Clinical aspects of allergic disease

Evaluation of devices for skin prick testing

Harold S. Nelson, MD, Jennie Lahr, BA, Andrea Buchmeier, BA, and David McCormick, MS *Denver, Colo.*

Background: Previous comparisons of devices for percutaneous skin testing have revealed statistically and clinically significant differences, from one device to another, in the size of reactions to histamine and allergen extracts and at negative control sites.
Objective: The objective of this study was to compare the performance of several skin test devices which are either new, modified, or used with a modified technique.
Methods: Twenty subjects were tested five to eight times with each of the devices both to glycerol-saline and to 10 mg/ml histamine base. The devices tested were the MultiTest II, Duo Tip-Test (prick and scarification), Quintest, DermaPik (prick and scarification), and small pox needle.
Results: There were highly significant differences among the devices for the size of the reaction to histamine (mean wheal diameter 4.28 to 8.59 mm, $p < 0.0001$), the standard errors of the wheals to histamine (0.82 to 1.45 mm, $p < 0.05$) and in the mean wheal size with glycerol-saline (0.00 to 2.48 mm, $p < 0.0001$).
Conclusions: Devices for performing skin prick testing vary greatly in several characteristics, including the size of reactions at both positive and negative test sites. Each skin test technician should be tested with the device used in that skin testing laboratory to establish criteria for positive and negative tests. (J Allergy Clin Immunol 1998;101:153-6.)

Key words: *Skin testing, skin test devices*

Previous comparisons of devices for percutaneous skin testing have revealed statistically and clinically significant differences, from one device to another, in the size of reaction to histamine or allergen extract and in the likelihood of a reaction at a negative control site.[1, 2] Manufacturers continue to introduce new devices. The principal impetus is to develop devices that allow simultaneous application of the extract to the skin and penetration of the epidermis. An additional quality sought by some manufacturers is to allow simultaneous application of multiple extracts by attaching several applicators to a rigid frame.

The purpose of this study was to compare the performance of several devices that were either new (Duo Tip Test, Quintest), had been modified (MultiTest), or were used with a modified technique from that which we had previously used in studies.[2, 3] For comparison we used a standard skin prick test method.[3]

From the National Jewish Medical and Research Center, Denver.
Received for publication July 3, 1997; revised Sept. 8, 1997; accepted for publication Sept. 11, 1997.
Reprint requests: Harold S. Nelson, MD, National Jewish Medical and Research Center, 1400 Jackson St., Denver, CO 80206.

0091-6749/98 $5.00 + 0 **1/1/86222**

METHODS

Study design

Each subject was tested with all devices in two sessions. At each session they were tested on one side of the back with glycerol-saline and on the other side with a 10 mg/ml histamine base in glycerol-saline prepared by the National Jewish pharmacy. Five tests were performed on each subject with both glycerol-saline and with histamine for each of the single testing devices and the Quintest, while eight were performed with the MultiTest. Sites were randomly rotated so that each device was tested in an equal distribution from the top to the bottom of the back. Both the Duo Tip-Test and the DermaPik were used both as a scarifier (rotated with light pressure) and as a prick device; the other devices were tested by only one method. Before each session, medications were withheld for the following periods: astemizole, 3 months; other antihistamines, 7 days; antidepressants, 14 days; and histamine H-2 antagonists, 24 hours.

To maintain objectivity the study was conducted with the technician who performed all the tests blinded as to whether the test at a particular site was with glycerol-saline or histamine. The results were recorded by a second technician who had not performed the tests.

Since the results with several of these devices vary with the technique of application, representatives of each of the three companies personally demonstrated the desired technique to the technician who performed the study.

Subjects

Twenty subjects, 18 to 70 years of age, either male or female and with or without allergies, were included in the study.

Devices

The devices tested were the Multi-Test II and Duo Tip-Test (Lincoln Diagnostics; Decatur, Ill.); Quintest and small pox needle (Bayer Allergen Products; Spokane, Wash.); and DermaPik (Greer Laboratories; LaNoir, N.C.) (Fig. 1).

Skin testing

Skin testing with all devices was performed on the back. The sites of the tests were rotated so that each device was tested equally in each section of the back.[2] The Quintest and Multi-Test were tested on the volar surfaces of the forearm as well as on the back.

The skin testing results were recorded at 8 minutes for histamine and at 15 minutes for glycerol-saline by outlining with a felt-tipped pen and transferring the outline to a permanent record with transparent tape. The longest and orthogonal diameters were measured, and the mean diameter was used for analysis.

Statistical analysis

The mean wheal sizes produced by testing with histamine were compared by two-way analysis of variance (ANOVA). Pooled estimates of the variance for each device were computed by removing variance between subjects. To make specific comparisons between the mean wheals produced by histamine

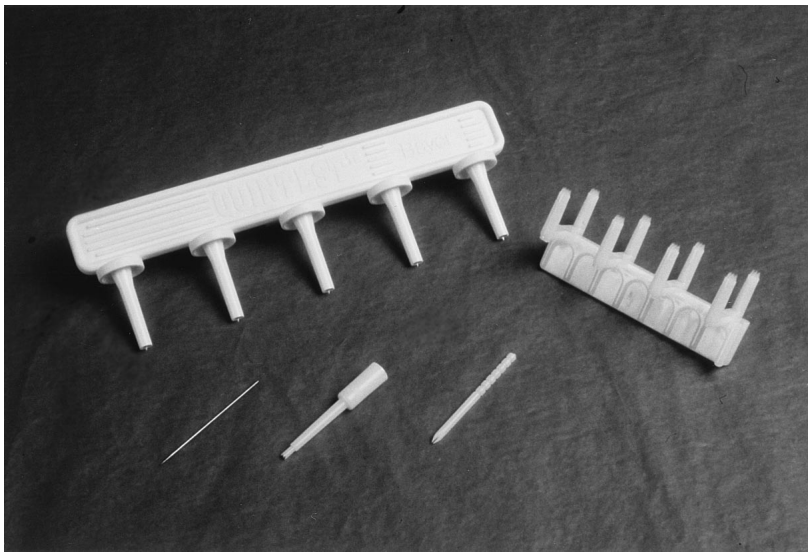

FIG. 1. Shown are the five devices used in the study. Top, left Quintest, right MultiTest. Bottom left to right small pox needle, DermaPik, Duo Tip Test.

with the different devices, Fisher's protected least significant difference multiple comparison procedure was used. Standard deviations were compared between devices with Cochran's test for homogeneity of variances. Specific comparisons were made with the use of Cochran's test under the Fisher's protected procedure.

Pseudo false-negative tests were defined as wheal less than 3 mm mean diameter or an erythema less than 10 mm diameter when another of the same set of tests yielded a wheal of 5 mm or greater and a flare of more than 10 mm. True false-negative tests were deemed present when both the wheal and the erythema failed to reach the prescribed size in the presence of another test site with a wheal of 5 mm or greater. The frequency of false-negative reactions between devices was computed with the use of Fisher's exact test.

Size of reactions at negative test sites were compared by three-way ANOVA, testing effects for subjects, device, and location. Quantiles corresponding to the probabilities of 0.95 and 0.99 were computed for the distribution of sizes of the glycerol-saline solution reactions. Differences resulting from subject and location were ignored, these quantiles estimate wheal size criteria required to achieve 0.95 and 0.99 specificity, respectively. The specificity achieved by use of the standard 3 mm cut-off was also computed for each device.

All tests of hypotheses and multiple comparison procedures were two-tailed at the 0.05 level of significance. All analyses were performed with JMP Version 3.2 software or SAS Version 6.10 software.

RESULTS

Twelve women and eight men participated in the study. Their mean age was 39.4 years, with a range of 29 to 66 years.

The results of the comparative testing are presented in Table I. There were highly significant differences among the devices for the size of the mean wheal with histamine ($p < 0.0001$). The mean histamine wheal from the DermaPik twist was significantly greater than with any other device ($p < 0.0001$). Most of the other differences in histamine wheal size were also significant, although less than for DermaPik. Only the wheals with the small pox needle and the Duo Tip prick and with the DermaPik prick and Duo Tip twist were not significantly different. The MultiTest produced a larger reaction than the Quintest ($p < 0.0001$) and both produced larger wheal size with histamine on the back than on the arm ($p < 0.0001$).

The reproducibility of the histamine reactions expressed as the standard deviation also revealed significant differences among devices. The variability of the MultiTest device was significantly less than that of the smallpox needle or the DuoTip and DermaPik by either method ($p < 0.01$) as well as the Quintest ($p < 0.05$). The Quintest, in turn, had significantly less variability than the DermaPik twist and Duo Tip prick ($p < 0.05$), whereas the small pox needle had significantly less variability than the DermaPik twist ($p < 0.01$), Duo Tip prick ($p < 0.01$), or Duo Tip twist ($p < 0.05$).

Pseudo false-negative reactions, those for which either the wheal was less than 3 mm or the flare was less than 10 cm but not both, occurred with significantly differing frequencies among the various devices ($p < 0.001$) on the back and ($p = 0.028$) on the arm. True false-negatives, in which both wheals and flares were below the threshold size, did not occur with significantly differing frequency on the back or arm ($p = 0.10$ and $p = 0.39$, respectively). There were 24 pseudo false-negative tests on the back, equally divided between wheal less than 3 mm and flare less than 10 mm, whereas there were only three "true" false-negative tests on the back in which both the wheal and flare failed to meet the required size. Two of the "true" false-negatives were with Quintest and one with the DuoTip prick method.

TABLE I. Performance of skin prick test devices

Device	Number of tests	Mean histamine wheal	Standard deviation	Pseudo-false negative (%)	True–false negative (%)	Wheal size specificity of 0.95	Yield specificity of 0.99	Specificity of 3 mm wheal
Back								
QuinTest	100	4.28 mm	1.019	5	2	0	0	1.00
Smallpox	200	4.85 mm	1.067	2	0	0	0	1.00
Duo Tip Prick	100	4.82 mm	1.430	6	1	0	1.5 mm	1.00
Duo Tip Twist	100	6.41 mm	1.301	1	0	3 mm	3.5 mm	0.97
MultiTest II	160	5.71 mm	0.816	6.25	0	3 mm	4 mm	0.95
DermaPik Prick	100	6.2 mm	1.191	0	0	0	3.25 mm	0.98
DermaPik Twist	100	8.59 mm	1.454	0	0	4.5 mm	5 mm	0.35
Arm								
QuinTest	100	3.94 mm	1.139	12	1	0	0	1.00
MultiTest II	160	5.22 mm	0.581	4	0	0	0	1.00

The table presents the results of tests performed on 20 subjects employing the devices and techniques listed in the left column. See Results section of text for discussion of 0.95 and 0.99 specificity and specificity of a 3 mm wheal.

There was a significant difference in the size of the wheals at the glycerol-saline sites among devices ($p < 0.0001$). The DermaPik used as a scarifier produced substantially larger reactions than those produced by the other devices ($p < 0.001$ in each case). The reactions to glycerol saline with Quintest and small pox needle were significantly smaller than to MultiTest and Duo Tip twist, whereas the reactions to Duo Tip prick were significantly smaller than to the Duo Tip twist.

The 95th and 99th quantiles for wheal size at the glycerol saline sites are given for each device in the Table. These quantiles estimate the wheal size required respectively to achieve 0.95 and 0.99 specificity, or "true positive." Ninety-five percent of reactions to glycerol saline were 0 mm by four methods of testing, less than 3 mm by two methods (MultiTest and Duo Tip twist), and less than 4.5 mm with the DermaPik twist. The 99 quantiles reflect the size wheal greater than 99% of the wheals elicited with that device when testing with glycerol saline. For the Quintest and small pox needle, the value is still 0 mm since in this study, neither produced any wheals with glycerol-saline. The other devices all produced some whealing with glycerol saline, hence a specificity of 0.99 required wheals of increasing size: 1.5 mm for DuoTip prick, 3.25 mm for DermaPik prick, 3.5 mm for DuoTip twist, 4 mm for MultiTest, and 5 mm for DermaPik twist.

A 3 mm wheal is often listed as the threshold for a positive skin prick test.[3] For all the devices and methods except the DermaPik used as a scarifier, the specificity of a 3 mm wheal in this study was in the range of 0.95 to 0.99 (Table). However, if 0.99 specificity is desired, the criterion of a 3 mm wheal would be appropriate only with the Quintest, smallpox needle, and Duo Tip used by pricking.

DISCUSSION

As has been true in previous comparisons, this study revealed highly significant differences in the size of skin test reactions obtained with the different devices and techniques studied. The mean wheal diameter with 10 mg/ml histamine base varied from 4.28 mm with the Quintest to 8.59 mm with the DermaPik used as a scarifier. As has been reported previously, devices that produced the smaller wheals with histamine were more apt to give false-negative reactions, whereas those producing the largest wheals with histamine were most apt to produce whealing at the negative control sites.[1, 2]

Different skin prick test devices offer different potential advantages to the user. Thus of the devices tested, only with the smallpox needle may one device be reused (with wiping) to perform all the tests on one patient. On the other hand, all of the other devices tested offer the feature of "dip and apply," so that application of the allergen extract to the skin is accomplished at the same time as the extract is introduced through the epidermis. Additionally, the Quintest and MultiTest allow application of multiple extracts with one operation. Another consideration is the importance of avoiding false-negative reactions versus having negligible reactions at the negative control sites. Two of three devices that provided 0.99 specificity for a "true" positive test with a 3 mm wheal also had a small (1% to 2% percent) incidence of false-negative reactions. Most of the devices that avoided any false-negative reactions would yield several false-positive reactions per 100 tests if the 3 mm criterion was used for a positive test. For the DermaPik twist, clearly a 3 mm wheal criterion would be inappropriate. The importance of each of these features must be considered by the user.

The advantage of using both the wheal and the erythema to judge a marginally positive test can be seen from the difference in rate of "pseudo" false-negative and "true" false-negative reactions. If both a 3 mm mean wheal diameter and a 10 mm flare were required for positive reactions there were 24 (2.8%) "pseudo false-negatives" on the back and 19 (7.5%) on the arm. On the other hand, if either a 3 mm wheal or a 10 mm flare was

accepted as positive, the "true false-negative" rate was only 3 (0.3%) on the back and 1 (0.4%) on the forearm.

The results reported here were achieved by one very experienced and careful technician who had been instructed by the manufacturers in the use of the devices. These results should be viewed as the best achievable. Physicians are encouraged to have their technicians perform similar testing with their device to set criteria for positive and negative skin prick tests in their own clinic.

REFERENCES

1. Adinoff AD, Rosloniec DM, McCall LI, Nelson HS. A comparison of six epicutaneous devices in the performance of immediate hypersensitivity skin testing. J Allergy Clin Immunol 1989;84:168-74.
2. Nelson HS, Rosloniec DM, McCall LI, Ikle D. Comparative performance of five commercial prick skin test devices. J Allergy Clin Immunol 1993;92:750-6.
3. Nelson H. Clinical application of immediate skin testing. In: Spector SL, editor. Provocative challenge procedures: background and methodology. Mount Kisco, N.Y.: Futura Publishing Co., Inc; 1989.