Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY

EAACI

Allergy

ORIGINAL ARTICLE EXPERIMENTAL ALLERGY AND IMMUNOLOGY

Comparison of five techniques of skin prick tests used routinely in Europe

M. S. Masse[1], A. Granger Vallée[2], A. Chiriac[1], H. Dhivert-Donnadieu[1], L. Bousquet-Rouanet[1], P.-J. Bousquet[1] & P. Demoly[1]

[1]Département de Pneumologie, Unité d'Exploration des allergies et INSERM U657, Hôpital Arnaud de Villeneuve; [2]Département de Néphrologie, Hôpital Lapeyronie, Montpellier Cedex, France

To cite this article: Masse MS, Granger Vallée A, Chiriac A, Dhivert-Donnadieu H, Bousquet-Rouanet L, Bousquet P-J, Demoly P. Comparison of five techniques of skin prick tests used routinely in Europe. *Allergy* 2011; **66**: 1415–1419.

Keywords
aeroallergens; device; diagnosis; lancet; sensitization; skin prick tests.

Correspondence
Pascal Demoly, CHRU de Montpellier – Hopital Arnaud de Villeneuve, 371 av. Du doyen Gaston Giraud, 34295 Montpellier Cedex 5, France.
Tel.: +33(0)467336127
Fax: +33(0)467042708
E-mail: pascal.demoly@inserm.fr

Accepted for publication 5 July 2011

DOI:10.1111/j.1398-9995.2011.02679.x

Edited by: Thomas Bieber

Abstract

Background: Skin prick tests represent indispensable tools in allergy, even more than 30 years after their introduction in clinical practice.
Objectives: Few recent European studies have focused on this topic and we thus wanted to compare the instruments most often used today.
Methods: Four instruments were investigated: the 23G intravenous (IV) needle, the ALK Lancet, the Stallergenes (STG) Prick Lancet and the Stallerpoint® (using two different methods). Sensitivity, reproducibility, and acceptability were evaluated. In 22 subjects, we calculated the sensitivity and reproducibility (both intra- and interpatient) of these methods by testing the positive control five times. In 50 subjects, we tested the single-blind acceptability of these same five techniques.
Results: In terms of sensitivity, the IV needle (100%) and metal lancets (96% for the ALK Lancet and 98% for the STG Prick Lancet) were superior ($P < 0.01$) to the two Stallerpoint® methods (20% and 57%). Intrapatient reproducibility was 16.2%, 14.6%, 15.0%, 97.1% and 18.1%, respectively. The instruments that were best tolerated by the patients were the IV needle and the two metal lancets.
Conclusion: Metal needles and/or lancets are the tools of choice for skin prick testing.

Background

Spectacular advances in the diagnosis of allergy have occurred in recent years, especially with the advent of recombinant allergens and determination of their blood-specific IgE. However, skin testing still remains the tool most often used by clinicians working in the allergy field, and its many advantages (ease of use, low cost, rapidity, and visibility of the results by the patient) render it unlikely to be replaced in the future (1). Furthermore, its sensitivity and specificity have not been surpassed by any other test. Over the years, many different techniques and instruments have been marketed to perform skin tests. Some techniques have been abandoned, including scratch tests, primarily owing to their low reproducibility and painful nature. On the contrary, others have been proven useful and continue to be used even more than 30 years after their introduction in practice. This is the case of needles [with the modified prick test technique suggested by Pepys in 1975 (2)] and metal lancets [introduced in the late 70s (3)].

In the medical literature of the 90s, there are several studies that compared various tools available to perform skin prick tests (4–9). In contrast, only a few quality studies have been published on this subject in the recent years (10, 11). Moreover, most of them were made in the US and do not apply to the European population, given that most of the tools studied are not marketed or distributed in Europe. The reverse is also true, namely that certain instruments are used almost exclusively in Europe, as is the case of the Stallerpoint®, which has been evaluated in only one study (4).

Consequently, we designed this study to evaluate the sensitivity, reproducibility, and acceptability of the most frequently used methods to perform skin prick tests in Europe.

Methodology

Sensitivity and reproducibility

Subjects

For the first part of the study, assessing the sensitivity and reproducibility of five different techniques of skin prick test in the same patient, we limited our selection to adults over 18 years old so as not to cause such inconvenience in children. In addition, patients had to have ceased taking antihistaminic drugs at least 8 days prior to the experiment and had to be exempt of dermographism and needle phobia. However, no selection was made as to the atopic status of the patient. Patients had to provide their free and informed consent after being taught about the risks and benefits of the study by the investigator. In total, 22 adults were included in this part of the study.

Materials and methods

We compared the reproducibility and sensitivity of four instruments routinely employed today in Europe to perform skin prick tests [intravenous (IV) needle, ALK Lancet, Stallergenes (STG) Prick Lancet and Stallerpoint®] (Fig. 1) with the use of the positive control codeine phosphate 9% (Stallergenes SA, Antony, France) for each technique.

Concerning the tests with the IV needle, we used 23 G needles 1″ (BD MicrolanceTM 3, TM3, Fraga, Spain) and the modified prick test method as suggested by Pepys (2). In short, this technique consists in crossing the drop, bevel up, before making a slight infringement of the skin without causing bleeding. Regarding the other three instruments, we crossed the drop being perpendicular to the skin and applied moderate pressure (corresponding to a depression of the skin of about 2–3 mm). We used this technique, in accordance with the manufacturer recommendations, to test the ALK Lancet (ALK-Abello, Horsholm, Denmark), the STG Prick Lancet (Stallergenes SA) and the Stallerpoint® (Stallergenes SA). However, because of the low sensitivity already demonstrated with the Stallerpoint® according to the method previously described (personal data), we also investigated it using a second technique. The latter consisted in applying a vertical pressure, using the technique described earlier, followed by a 90° rotation of the instrument's axis while maintaining the depression of the skin (Stallerpoint® 90).

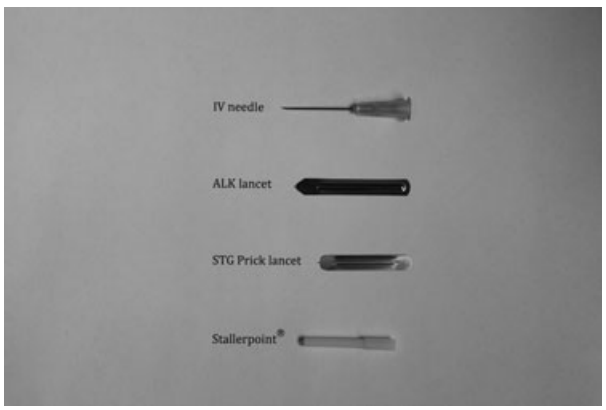

Figure 1 Skin prick test instruments evaluated in the present study.

In total, we realized five skin prick tests to 9% codeine phosphate (CP) with each technique in every subject (25 prick tests per subject). All tests were performed on the patients' forearms by two experimented investigators, taking care to distance each test by at least 3 cm. A different sterile instrument was used for each prick test. The drops were wiped off 1 min after the prick test was performed, and the wheals were measured by the same technician who had previously performed the tests and after a period of 10 min. To do this, each test was first encircled with a pen and then transferred to a permanent record with transparent tape. The longest diameter (D1) was first measured, followed by the perpendicular diameter passing through its middle (D2). The mean diameter (MD) was then calculated using the formula: MD = (D1 + D2)/2. The latter was used for final analysis of the results.

Acceptability

Subjects

For the second part of the study, assessing blindly the pain inflicted by five different techniques of skin prick tests, we decided to include adults and children over 10 years, given the small number of tests performed in each subject. In total, 50 subjects aged 11–80 years were included in this portion of the study. Once more, a free and informed agreement was previously obtained from each subject (including parents of children) by the investigator.

Materials and methods

We performed the five techniques of prick tests (IV needle, ALK Lancet, STG Prick Lancet, Stallerpoint® and Stallerpoint® 90) using the same procedural rules as described earlier. However, this part of the study was conducted without PC to assess only the pain caused by the instruments themselves and not the one who might have been caused by the PC. The patient was instructed to keep his or her eyes closed during experimentation (single blind) and the order with which the techniques were tested was determined randomly for each subject. After each manipulation, the subject was asked to score the pain experienced from 0 to 10, according to a numerical scale of pain: 0 representing no pain; 5, a moderate pain; and 10, the worst pain imaginable (12–14).

Statistical analysis

To calculate the sensitivity of each technique, we used a positive threshold of 3 mm for the mean diameter (MD) of each wheal. Sensitivity was calculated as equal to the ratio between the number of true-positive tests and the sum of true-positive and false-negative tests.

To assess reproducibility in a same patient (intrapatient reproducibility), we calculated the coefficient of variation

(CV) between the MD of the five papules induced by the same technique using the following formula:

$$CV_{intrapatient} = SD_{intrapatient}/\mu_{intrapatient}$$

where SD and μ are, respectively, the standard deviation and the average of the MD of the five papules.

Finally, we were also interested in the variation between the average size of the wheals induced by the same technique from one patient to another (interpatient reproducibility). To do this, we calculated the interpatient CV using the following formula:

$$CV_{interpatient} = SD_{interpatient}/\mu_{interpatient}.$$

We used nonparametric descriptive statistics (median, inter-quartile range) because of the non-normal distribution of the data (Shapiro–Wilk test). Friedman tests followed by Dunn's post-tests were used to compare the five techniques according to their $CV_{intrapatient}$, mean size of papules and pain ratings. Sensitivities, for their part, were compared using Fisher's exact test. *P* values below 0.05 were considered statistically significant.

Results

Sensitivity of the different instruments

In terms of sensitivity (Table 1), the IV needle and the two metal lancets have yielded excellent and equivalent results: 100% for the IV needle, 96% for the ALK Lancet and 98% for the STG Prick Lancet ($P > 0.05$). The Stallerpoint® has revealed itself less sensitive (20%) than the other instruments ($P < 0.01$). While the Stallerpoint® 90 permitted to improve the sensitivity of the instrument (57% *vs* 20%, $P < 0.01$), it nevertheless remained statistically less sensitive than the IV needle and metal lancets ($P < 0.01$).

Mean size of the wheals induced by the different instruments

The mean wheal sizes (average of the five MDs) induced by each technique are presented in Table 2 and Fig. 2. The IV needle (median 5.4 mm) induced the largest mean wheals ($P < 0.001$ with both the Stallerpoint® and Stallerpoint® 90, $P < 0.05$ with the ALK Lancet, $P > 0.05$ with the STG Prick Lancet). The Stallerpoint® gave significantly smaller papules than the other instruments (median 1.4 mm, $P < 0.001$ with IV needle and metal lancets), whereas the addition of a quarter turn (Stallerpoint® 90) tended to yield larger papules (median 2.8 mm, $P > 0.05$).

Table 1 Sensitivities of different techniques of skin prick tests

Technique	True-positives, *n*	False-negatives, *n*	Sensitivity, %
Intravenous needle	110	0	100
ALK Lancet	106	4	96
Stallergenes Prick Lancet	108	2	98
Stallerpoint®	22	88	20
Stallerpoint® 90	63	47	57

Table 2 Skin prick test results: mean size of wheals and reproducibility

	Mean papule sizes, mm		Coefficient of variation, %	
Technique	Median	IQR	Intrapatient	Interpatient
Intravenous needle	5.4	4.3–6.3	16.2	21.3
ALK Lancet	4.3	4.0–4.7	14.6	13.0
Stallergenes Prick Lancet	4.9	4.2–5.7	15.0	16.4
Stallerpoint®	1.4	0.7–2.2	97.1	79.9
Stallerpoint® 90	2.8	2.6–3.5	18.1	24.7

IQR, interquartile range.

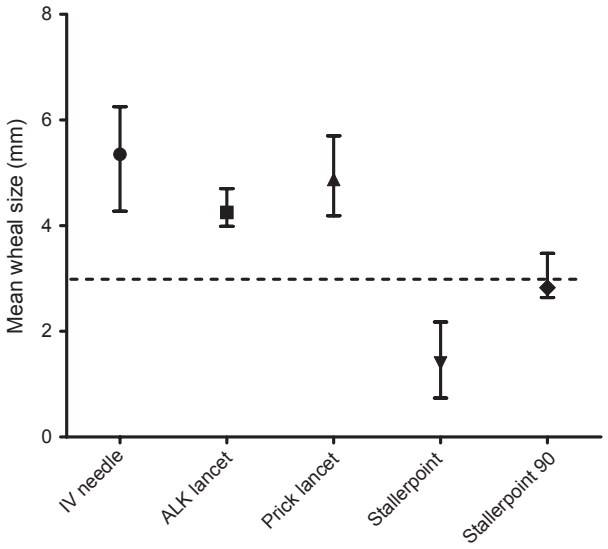

Figure 2 Mean wheal sizes (average of the five mean diameters) produced by the different skin prick techniques. Symbols represent the median and the whiskers, the interquartile range. The dashed line represents the positivity threshold used (3 mm).

Reproducibility of skin prick tests

The results are presented in Table 2. Regarding the intrapatient reproducibility, the table shows that the three instruments that gave the best median $CV_{intrapatient}$ were the IV needle ($CV_{intrapatient}$ 16.2%), the ALK Lancet ($CV_{intrapatient}$ 14.6%) and the STG Prick Lancet ($CV_{intrapatient}$ 15.0%). There was no statistically significant difference between them. The Stallerpoint® displayed poor reproducibility with a median intrapatient CV of 97.1%, being statistically significantly worse than the three other instruments ($P < 0.05$ with the IV needle and $P < 0.001$ with the ALK Lancet and STG Prick Lancet). Then again, we obtained a better result than

Table 3 Pain evaluation (numeric scale of pain)

Technique	Median	Interquartile range
Intravenous needle	1.0	0.0–2.0
ALK Lancet	1.0	0.8–2.3
Stallergenes Prick Lancet	1.0	1.0–3.0
Stallerpoint®	2.0	1.0–3.0
Stallerpoint® 90	4.0	2.0–5.0

that associated with the standard technique when using the Stallerpoint® 90 ($CV_{intrapatient}$ 18.1%, $P < 0.05$).

In regard to the interpatient reproducibility, the results were in essence the same: metal lancets were highly reproducible from one subject to another, with interpatient coefficients of variation of 13.0% and 16.4% for the ALK Lancet and the STG Prick Lancet, respectively. The IV needle ($CV_{interpatient}$ 21.3%) and Stallerpoint® 90 ($CV_{interpatient}$ 24.7%) also gave acceptable results. Once more, the Stallerpoint® using the standard method appeared less performing than the other techniques with a $CV_{interpatient}$ of 79.9%.

Acceptability

Analyses of the results for the second part of our study, assessing the pain induced by each of these techniques (compared blindly in the same patient), are presented in Table 3. The instruments that were best tolerated by the patients were the IV needle and the two metal lancets (median pain score of 1.0 for the three techniques, P = NS). Regarding the Stallerpoint® (median pain score of 2.0), it was more painful than the IV needle ($P < 0.05$), but the difference did not reach statistical significance when compared to the two metal lancets. Concerning the Stallerpoint® 90, it was by far the most painful technique (median pain score of 4.0), being ranked higher on the pain scale than the other three instruments ($P < 0.001$ with each) and the Stallerpoint® without the quarter-turn technique ($P < 0.01$).

Discussion

Daily diagnostic and therapeutic allergy decisions are dictated by the results of skin tests: be it a food to be avoided, environmental changes to be adopted during a respiratory allergy or an anaphylaxis or even an allergen immunotherapy to be undertaken. These decisions are of paramount importance, and sometimes the vital prognosis of a patient may depend on them. We therefore require both accurate and well-tolerated instruments to perform these skin tests. Moreover, allergists have a duty to know the limits of the instruments they use; unfortunately, few recent European studies have addressed this issue. This prospective study comparing four of the most frequently used instruments in Europe was therefore aimed to delineate the precision of these instruments and their acceptability (by measuring pain on a numerical scale). Very relevant results have emerged with the following main conclusions.

First of all, the Stallerpoint®, with the method of either simple prick or prick followed by a quarter turn, presents much lower sensitivities than the IV needle or metal lancets (ALK Lancet and STG Prick Lancet), with values of 20% and 57%, respectively. The fixed 10-min delay between prick test performance and reading cannot explain these poor results: no early but evanescent positive results have been noticed by the technicians (who remained with each subject during the whole waiting period), and no late positive results have been witnessed by the examiners (while performing the acceptability part of the protocol, following the reading of the wheals). The allergist must therefore be advised that a number of skin tests performed with the Stallerpoint® may prove to be false-negatives; the clinical approach and the ensuing therapeutic decisions should consequently take that important information into account. On the contrary, both the IV needle and the metal lancets yield very high and virtually equivalent sensitivities (from 96% to 100%).

Moreover, unlike other studies published in this topic, we have gone further by calculating the reproducibility of each technique in the same patient (intrapatient) but also between subjects (interpatient). Regarding this important aspect of performance, the IV needle and the metal lancets gave, again, superior results with the lowest intra- and interpatient coefficients of variation. Concerning the Stallerpoint® (standard), it appeared insufficiently reproducible to be clinically useful (intra- and interpatient CV of 97.1% and 79.9%, respectively).

In the final part of this study, we demonstrated that there were significant differences in the pain produced by the different skin prick test techniques. In fact, the Stallerpoint® 90 was the least acceptable procedure with a median value of 4.0 on the numerical scale of pain. This is clinically significant considering that a value >3.0 usually represents moderate pain, while values ranging from 1.0 to 3.0 are considered equivalent to mild pain. In contrast, the IV needle and the metal lancets were, again, comparable and very well tolerated (median of 1.0 on the numerical scale of pain).

In fact, we have demonstrated the same phenomenon that Carr (10) and Montalvo (5) also noticed in their respective studies: the instruments that caused the greatest diameter of papules (IV needle, metal lancets) were not those most painful but, conversely, the best tolerated. This is contrary, however, to other results that had been published previously (6).

We recognize that this study has some limitations. First, we did not include instruments that have the possibility to perform several skin tests simultaneously (multi-allergen tests). However, their use in Europe is extremely limited. In addition, they have been shown to be more painful, to produce more false-negatives and to be less reproducible (10, 11). Secondly, we tested the five skin prick techniques with the solution of 9% codeine phosphate, and no testing has been realized with allergens, thereby preventing direct exportation of the data to the patient population we evaluate in the clinic. Also, we did not test with saline (negative control), which would have allowed us to measure the specificity of our techniques. Finally, tests were performed by two technicians with extensive experience with the instruments being

evaluated, while the skin prick tests performed in the clinic are sometimes realized by less skilled personnel. The external validity of our results could therefore be somewhat lessened.

Despite these limitations, important conclusions can nonetheless be drawn from this study, and it underlines the importance of being aware of the characteristics of each instrument and of teaching the technicians on the proper technique to use.

In conclusion, we have shown, in this prospective study evaluating the sensitivity, reproducibility and acceptability of five of the most widely used techniques of skin prick test in Europe, that all techniques are not equivalent. In fact, the IV needle (with the modified prick technique) and metal lancets (ALK Lancet and STG Prick Lancet) have once again been proven to be the most sensitive, reproducible and acceptable techniques. They can thus continue to occupy a prominent place in allergy practice, even over 30 years after their introduction on the market.

Conflict of interest

None of the authors has any conflict of interest to disclose.

References

1. Demoly P, Bousquet J, Romano A. *In vivo* methods for the study of allergy. In: Adkinson NF Jr, Bochner BS, Busse WW, Holgate ST, Lemanske RF Jr, Simons FER, editors. Middleton's Allergy, Principles & Practice, 7th edn. Philadelphia: Mosby Elsevier Inc. 2009: 1267–1280.
2. Pepys J. Skin testing. *Br J Hosp Med* 1975;**14**:412–417.
3. Osterballe O, Weeke B. A new lancet for skin prick testing. *Allergy* 1979;**34**:209–212.
4. Demoly P, Bousquet J, Manderscheid JC, Dreborg S, Dhivert H, Michel FB. Precision of skin prick and puncture tests with nine methods. *J Allergy Clin Immunol* 1991;**88**:758–762.
5. Montalvo A, Martin S, Mesa A, Cortés C, Rodriguez M, Laso MT. Comparative study of 3 types of lancets for performing prick tests. *Allergol Immunopathol (Madr)* 1996;**24**:58–64.
6. Nelson HS, Rosloniec DM, McCall LI, Iklé D. Comparative performance of five commercial prick skin tests devices. *J Allergy Clin Immunol* 1993;**92**:750–756.
7. Illy 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.
8. Nelson HS, Lahr J, Buchmeier A, McCormick D. Evaluation of devices for skin prick testing. *J Allergy Clin Immunol* 1998;**101**:153–156.
9. Engler DB, DeJarnatt AC, Sim TC, Lee JL, Grant JA. Comparison of the sensitivity and precision of four skin test devices. *J Allergy Clin Immunol* 1992;**90**:985–991.
10. Carr WW, Martin B, Howard RS, Cox L, Borish L, the Immunotherapy Committee of the American Academy of Allergy, Asthma and Immunology. Comparison of test devices for skin testing. *J Allergy Clin Immunol* 2005;**116**:341–346.
11. Nelson HS, Kolehmainen C, Lahr J, Murphy J, Buchmeier A. A comparison of multiheaded devices for allergy skin testing. *J Allergy Clin Immunol* 2004;**113**:1218–1219.
12. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis* 1978;**37**:378–381.
13. Breivik EK, Björnsson GA, Skovlunde E. A comparison of pain rating scales by sampling from clinical trial data. *Clin J Pain* 2000;**16**:22–28.
14. Williamson A, Hoggart B. Pain: a review of three commonly used bread rating scales. *J Clin Nurs* 2005;**14**:798–804.