
Skin testing: making the most of every prick

There is no question that great progress has been made in diagnostic testing for allergy in recent years. The most significant advances have come with the increasing availability of standardized extracts for skin testing and refinements in radioallergosorbent test technology. However, although these advances are clearly significant, they should not lull us into believing that the diagnosis of allergy is even close to an exact science. This fact is clearly demonstrated in the excellent paper by Rhodius et al¹ in this issue of the *Annals*, which reveals striking inconsistencies in even the most basic skin testing methods.

In this simple but well conceived study, the authors compared two different prick skin test devices, the standard lancet and the Quintest (Bayer Corporation, West Haven, CT), and the extracts from two reputable manufacturers (Bayer and ALK Allergologisk Laboratorium A/S, Hørsholm, Denmark) for the diagnosis of cat and dust mite allergy. The study provides several important findings. First, they found that wheal sizes differed significantly between the different skin test devices. For example, when using the Bayer extracts, the mean wheal size was 60% to 150% larger with the lancet than with the Quintest. Second, these differences in wheal size were magnified even further when the different extracts were used, with greater responses occurring with the Bayer extracts for dust mite and the ALK extracts for cat. An example of these combined differences is that the mean wheal size was 8.0 mm for *Dermatophagoides farinae* when using the lancet with the Bayer extract, compared with just 2.4 mm when using the ALK extract with the Quintest.

Although the differences in wheal size were striking, and better agree-

ment would certainly be desirable, these findings alone are not necessarily of great significance. As long as the test provides an accurate answer as to the presence or absence of a specific allergy, the actual size of the skin test may not be terribly meaningful. This would especially be true if there is overall consistency in one's test methods, such that even the relative magnitude of one skin test versus another should be apparent with whatever test method is used.

Unfortunately, however, the authors also found that there were major discrepancies in the ability of the Quintest device, especially when used with ALK extracts for dust mite and Bayer extracts for cat, to produce a positive skin test when compared with the lancet. These discrepancies are far more concerning. For example, in this study the combination of the Quintest device and ALK extracts provided only 62% sensitivity in the overall identification of atopy compared with the reference combination of the lancet and Bayer extracts. The authors stress that these differences would have major implications in the identification of atopy in research studies, and that the results of one study may not be comparable with the results of another if different devices and extracts are used.

Although I agree that the identification of atopy for research purposes is indeed very important, I believe that the implications of these findings have far greater significance for the day to day care of allergic patients. For example, based on this study, if I chose to use the Quintest and ALK antigens for dust mite testing in my office, only 50% of the mite allergic patients would be identified. Even more striking, if I used the Quintest and Bayer antigens for cat testing, only 27% of my patients with cat allergy would be cor-

rectly diagnosed. These discrepancies could have a staggering impact on my ability to provide appropriate patient care, particularly in providing advice about allergen avoidance and in the prescription of immunotherapy.

To some, these results may seem surprising or even unbelievable. It is easy to become so convinced in the infallibility of our skin testing that inaccurate results seem rare or even impossible. In reality, however, it is likely that similar discrepancies in test results occur on a daily basis from one allergy office or allergy study to another. The fact is that skin testing is not, and, in all probability can never be, a precise science. The reason for this pessimism is that even in the best case, every skin test result depends on the combination of numerous variables, only some of which are under our control. To name but a few, the variables that must be considered whenever a skin test is applied include extract potency,^{1,2} the specific device used,^{1,3-6} the age of the patient,⁷ the location on the body that is used,⁵ the season of the year,⁸ the health of the skin, the skill of the tester, the accuracy of the interpretation, and the influence of medications on test results.

The first two of these variables, extract potency and skin test devices, deserve special attention. The quality of extracts have clearly improved over time, particularly because of the increasing availability of standardized products. However, it should be recognized that standardized does not necessarily mean equal in potency. Nelson² recently compared a variety of standardized extracts used in allergen immunotherapy and found rather striking variability in their content of major allergen. For example, in a comparison of 28 *Dermatophagoides pteronyssinus* extracts, all standardized to contain 10,000 AU/mL, the major allergen

content ranged from 68 to 385 $\mu\text{g/mL}$. The best results were found for cat, in which major allergen content ranged only twofold in the 12 extracts tested, from 26 to 52 $\mu\text{g/mL}$, whereas the worst results occurred for orchard grass extracts, which contained from 294 to 2,414 $\mu\text{g/mL}$ of major allergen. These results are not unlike those reported in the Rhodius study, where they reported comparable potency of the cat extracts but approximately 10-fold differences in potency in the dust mite extracts from the two manufacturers.

Therefore, it is clear that even standardized extracts are far from perfect and may be a significant source of error in skin testing. The explanation for this shortcoming lies in the means by which extracts are standardized. Currently, standardization is based on arbitrary units that are determined by a combination on skin test reactivity or by an immunochemical evaluation of major allergen content.⁹ These techniques may include competitive binding assays, gel electrophoresis, and isoelectric focusing, and do not necessarily include specific allergen assays, which would provide the most accurate measurement of major allergen content. Although there are reasons for these approaches, including a desire not to exclude minor allergens from an extract, there is clearly room for improvement. The technology is certainly available to include an accurate measure of major allergen content in the process of allergen standardization; this would seem a logical and desirable step.

As evidenced in the Rhodius study, skin test devices are another major source of potential error in skin test results. They found that the Quintest device provided smaller skin test responses and more false-negative results compared with the lancet. There are now a wide variety of skin test devices available for use and new devices are being introduced regularly. No two devices are exactly alike and we now have choices that include single and multitest devices that can be constructed of either metal or plastic. Some make a single puncture into the

skin whereas others make multiple punctures and, in addition, the puncture depth and the amount of allergen that is introduced can vary depending on the device used and the pressure applied by the tester.

Some of these devices have been formally compared, and these studies have shown significant interdevice differences with regard to wheal size and both false-positive and false-negative results.³⁻⁶ It was also shown that there can be significant differences depending on the location that the test was applied. For example, there is a significant gradient in skin test responses on the back, with decreasing wheal size as one proceeds from the upper to the lower back.⁵ It is important that anyone performing skin testing review the available data on skin test devices to make an optimal decision for the testing to be performed, whether for research use or patient care.

Given the inherent difficulties with skin testing, some have argued that in vitro measures of allergen-specific immunoglobulin E might be preferable. Whereas the performance of these tests also relies directly on the quality of the allergen extracts that are available, the rest of the potential sources of error in skin testing are eliminated. However, although in vitro tests may have some advantages for the diagnosis of food allergy,¹⁰ skin testing is still the preferred method for the diagnosis of inhalant allergy. In addition to significant cost issues, studies have shown that in vitro testing is overall less sensitive than prick skin testing,^{11,12} and some have argued that a skin test result is more likely to correlate with true allergic sensitivity. However, it should also be noted that the studies comparing skin tests and in vitro tests are likely to have been performed in ideal circumstances, thereby maximizing the sensitivity of the skin test results. If the data from the Rhodius are considered, it is very likely that an in vitro test would have been far more sensitive than skin testing using the Quintest device.

How then should we proceed? Despite the potential inconsistencies, skin

testing still has tremendous value in the diagnosis of allergy for both clinical and research purposes. Radioallergosorbent tests may serve as a useful adjunct to skin testing, but are unlikely to ever supplant skin testing as the first-line method of allergy testing. However, it is clear that further efforts are needed to make skin testing as accurate and reproducible as possible. These efforts need to include both the industry and the Food and Drug Administration to improve standardization not just of allergen extracts but also of skin test devices. In addition, it is essential that all of us who perform skin testing take initiative to recognize the variables that may affect our test results and make every possible effort to overcome them.

ROBERT A. WOOD, MD
Department of Pediatrics
Division of Allergy and Immunology
Johns Hopkins University School of Medicine
Baltimore, Maryland

REFERENCES

1. Rhodius R, Wickens K, Cheng S, Crane J. A comparison of two skin test methodologies from two different manufacturers. *Ann Asthma Allergy Immunol* 2002;88:374-379.
2. Nelson HS. The use of standardized extracts in allergen immunotherapy. *J Allergy Clin Immunol* 2000;106:41-45.
3. Nelson HS, Lahr J, Buchmeier A, McCormick D. Evaluation of devices for prick skin testing. *J Allergy Clin Immunol* 1998;101:153-156.
4. Adinoff AD, Rosloniec DM, McCall LL, Nelson HS. A comparison of six epicutaneous devices in the performance of immediate hypersensitivity skin testing. *J Allergy Clin Immunol* 1989;84:168-174.
5. Nelson HS, Rosloniec DM, McCall LI, Ikle D. Comparative performance of five commercial prick skin test devices. *J Allergy Clin Immunol* 1993;92:750-756.
6. Illi S, Garcia-Marcos L, Hernando V, et al. Reproducibility of skin prick test results in epidemiologic studies: a comparison of two devices. *Allergy* 1998;53:353-358.

-
-
7. Skassa-Brociek W, Manderscheid JC, Michel FB, Bousquet J. Skin test reactivity to histamine from infancy to old age. *J Allergy Clin Immunol* 1987;80:711–716.
 8. Oppenheimer JJ, Nelson HS. Seasonal variation in immediate skin test results. *Ann Allergy* 1993;71:227–229.
 9. American Academy of Allergy, Asthma and Immunology (AAAAI). The use of standardized allergen extracts. *J Allergy Clin Immunol* 1997;99:583–586.
 10. Yunginger JW, Ahlstedt S, Eggleston PA, et al. Quantitative IgE antibody assays in allergic diseases. *J Allergy Clin Immunol* 2000;105:1077–1084.
 11. Wood RA, Phipatanakul W, Hamilton RG, Eggleston PA. A comparison of skin prick tests, intradermal skin tests, and RASTs in the diagnosis of cat allergy. *J Allergy Clin Immunol* 1999;103:773–779.
 12. Kelso JM, Sodhi N, Gosselin VA, Yunginger JW. Diagnostic performance characteristics of the standard Phadebas RAST, modified RAST, and Pharmacia CAP system versus skin testing. *Ann Allergy* 1991;67:511–514.
-