A comparison of two single-headed and two multi-headed allergen skin test devices

In-Kyu Yoon, M.D.,* Bryan L. Martin, D.O.,* and Warner W. Carr, M.D.#, The Immunotherapy Committee of the American Academy of Allergy, Asthma, and Immunology

ABSTRACT

Given the potential differences in performance of skin test devices, the purpose of this study was to prospectively assess the performance of two single-headed and two multiheaded devices for allergy skin testing in terms of wheal size, sensitivity, specificity, intradevice variability, and pain level. Two single-headed devices (Greer Pick, Duotip-Test) and two multiheaded devices (Multi-Test II, OMNI) were tested in 15 subjects in a prospective partially blind fashion looking at wheal reactions and pain using histamine and glycerol-saline on the arms and back. Differences among devices in wheal size and pain were noted. Sensitivity, specificity, and intradevice variability were calculated. Differences between corner and interior heads in multiheaded devices were analyzed. No significant differences were observed in wheal size between Greer Pick (7.1 ± 1.4 mm) and Duotip-Test (7.2 ± 1.6 mm). Multiheaded devices were significantly different in wheal size compared with each other and to the single-headed devices (Multi-Test II, 5.4 ± 1.7 mm; OMNI, 3.3 ± 1.2 mm). Single-headed devices were more sensitive (100% each, 95% CI of 92–100%) than the multiheaded devices. Multi-Test II was significantly more sensitive (83%, 95% CI of 78–87%) than OMNI (57%, 95% CI of 51–62%). There was significant intradevice variability for the multiheaded devices with corner heads being significantly more sensitive than interior heads. Specificities for all devices were equally good (≥97%). Pain was greater for multiheaded devices than single-headed devices but was generally mild. In conclusion, this study supports the idea that single-headed devices may be more sensitive and consistent than multiheaded devices. Multi-Test II is more sensitive than OMNI. In multiheaded devices, corner heads are more sensitive than interior heads.

(Allergy Asthma Proc 27:473–478, 2006; doi: 10.2500/aap.2006.27.2937)

Percutaneous allergy skin testing is an important tool used both in clinical allergy practice and in epidemiological and research studies. Despite the importance of percutaneous allergy testing in clinical decision making and study outcomes, there has not been consistency in the types of devices used for these tests.[1] This may be particularly relevant because of the variability in performance of different skin test devices when they have been evaluated in the past.[2–7] The situation is further clouded by the ongoing introduction of modified and new skin test devices by different manufacturers.

Skin test devices have traditionally consisted of single-headed devices that deliver one prick or puncture at a time. Newer devices consist of modified versions of single-headed devices and, increasingly, of multiheaded devices that deliver several pricks or punctures at a time. Multiheaded devices have become more widely used because they may lead to greater efficiency and productivity and may be better tolerated by young children.

In previous comparisons of different skin test devices, there has been significant variability noted in wheal size and sensitivity.[5–7] Some of these studies have shown a trend toward higher sensitivity and consistency for single-headed devices than multiheaded devices.[7,8] However, this has not been a uniform finding in all studies.[1]

Additionally, the intradevice variability of the multiple heads in multiheaded devices has been evaluated. In a study by Carr *et al.*,[7] this intradevice variability was noted to be greater in the four interior heads than in the four corner heads in three different eight-headed devices studied.

Given the potential differences in performance of skin test devices, the purpose of this study was to assess prospectively the performance of two single-headed devices and two multiheaded devices in terms of wheal size, sensitivity, specificity, variability, and pain level. For the multiheaded devices, we sought to compare the sensitivity of corner heads versus interior heads. We chose the two single-headed devices, Greer-Pick (Greer Labs, Lenoir, NC) and Duotip-Test (Lincoln Diagnostics, Decatur, IL), because of good performance in previous studies in terms of sensitivity and variability.[5–7,9,10] We chose the multiheaded device,

*From the *Walter Reed Army Medical Center, Washington, D.C., and #Walter Reed Army Institute of Research, Silver Spring, Maryland*

Supported by the Department of Clinical Investigation and the Department of Allergy/Immunology, Walter Reed Army Medical Center, 6900 Georgia Avenue, Washington, D.C. 20307

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense

Address correspondence and reprint requests to In-Kyu Yoon, M.D., Walter Reed Army Medical Center, 6900 Georgia Avenue, Washington, D.C. 20307

Multi-Test II (Lincoln Diagnostics), because it also had performed well in previous studies.[5–7] The remaining multiheaded device, OMNI (Greer Labs), was chosen because it had not previously been evaluated in the scientific literature.

METHODS

Study Design

This was a prospective, partially blind clinical trial approved by the Walter Reed Army Medical Center Clinical Investigation Committee and the Human Use Committee. Informed consent was obtained from all subjects before enrollment in the study. For each subject, skin testing was performed during a single session. Devices were applied on the volar surface of the forearm and on the skin of the upper back using both histamine (10 mg/mL; Hollister-Stier, Spokane, WA) and glycerol-saline (50%; Hollister-Stier). Each device was applied four times on each subject (one histamine and one glycerol-saline on the arm; one histamine and one glycerol-saline on the back). For the multiheaded devices, a single application consisted of either all histamine or all glycerol-saline. The order in which devices were applied was randomly rotated on each body site. No distinction was made between the right and left side of the body. Single-headed devices were spaced at least 30 mm apart. Multiheaded devices were spaced according to the fixed design of the multiple heads of the device, which ranged from 20 to 30 mm. The total number of skin pricks for an individual subject was 80. The total number of skin pricks for the entire study was 1200.

A single technician performed all device applications and another technician performed all readings. The technician applying the pricks was blinded to the content of the applied solution (whether histamine or glycerol-saline) but, for obvious reasons, was not blinded to the device. The technician performing the readings was blinded to the content of the applied solution but was not blinded to the device because of the characteristic imprints left by some of the devices. Before study initiation, representatives of the manufacturers trained both technicians whose testing techniques were observed to the satisfaction of the representatives.

Pain was assessed on a scale from 0 to 10 with 0 being no pain and 10 being the worst imaginable pain.[11] The subject reported the level of pain immediately after every application of each device.

Subjects

Fifteen male or female subjects aged 18–70 years were recruited for the study. Subjects were excluded if they had dermatographism, severe atopic dermatitis, or asthma, or were taking antidepressants or tricyclic medicines. Antihistamines were withheld for 1 week and leukotriene antagonists and H_2-blockers were withheld for 72 hours before testing.

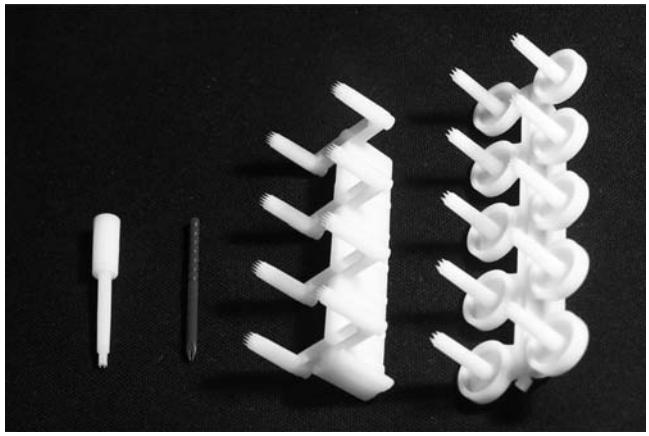


Figure 1. *Shown from left to right is the Greer Pick, Duotip-Test, Multi-Test II, and OMNI.*

Property of
Greenside Publications
IP: 24.221.42.242

Devices

Two single-headed and two multiheaded devices were tested. The single-headed devices were the Greer Pick and the Duotip-Test. The multiheaded devices were the Multi-Test II and the OMNI (see Fig. 1).

Skin Testing

Devices were applied and readings were taken first on the arm and then on the back. All devices were applied according to the techniques recommended by representatives of the manufacturers. The Greer Pick was applied using the upward prick method. The Duotip-Test was applied using light vertical pressure with rotation. The Multi-Test II was applied using vertical pressure with a rocking motion one time side-to-side and one time up-and-down ("north-south-east-west" motion). The OMNI was applied using vertical pressure with a rocking motion three times side-to-side ("east–west" motion). Wheal and flare reactions were read 15 minutes after application by measuring the longest orthogonal diameters.

Statistical Analysis

Based on prior studies,[7] the expected SD of the difference between devices (within a subject) for a wheal reaction was estimated to be ~2 mm. Because four devices were tested, six pairwise comparisons were possible. The comparison-wise type I error was, therefore, set at 0.008 using a Bonferroni correction. A sample size of 15 subjects was determined to have 80% power to detect a difference of 2 mm in the wheal reaction.

The wheal and flare results for histamine application were analyzed using repeated measures analysis of variance, with the within-subject factors being body

Table 1 **Wheal and flare reactions for four devices**

	Histamine Wheal,* Mean (SD)	Histamine Flare,* Mean (SD)	Histamine Wheal, CV	Saline Wheal,* Mean (SD)	Saline Flare,* Mean (SD)	Sensitivity (95% CI)	Specificity (95% CI)
Single-headed devices							
Greer Pick	7.1 (1.4)	28.4 (8.9)	0.20	0.0 (0.0)	4.4 (1.4)	100% (92–100%)	100% (92–100%)
Duotip-Test	7.2 (1.6)	28.6 (8.0)	0.22	0.2 (0.8)	4.4 (2.0)	100% (92–100%)	97% (82–100%)
Multiheaded Devices							
Multi-Test II	5.4 (1.7)	20.5 (5.9)	0.31	0.0 (0.1)	4.5 (0.7)	83% (78–87%)	100% (98–100%)
OMNI	3.3 (1.2)	14.3 (4.5)	0.37	0.0 (0.0)	4.4 (1.2)	57% (51–62%)	100% (99–100%)

**Size in millimeters.*

site (arm and back) and device. When calculating sensitivity and specificity, a true positive was considered to be a histamine wheal of ≥3 mm and a true negative was considered to be a glycerol-saline wheal of <3 mm. A reaction was considered to be a false negative if a histamine wheal was <3 mm. A reaction was considered to be a false positive if a glycerol-saline wheal was ≥3 mm. For multiheaded devices, each individual head was used in the calculation of sensitivity and specificity. Sensitivity and specificity of each device were presented as proportions with 95% CIs and devices were compared using Fisher's exact test (two-tailed). When analyzing the multiheaded devices, intradevice variability was described using the mean of within-subject CVs for each device. For the multiheaded devices, the wheal produced by each head was compared using repeated measures analysis of variance.

Pain scores were compared among devices using the Wilcoxon signed-rank test. Median pain scores were presented as well. For interdevice comparisons of pain and wheal and flare sizes within groups, there were six possible pairwise analyses. Using a Bonferroni correction of the overall experimental p value of 0.05, a value of $p = 0.008$ (0.05/6) was considered significant.

RESULTS

No significant differences were observed in histamine wheal size between the two single-headed devices, Greer Pick and Duotip-Test ($p = 0.61$). However, the two multiheaded devices, Multi-Test II and OMNI, were significantly different in histamine wheal size compared with each other and with each of the single-headed devices. The single-headed devices had wheal sizes that were larger than the multiheaded devices (p values ranging from <0.0005 to 0.002), and the Multi-Test II had a wheal size that was larger than the OMNI ($p < 0.0005$). The same comparisons also held true for histamine flare reactions. The histamine flare sizes did not differ between the two single-headed devices ($p = 0.91$). The two multiheaded devices had significantly different histamine flare sizes compared with each other and with each of the single-headed devices (all values of $p < 0.0005$; see Table 1).

The single-headed devices were significantly more sensitive than the multiheaded devices. The Greer Pick and Duotip-Test had sensitivities of 100% each (95% CI of 92–100%). The Multi-Test II was significantly more sensitive than the OMNI with sensitivity values of 83% (95% CI of 78–87%) and 57% (95% CI of 51–62%), respectively. Specificity values for all devices were ≥97% (Table 1).

Differences in histamine wheal size between the arm and the back were noted also. In general, the mean histamine wheal size for each device was somewhat greater on the arm than on the back. However, this difference reached statistical significance only for the Duotip-Test ($p = 0.004$) and approached statistical significance for the Multi-Test II ($p = 0.027$; Table 2).

There was significant intradevice variability in both multiheaded devices as shown by the intradevice CV values listed in Table 3. Intradevice variability was significantly greater for the OMNI than the Multi-Test II (CV of 0.94 versus 0.43 on the arm, $p < 0.0005$; 1.09 versus 0.62 on the back, $p = 0.003$).

There were significant differences in wheal size between the corner heads and interior heads in both multiheaded devices. For the Multi-Test II, the mean wheal size for the four corner heads was 6.3 mm (95% CI of 5.7–7.0 mm) as compared with 4.4 mm (95% CI of 3.6–5.2 mm) for the four interior heads (difference with $p = 0.009$). For the OMNI, the mean wheal size for the

Property of OceanSide Publications IP: 94.221.45.242

Table 2 **Histamine wheal reactions on arm and back**

	Histamine Wheal,* Mean (SD)
Greer Pick	
Arm	7.4 (1.6)
Back	6.7 (1.2)
Duotip-Test	
Arm	7.9 (1.6)
Back	6.5 (1.3)
Multi-Test II	
Arm	6.0 (1.6)
Back	4.7 (1.6)
OMNI	
Arm	3.6 (0.9)
Back	2.9 (1.3)

**Size in millimeters.*

Table 3 **Intradevice variability in histamine wheal in two multiheaded devices**

	Mean Intradevice CV (SEM)
Multi-Test II	
Arm	0.43 (0.08)
Back	0.62 (0.10)
OMNI	
Arm	0.94 (0.07)
Back	1.09 (0.11)

four corner heads was 4.6 mm (95% CI of 4.1–5.1 mm) as compared with 2.4 mm (95% CI of 1.8–3.0 mm) for the six interior heads (difference with $p = 0.001$). In addition, the sensitivity of the corner heads was significantly higher than the interior heads in both multiheaded devices. In the Multi-Test II, the four corner heads had a sensitivity of 91% (95% CI of 84–95%) and the four interior heads had a sensitivity of 75% (95% CI of 67–82%). In the OMNI, the four corner heads had a sensitivity of 77% (95% CI of 68–83%) and the six interior heads had a sensitivity of 43% (95% CI of 36–51%; Table 4).

Mean pain as measured on a scale from 0 to 10 was greater for the multiheaded devices than for the single-headed devices but was generally mild. The two single-headed devices had mean pain scores of 1.3 for the Greer Pick and 1.4 for the Duotip-Test. The two multiheaded devices had mean pain scores of 2.1 for the Multi-Test II and 2.1 for the OMNI. There were no significant differences in the mean pain scores between the two single-headed devices or between the two multiheaded devices. Both multiheaded devices had

Table 4 **Histamine wheals for corner versus interior heads in 2 multiheaded devices**

	Histamine Wheal,* Mean (95% CI)	Sensitivity (95% CI)
Multi-Test II		
Corner heads	6.3 (5.7–7.0)	91% (84–95%)
Interior heads	4.4 (3.6–5.2)	75% (67–82%)
OMNI		
Corner heads	4.6 (4.1–5.1)	77% (68–83%)
Interior heads	2.4 (1.8–3.0)	43% (36–51%)

**Size in millimeters.*

Table 5 **Pain scores for four devices**

	Pain,* Mean (SD)	Pain,* Median
Single-headed devices		
Greer Pick	1.3 (1.4)	1
Duotip-Test	1.4 (1.4)	1
Multiheaded devices		
Multi-Test II	2.1 (1.7)	1
OMNI	2.1 (2.0)	1

**Pain scores based on a scale from 0 to 10.*

significantly higher pain scores than either of the single-headed devices (p values ranging from <0.0005 to 0.005). However, all mean pain scores were ≤ 2.1 and the median pain score for each of the devices was 1. There were no significant differences in pain scores between the arm and the back for any of the devices (Table 5).

DISCUSSION

This study revealed significant differences between single-headed and multiheaded devices and within the different heads of multiheaded devices. Both of the single-headed devices when used according to manufacturer suggested instructions had higher sensitivities than either of the multiheaded devices. The same trend was noted by Carr *et al.*[7] and may reflect an inherent advantage in single-headed devices over multiheaded devices. The wheal sizes of the single-headed devices were larger overall than the multiheaded devices and this finding probably contributed to the higher sensitivity values for the single-headed devices. However, this higher sensitivity did not come at the expense of lower specificity. Specificity values were similar in all four devices as illustrated by the universally low wheal and flare reactions at the glycerol-saline sites.

Variability from subject to subject was lower for the single-headed devices than the multiheaded devices, indicating overall better reproducibility of the single-

Property of
OneAmSide Publications
IP: 54.221.42.242

headed devices. The single-headed devices had consistent performance from application to application, whereas the various heads of the multiheaded devices had a wide range of wheal and flare reactions ranging from no reaction to reactions equivalent to the single-headed devices. Intradevice variability was significant in the multiheaded devices and may have contributed to the lower sensitivity of the multiheaded devices. Of the two multiheaded devices, the higher intradevice CV values of the OMNI versus the Multi-Test II probably contributed to the lower sensitivity of the former versus the latter device.

This study found a significant difference in performance between the corner heads and interior heads in the multiheaded devices. The corner heads in both multiheaded devices had significantly larger wheal sizes and higher sensitivities than the interior heads. These findings may have implications for the reliability of results from individual heads of multiheaded devices and, therefore, may have consequences in clinical allergy practice. In clinical practice, if the corner heads and interior heads are considered to be equivalent, incorrect assumptions might be made about the relative sensitivities of the various heads of multiheaded devices. For example, if the corner heads are used to place positive and negative controls and the interior heads are used for suspected allergens, the results from the allergens may not be as sensitive as would be indicated by the positive and negative controls.

Although this study was not designed to determine the reasons for the higher sensitivity of the corner heads versus interior heads, we conjecture that certain factors may have played a role. In the Multi-Test II, the north-south-east-west technique of application may have placed greater pressure at the ends of the device, leading to larger wheals at the corner heads. The OMNI was applied only with an east-west technique so this same explanation would not be as applicable to the OMNI. However, there may have been subtle movement of the device in a north-south direction contributing to increased pressure at the corner heads. As was already noted, the Multi-Test II performed significantly better than the OMNI in terms of both sensitivity and intradevice variability. An important design difference between the two devices is that the 10-headed OMNI contains two more interior heads than the eight-headed Multi-Test II. Given the likelihood that interior heads might be less sensitive than corner heads, the two extra interior heads of the OMNI may have impacted negatively on the sensitivity and consistency of the OMNI. In addition, the greater number of heads contained on the OMNI may have led to less pressure being generated at each individual head given the same amount of force applied to the entire device. The lower pressure generated at each head of the OMNI may have then contributed to the smaller wheal and flare sizes and poorer sensitivity as compared with the Multi-Test II.

Although mean pain was noted to be higher in the multiheaded devices than in the single-headed devices, the overall low level of pain would indicate that this difference is unlikely to be of clinical significance. A previous study by Nelson *et al.*[6] resulted in higher overall pain scores than in the current study. For instance, when looking at the same Multi-Test II device, the mean pain score was 4.4 in the Nelson study as compared with 2.1 in this study. Both pain scales were from 0 to 10. Although comparisons between different studies are often difficult to make, it is possible that the reason for this discrepancy was that the force of application may have been somewhat greater in the Nelson study as compared with this study and was reflected in the higher pain scores. This possible greater force of application may explain the lack of significant difference seen in corner versus interior heads in the Nelson study that became more apparent when less force was applied in this study.

Wheal and flare reactions were either larger or statistically equivalent on the arm versus the back for all four devices. Even when statistically equivalent, there was a trend toward larger reactions on the arm than the back for all devices. This finding conflicts with previous studies[7] and is not readily explained. It is possible that this difference was a technician-dependent finding rather than a reflection of the inherent properties of the skin of the arm or the back. This finding highlights the difficulties in making any generalizations about comparative wheal and flare reactions in various regions of the body when so many additional factors may be playing a role.

In conclusion, this study supports the idea that single-headed devices may be more sensitive and consistent than multiheaded devices. For multiheaded devices, caution should be used when interpreting results from different individual heads because corner heads appear to be more sensitive than interior heads. In particular, the newly introduced OMNI had the poorest sensitivity and consistency of all four devices tested. As with any testing procedure, allergy skin tests should be evaluated in the context of the clinical situation. Thus, when using multiheaded devices, a possible strategy might be to retest negative results using a single-headed device if those negative results are of questionable validity based on the clinical history.

Property of
OceanSide Publications
IP: 54.221.42.21

ACKNOWLEDGMENTS

Thanks to Levi Harvey and Clarence Yarborough who administered and read the skin tests, and to Robin Howard who assisted with statistical analysis.

REFERENCES

1. Illi S, Garcia-Marcos L, Hernando V, et al. Reproducibility of skin prick test results in epidemiologic studies: A comparison of two devices. Allergy 53:353–358, 1998.
2. Adinoff AD, Rosloniec DM, McCall LL, and Nelson HS. A comparison of six epicutaneous devices in the performance of immediate hypersensitivity skin testing. J Allergy Clin Immunol 84:168–174, 1989.
3. Demoly P, Bousquet J, Manderscheid JC, et al. Precision of skin prick and puncture tests with nine methods. J Allergy Clin Immunol 88:758–762, 1991.
4. Nelson HS, Rosloniec DM, McCall LI, and Ikle D. Comparative performance of five commercial prick skin test devices. J Allergy Clin Immunol 92:750–756, 1993.
5. Nelson HS, Lahr J, Buchmeier A, and McCormick D. Evaluation of devices for skin prick testing. J Allergy Clin Immunol 101: 153–156, 1998.
6. Nelson HS, Kolehmainen C, Lahr J, et al. A comparison of multiheaded devices for allergy skin testing. J Allergy Clin Immunol 113:1218–1219, 2004.
7. Carr WW, Martin B, Howard RS, et al. Comparison of test devices for skin prick testing. J Allergy Clin Immunol 116:341–346, 2005.
8. Ates A, Kinikli G, Turgay M, et al. The results of skin prick testing in patients with allergic rhinitis: A comparison between a multiple lancet device and a single lancet. Asian Pac J Allergy Immunol 22:109–114, 2004.
9. Sangsupawanich P, Chamnanphol S, and Koonrungsrisomboon D. Evaluation of three methods for using the Duotip-Test device for skin testing. Asian Pac J Allergy Immunol 18:153–156, 2000.
10. Stryk SV, and Slavin RG. Evaluation of Duotip-Test. J Allergy Clin Immunol 99:149–150, 1997.
11. Wong DL, Hockenberry-Eaton M, Wilson D, et al. Wong-Baker FACES Pain Rating Scale. In Wong's Essentials of Pediatric Nursing, 6th ed. St. Louis: Elsevier, 1301–1302, 2001. □

Property of

OceanSide Publications

IP: 84.221.42.242