.0 $\pm$ 3.6	9.9 $\pm$ 3.5	10.2 $\pm$ 3.6
Female: male	19:19	19:18	3:14
No. of skin pricks	528	324	204

mean wheal size of *D. pter.* The result was called "standardized mean wheal size". Analogously to the mean wheal size, three dichotomous variables were introduced, by defining a positive reaction to *D. pter.* as standardized mean wheal sizes ≥ 1 , ≥ 2 , and ≥ 3 mm. For all six of these dichotomous variables, the proportion of concordant test results among fieldworkers in weeks 1 and 2 was computed as the sum of the number of positive reactions on both occasions plus the number of negative reactions on both occasions divided by the total number of paired tests.

Results

Of the 28 children, 16 were tested with the Multi-Test device as well as with the ALK lancet, 11 were tested only with the Multi-Test device, and one child was tested only with the ALK lancet (Table 1). For both prick test methods, each solution was tested on two occasions per child and twice per occasion, leading to a total of 528 skin pricks.

The results shown in Tables 1 and 2 are those of reading 2. The results of reading 1 yielded smaller mean wheal sizes, but did not differ significantly from those of reading 2 with respect to the coefficient of variation, the differences in fieldworker results, and the number of concordances between tests. Nor were any differences seen between sexes.

Table 2 shows the number of performed skin prick tests, and the mean, the standard deviation, and the coefficient of variation of the three solutions for both skin prick test devices. The skin pricks applied by the Multi-Test device resulted in significantly larger mean wheal sizes for two of the three solutions tested; namely, negative control ($P \leq 0.01$) and histamine ($P \leq 0.001$). The coefficient of variation not only varied between devices, but was also dependent on the corresponding solution used. For both test devices, the coefficients of variation were lowest for histamine and highest for the negative control, with *D. pter.* ranging in the middle. The coefficient of variation for histamine for the ALK lancet (CV=17.3%) met the criteria recommended in the literature (11); i.e., CV $\leq$ 20%.

The corresponding coefficient of variation for the Multi-Test (CV=21.8%) was slightly above this recommendation. However, a different picture appeared if the reactions to the allergen extract were considered. Both test devices resulted in higher coefficients of variation for *D. pter.* The ALK lancet exceeded this limit only slightly (CV=24.6%), whereas the Multi-Test device attained a CV of 47.4%.

To assess the fieldworker effect on differences in prick test results, we compared the fieldworker results with respect to measured wheal size. For the Multi-Test device, two of the three solutions resulted in significantly different fieldworker results; namely, negative control and histamine ($P \leq 0.05$). For the ALK lancet, all three solutions differed significantly among fieldworkers; *D. pter.*: $P \leq 0.05$; negative control and histamine: $P \leq 0.01$ (Table 3).

Table 4 shows the percentage of concordant results for the allergen extract *D. pter.* for each of the two skin prick test methods, according to the chosen cutoff point. The proportion of concordant results is shown for the standardized mean wheal size and for the actual mean wheal size. For the Multi-Test device, as well as for the ALK lancet, the lowest cutoff point achieved the highest percentage of concordances, and vice versa. This held for the standardized mean wheal size as well as for the actual mean wheal size. However, the ALK lancet showed, for both mean wheal size and standardized wheal size, a higher percentage of concordances than the Multi-Test for all three of the chosen cutoff points, varying between 100.0% and 58.8% for the standardized wheal size and between 100.0% and 88.2% for the mean wheal size. The Multi-Test varied between 88.9% and 57.4% for the standardized wheal size and between 92.6% and 77.8% for the mean wheal size.

Table 2. Mean wheal sizes, standard deviation (SD), and coefficients of variation (CV)

Test substance	Wheal size	Prick test method	
		Multi-Test	ALK lancet
Negative control	N	108	68
	Mean (mm)**	1.24	0.60
	SD (mm)	1.55	1.12
	CV (%)	124.38	186.01
Histamine	N	108	68
	Mean (mm)***	6.21	5.64
	SD (mm)	1.35	0.97
	CV (%)	21.83	17.25
<i>D. pter.</i>	N	108	68
	Mean (mm)	4.76	4.58
	SD (mm)	2.26	1.13
	CV (%)	47.36	24.63

** $P \leq 0.01$, *** $P \leq 0.001$; Wilcoxon rank-sum test.

Table 3. Analysis of fieldworker effect: number of pricks, mean wheal size, and standard deviation for four fieldworkers

Prick test method	Test substance	Fieldworker			
		1	2	3	4
Multi-Test	Negative control*	1.11±1.41 (n=32)	1.76±1.57 (n=36)	0.57±1.47 (n=22)	1.28±1.59 (n=18)
	Histamine*	6.66±1.25 (n=32)	6.19±1.62 (n=36)	5.84±0.99 (n=22)	5.89±1.19 (n=18)
	<i>D. pter.</i>	5.23±1.47 (n=32)	5.06±2.96 (n=36)	3.77±2.03 (n=22)	4.56±1.75 (n=18)
ALK lancet	Negative control*	0.37±0.79 (n=30)	1.25±1.53 (n=22)	0.00±0.00 (n=6)	0.25±0.54 (n=10)
	Histamine**	6.07±0.91 (n=30)	5.45±1.04 (n=22)	5.42±0.38 (n=6)	4.90±0.66 (n=10)
	<i>D. pter.</i> *	4.95±1.02 (n=30)	4.25±1.34 (n=22)	4.83±0.60 (n=6)	4.05±0.80 (n=10)

* $P \leq 0.05$, ** $P \leq 0.01$; Kruskal-Wallis test.

When asked which of the two devices they preferred, children and parents, as well as fieldworkers, inclined to Multi-Test rather than to the ALK lancet, since the former was less uncomfortable and easier and quicker to apply.

Discussion

In epidemiologic multicenter studies of the prevalence of atopy, one major aim is the comparison of prevalences among centers. However, comparability is ensured only if the criteria for a positive diagnosis of atopy are well defined and if the diagnosis itself is reproducible.

Several studies have investigated the reproducibility (2, 5, 7) and precision of the Multi-Test device (4, 12), either as a comparison with other skin prick test devices (4, 5, 7, 12) or as an evaluation of the Multi-Test by itself (2). Most of these studies concluded that the Multi-Test device is a highly reproducible device (2, 4, 5, 7). However, none of these studies were performed under field conditions; i.e., the skin pricks were applied by trained staff and mostly by a single physician or nurse. Therefore, the main aim of this study was to

compare the skin prick test device Multi-Test, which enables simultaneous testing of eight allergen extracts, with the widely used and recommended ALK lancet for single skin pricks with respect to their reproducibility under the usual field conditions of epidemiologic work.

Compared to previous studies (1–4, 7, 8, 12) of the reproducibility and precision of various skin prick test devices, our study resulted in similar coefficients of variation for histamine and *D. pter.* for both the ALK lancet and Multi-Test, although those achieved by the ALK lancet in our study were slightly smaller than those achieved by Multi-Test. As in these other studies, the coefficients of variation for *D. pter.* were higher than those for histamine, since the natural variability of sensitization to *D. pter.* of the children in the study adds up to the variability of the testing device. With respect to the high variation for the negative control, it has to be taken into account that it is extremely difficult to distinguish between wheal sizes as small as 1 mm or even smaller. Hence, the variation for the negative control not only depicts the actual variability of the device used for testing and of the children's reactions to the negative

Table 4. Concordance of test results (in percentage) between first and second test for *D. pter.* according to cutoff point for positive reaction

Cutoff point	Prick test method	
	Multi-Test	ALK lancet
	Percentage of concordant test results	Percentage of concordant test results
Standardized mean wheal ≥ 1 mm	88.9%	100.0%
Standardized mean wheal ≥ 2 mm	74.1%	88.2%
Standardized mean wheal ≥ 3 mm	51.9%	58.8%
Mean wheal ≥ 1 mm	92.6%	100.0%
Mean wheal ≥ 2 mm	88.9%	100.0%
Mean wheal ≥ 3 mm	77.8%	88.2%

control, but also of the ability of the fieldworkers to measure these small wheal sizes precisely. However, this high variation for negative control holds for both skin prick methods in our study.

Another important aspect to consider in epidemiologic fieldwork is the fieldworker effect. In an epidemiologic study of childhood allergy, Braun-Fahrlander et al. (9) showed that in spite of careful training and standardized test protocols, differences in skin prick test results between fieldworkers may occur. This held for our study as well. The skin prick tests applied by the four fieldworkers resulted in small, but statistically significant differences in their outcome. This was true for the ALK lancet as well as for the Multi-Test device. The differences remained significant even when fieldworker 3, who achieved a mean wheal size and standard deviation of 0 mm for the negative control, was excluded from the analysis.

Previous papers on the performance of skin prick tests analyzed only the reproducibility of wheal sizes. Of equal importance, however, is the reproducibility of results defined as "positive" or "negative". In our study, different definitions of a positive test result for *D. pter.* were used. First, we considered the actual wheal size. Second, for consideration of nonspecific reactions caused by the trauma to the skin, the wheal size caused by the allergen was set in relation to the reaction induced by the negative control solution. An appropriate cutoff point had then to be chosen for the actual wheal size as well as the standardized wheal size in order to evaluate results categorically by concordance rates. We studied three different cutoff points with respect to the resulting concordances of positive and negative test results. The two skin prick test methods achieved different percentages of concordant results. The ALK lancet achieved better results for all three of the chosen cutoff points for either way of defining the wheal size; i.e., actual wheal size as well as standardized wheal size. However, our results show for both skin prick methods that, with respect to conditions arising in fieldwork, a positive definition based on a cutoff point of 1 mm might be preferred when comparing prevalences of atopy among centers. This agrees with the results of Chinn et al. (10), who reported results of skin prick tests performed within the European Community Respiratory Health Study (ERHS), an epidemiologic multicenter study on respiratory health. In this study, allergen-coated lancets (Phazets) were used for skin prick testing and the results were compared to the corresponding specific IgE values. The authors concluded that skin prick tests should be reported as positive if greater than zero, since this cutoff for a positive result gave a more consistent relation with the

corresponding specific IgE value among fieldworkers than a cutoff of 3 mm. However, it is important to note that if a detection of clinically relevant atopy with respect to the relation between allergen sensitivities and respiratory illness is of major interest, a larger cutoff point should be chosen. Peat & Woolcock (13), who tested 13 common allergens, reported that children with one or more positive skin wheals of greater than or equal to 3 mm had significantly more recent wheeze, hay fever, eczema, and bronchial hyperresponsiveness than children with smaller wheals. Furthermore, Chinn et al. (10) found no consistent relation between allergen wheal diameter and histamine wheal diameter, and thus concluded that one should not adjust allergen wheal to histamine wheal. Our results with respect to adjusting for the size of the negative control wheal were similar. The proportion of concordant results was higher for all chosen cutoff points if the unadjusted wheal sizes were considered.

In conclusion, the results of this study suggest that the ALK lancet performs slightly better than the Multi-Test device in reproducibility under field conditions. However, the choice of one or the other device has to be made after consideration of additional aspects such as easy applicability, cost, availability, etc. However, all centers participating in a study must use an identical skin prick test method if results are to be compared among the centers.

Acknowledgment

This work was partly supported by Center Laboratories, USA.

References

1. Basomba A, Sastre A, Palaez A, Romar A, Campos A, Garcia-Villalmanzo A. Standardization of the prick test: a comparative study of three methods. *Allergy* 1985;40:395-9.
2. Berkowitz RB, Tinkelman DG, Lutz C, et al. Evaluation of the Multi-Test device for immediate hypersensitivity skin testing. *J Allergy Clin Immunol* 1992;90:979-85.
3. Demoly P, Bousquet J, Manderscheid JC, et al. Precision of skin prick and puncture tests with nine methods. *J Allergy Clin Immunol* 1991;88:758-62.
4. Engler DB, De Jarnatt AC, Sim TC, et al. Comparison of the sensitivity and precision of four skin test devices. *J Allergy Clin Immunol* 1992;90:985-91.
5. Mahan C, Spector SL, Siegel SC, et al. Validity and reproducibility of Multi-Test skin test device. *Ann Allergy* 1993;71:25-8.
6. Menardo JL, Bousquet J, Michel FB. Comparison of three prick test methods with the intradermal test and the RAST in the diagnosis of mite allergy. *Ann Allergy* 1982;48:235-9.
7. Nelson HS, Rosloniec DM, McCall LL, et al. Comparative performance of five commercial prick skin test devices. *J Allergy Clin Immunol* 1993;92:750-6.
8. Perrin LF, Dechamp CH, Deviller P, Joly P. Reproducibility

- of skin tests: a comparative study of the Pepys prick test and the Morrow-Brown needle and their correlation with the serum IgE level. *Clin Allergy* 1984;14:581-8.
9. Braun-Fahrländer Ch, Sennhauser FH, Wüthrich B. Significant variations of skin prick test results between five fieldworkers in a multi-center study. *Am J Respir Crit Care Med* 1995;151:A565.
10. Chinn S, Jarvis D, Luczinska C, Lai E, Burney P. Measuring atopy in a multi-center epidemiological study. *Eur J Epidemiol* 1996;12:155-62.
11. Dreborg S, Backman A, Basomba A, Bousquet J, Dieges P, Malling H-J. Skin tests used in type I allergy testing. Position paper of the European Academy of Allergology and Clinical Immunology. *Allergy* 1989;44 Suppl 10:1-59.
12. Adinoff AD, Rosloniec DM, McCall LL, Nelson HS. A comparison of six epicutaneous devices in hypersensitivity skin testing. *J Allergy Clin Immunol* 1989;84:168-74.
13. Peat JK, Woolcock AJ. Sensitivity to common allergens: relation to respiratory symptoms and bronchial hyper-responsiveness in children from three different climatic areas of Australia. *Clin Exp Allergy* 1991;21:573-81.

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.