

EDITORIAL

Food allergy—getting more out of your skin prick tests

Skin prick testing has been used for over a 100 years [1] and is of fundamental importance in the practice of clinical allergy. While it is an important tool in the context of a clinical history, there are limits to its value. We are therefore often unable to establish definitive diagnoses of food allergy without resorting to food challenges.

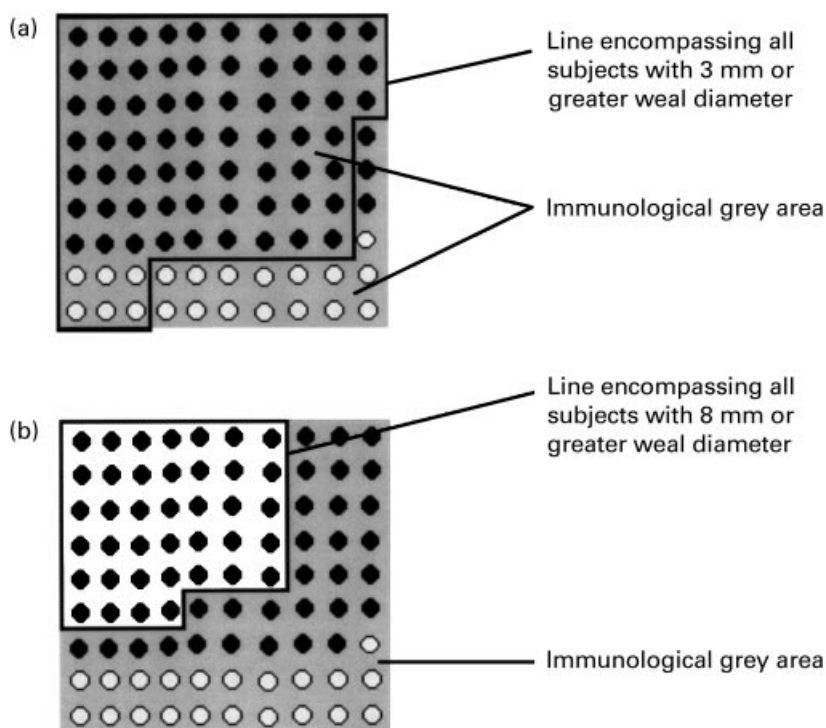
In this issue of the Journal, Sporik *et al.* [2] present a case series of 467 children, aged 1 month to 16 years, evaluated for suspected food allergy in a tertiary paediatric allergy clinic. They correlate skin prick wheal sizes with the results of open food challenges to cows milk, egg or peanut. Their data demonstrate that if the traditional 3 mm wheal is used as the lower limit for defining allergy, there are many false-positive and -negative results (Fig. 1a). This creates a large immunological grey area where we cannot make a definitive diagnoses on the basis of the test result. Their paper demonstrates how this problem can be overcome by increasing the cut-off value, thereby reducing the immunological grey area (Fig. 1b). They identify skin wheal diameters at and above which negative reactions do not occur on food challenge: cow's milk, 8 mm; egg, 7 mm; and peanut, 8 mm. The need for a corroborative food challenge to make the diagnosis could have been avoided in 33%, 56%

and 68%, respectively, of the children in this population. Such an approach has great clinical significance, considerably reducing the need for costly and time-consuming food challenges.

The concept of a cut-off value, above which a test has a very high positive predictive value for food allergy, is not entirely new. Specific IgE values, generated by the Pharmacia Cap system, have been correlated with the results of double-blinded food challenges [3]. Specific IgE results giving positive predictive values of at least 95% were: egg, 6 kUA/L; milk, 32 kUA/L; peanut, 15 kUA/L; and fish, 20 kUA/L. Skin prick testing has considerable advantages in terms of cost and the time taken for a result to be available.

How applicable are these skin prick cut-off values [2] to other patients in different clinical settings? Cutaneous reactivity is known to vary with age. The histamine-induced wheal diameter increases by 125% from 4 days to 24 months [4] with further 50% increase over the childhood and teenage years [5]. Similar changes with age occur in response to codeine, a mast cell degranulator [4]. There is also data to suggest that time of day [6], season [7] and patient's sex [8] affect cutaneous reactivity. Together, these data suggest that

Fig. 1. The value of using different cut-off levels to diagnose peanut allergy with skin prick testing. Data taken from Sporik's population [2]. (a) The low cut-off value of 3 mm for skin prick testing, that is traditionally used, is associated with a significant number of false negative and false positive results. Whether the wheal diameter is more or less than 3 mm, we cannot use the result to definitively rule in or rule out the diagnosis of peanut allergy. This creates a large immunological grey area where the test results are of limited value as they cannot be used diagnostically. (b) With a high cut-off value of 8 mm, we are able to establish a definitive diagnosis of peanut allergy in 43% of Sporik's population [2]. This considerably reduces the size of the immunological grey area and the need for food challenges. (●) Food-allergic subjects; (○) normal subjects; background grey colour represents the immunological grey area.



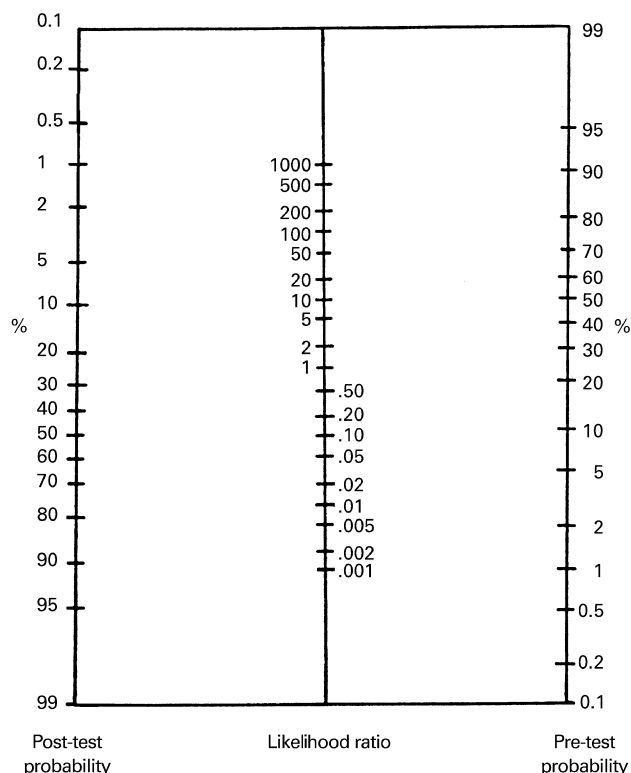


Fig. 2. The use of likelihood ratios to determine the predictive value of different skin test results in different subjects. The Fagan likelihood ratio nomogram [18] is used to determine the probability that a subject has peanut allergy (post-test probability) taking into account their pretest probability and the likelihood ratio associated with their skin prick test result. By joining the pretest probability in the first column to the likelihood ratio in the middle column with a ruler, the post-test probability can be read from the third column. Adapted with permission from [18].

if skin prick testing is to be standardized, results may need to be interpreted in terms of factors such as the subject's age and sex as well as the timing of the test. To achieve optimal specificity, different cut off values are likely to be required for different subpopulations of children.

For a test to be widely applicable, it must be performed using a standardized technique. The single headed lancet, used in this study [2], has a lower intra- and interoperator variability than other methods, such as the bifurcated or multiheaded lancet systems [9,10]. In addition to the technique, the site chosen is critical with a 20% variation in weal size, with histamine or allergen, depending on which part of the body is used for skin testing [9].

Whilst standardization of skin prick testing in terms of the population characteristics and technique employed is important, standardization of the allergen content of the reagent is the single most important determinant of outcome. Despite

the efforts of international committees [11], there is a lack of reagent standardization for food allergens. Commercial skin prick tests for peanut have been found to have different RAST inhibition curves, suggesting that they have different IgE-binding determinants, and concentrations of Ara h 1 and Ara h 2 that vary by a 100-fold [12]. Similar results have been seen with other food extract preparations [13]. For each food, there is an optimal allergen concentration for skin prick testing at which false-positive results (associated with a concentrated solution) and false-negative results (associated with a dilute solution) are minimized. This optimal concentration can be derived by generating a receiver-operator curve; such an approach is used to standardize inhalant allergens such as grass pollen [14].

Once standardization is achieved for food extracts, we may be able to use threshold concentrations (the minimum allergen concentration required to produce a particular weal response) to extract more information from the test results. Such an approach has been shown to be particularly useful for defining how clinical reactivity changes with time [11]. This information may be especially helpful with childhood food allergy where there is a need to distinguish between persisting food allergy and resolution of the condition [15].

The fundamental question clinicians must consider when evaluating a diagnostic test is the probability that their patient has the condition, given a particular test result. A test result is useful if the results enable the clinician to be reasonably certain ($\geq 95\%$ confidence) that a patient does or does not have the disease. In practice, clinicians refer to the sensitivity and specificity of a test. However, these measures are more useful in study populations and tell us less about how the test performs at the level of the individual.

Positive and negative predictive values lend themselves more easily to interpretation at the level of the individual. They give the probability that a patient has, or does not have, the condition given a particular test cut-off value. Their main limitation is that they vary according to the prevalence of the condition in the population. A test that has a high positive predictive value in a tertiary allergy clinic will have a much lower one in a primary care population because of the lower rate of food allergy and a corresponding higher rate of false-positive results. Authors have tried to overcome this problem by correcting their results for a population with a lower prevalence of food allergy [3]. However, this does not help the practising allergist who sees a variety of patients each with a different clinical presentation and background of risk factors.

An alternative approach, introduced by the authors [2], is to use likelihood ratios (LR). The LR for a positive test result is the likelihood that a positive test would be expected in a patient with the food allergy compared to the likelihood that the same result would be expected in a patient without

food allergy [16,17] (see appendix 1). The advantage of this approach is that LR values are independent of the prevalence of the condition tested. They can therefore be used to calculate the likelihood of a disease both within a tertiary care centre as well as in the primary care setting. Clinically useful tests generally have a LR of more than 10 [16]. Before a patient's test result can be interpreted using the LR, their pretest probability must be estimated. This is the chance that they are food allergic based only on their clinical presentation and risk factors prior to any test result. The post-test probability of food allergy for a subject can be derived from a statistically derived nomogram [18] (Fig. 2). This takes into account the subject's pretest probability and the LR corresponding to the test result.

The use of likelihood ratios gives us the ability to interpret skin testing results within the subject's clinical context. This can be demonstrated by considering the following clinical presentations (Table 1). For a child (case D, Table 1) with a convincing history of peanut allergy (pretest probability (80%), a skin prick weal diameter of 6 mm can be regarded as being diagnostic as it achieves a post-test probability of allergy of at least 95%. However a 6-mm weal is far less clinically useful in the clinical context of an asymptomatic child whose sibling has peanut allergy (case B, Table 1). In this case, a pretest probability of 10% in association with a 6-mm weal gives a post-test probability of 45%. This means that the child has less than a 1 in 2

chance of being peanut allergic and needs to undergo a peanut challenge to definitely decide whether or not they are food allergic. In practice, allergists tend to use a 3-mm or greater weal as being predictive of allergy. However in the context of diagnosing food allergy, this traditional, low cut-off value introduces a large potential for diagnostic error. For example, consider a child with a poor history of peanut allergy (case A, Table 1) whose pretest probability of allergy is likely to be less than the population prevalence (approximately 0.5%). For this subject, a 3-mm weal to peanut will actually reduce their chance of being allergic (post-test probability 0.2%) as most subjects with peanut allergy have much larger weal sizes. Therefore in the context of a poor clinical history, a small 3 mm skin prick weal does not support a diagnosis of peanut allergy. The problems associated with interpreting a negative test can be demonstrated by looking at the post-test probabilities associated with a zero weal size. A child with a good history for peanut allergy (pretest probability (80%) and a negative skin prick test still has a one in seven chance of reacting on challenge (case D, Table 1). This explains why a child with a history suggestive of a food allergy but a negative skin prick test must have a food challenge.

The paper presented in this journal [2] demonstrates how skin prick testing results can be used in a more analytical way to gain additional diagnostic information about a patient. The data presented suggest that, on the basis of

Table 1 The use of likelihood ratios to determine the predictive value of different skin test results in different subjects

Clinical presentation	Pre-test probability	Likelihood ratio <i>Sensitivity</i> <i>Specificity</i>	Post-test probability given skin prick testing weal sizes			
			0 mm	3 mm	6 mm	10 mm
			0.05 100% 0%	0.5 95% 70%	6.5 85% 95%	∞ 45% 100%
Case A – Child who developed a headache and starts vomiting 4 h after eating a peanut butter sandwich.	0.5%		<0.1%	0.2%	5%	>99%
Case B – Sister of a child with proven peanut allergy.	10%		0.5%	5%	45%	>99%
Case C – Child with asthma who has developed urticaria 45 min after eating a small piece of peanut butter sandwich.	50%		5%	35%	90%	>99%
Case D – Child who developed urticaria and wheeze on two occasions within minutes of accidental exposure to peanuts.	80%		15%	70%	96%	>99%

These four cases illustrate the use of likelihood ratios in the interpretation of skin prick results with peanut extract. The likelihood ratios have been calculated from the data presented by Sporik *et al.* using skin prick test values of different diameters [2].

skin prick tests, we may be able to avoid a diagnostic challenge in a substantial number of children presenting with food allergy. A cautionary note is required though if these results are to be utilized in other clinics, given the many variables that may affect skin prick testing results. These data will hopefully give further impetus to the standardization of skin prick testing reagents to foods.

Appendix 1.

Considering the subject groups

	Disease	
	Present	Absent
<i>Test result</i>		
Positive	True-positives (a)	False-positives (b)
Negative	False-positives (c)	True-negatives (d)

Sensitivity is the proportion of subjects with the disorder who test positive [$a/(a + c)$]. Specificity is the proportion of subjects without the disorder who test negative [$d/(b + d)$]. Positive predictive value is the probability that a subject has a disorder if they test positive [$a/(a + b)$]. Negative predictive value is the probability that a subject does not have a disorder if they test negative [$d/(c + d)$]. Likelihood ratio for a positive test result is the likelihood that a given test would be expected in a subject with the target condition compared to the likelihood that the same result would be expected in a subject without the condition [$\{a/(a + c)\}/\{b/(b + d)\}$].

References

- 1 Blackey CH. Experimental researches on the causes and nature of Catarrhus aestiuxus (hay fever and asthma). London: Bailliere, Tindall and Cox, 1873.
- 2 Sporik R, Hill D, Hosking CS. Specificity of allergen skin testing in predicting positive open food challenges to milk, egg and peanut in children. *Clin Exp Allergy* 2000; 30:1539–1545.
- 3 Sampson HA, Ho DG. Relationship between food-specific IgE concentrations and the risk of positive food challenges in children and adolescents. *J Allergy Clin Immunol*, 1997; 100:444–51.
- 4 Ménardo JL, Bousquet J, Rodière M, Astruc J, Michel F-B. Skin testing in infancy. *J Allergy Clin Immunol*, 1985; 75:646–51.
- 5 Skassa-Brociek W, Manderscheid J-C, Michel F-B, Bousquet J. Skin reactivity to histamine from infancy to old age. *J Allergy Clin Immunol*, 1987; 80:711–6.
- 6 Lee RE, Smolensky MH, Leach CS, McGovern JP. Circadian rhythms in the cutaneous reactivity to histamine and selected antigens, including phase relationship to urinary cortisol excretion. *Ann Allergy*, 1969; 38:231.
- 7 Oppenheimer JJ, Nelson HS. Seasonal variation in immediate skin test reactions. *Ann Allergy*, 1993; 71:227–9.
- 8 Freidhoff LR, Meyers DA, Marsh DG. A genetic-epidemiologic study of human immune responsiveness to allergens in an industrial population. II The association among skin sensitivity, total serum IgE, age, sex and the reporting of allergies in a stratified random sample. *JACI*, 1984; 73:490.
- 9 Nelson HS, Rosloniec DM, McCall LI, Iklé D. Comparative performance of five commercial prick skin test devices. *J Allergy Clin Immunol*, 1993; 92:750–6.
- 10 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; 53:353–8.
- 11 Dreborg S, Frew A. Position paper: allergen standardization and skin tests. *The European Acad Allergol Clin Immunol Allergy*, 1993; 48 (14 (Suppl.):48–82.
- 12 Hefle SL, Helm RM, Wesley Burks A, Bush RK. Comparison of commercial peanut skin test extracts. *J Allergy Clin Immunol*, 1995; 95:837–42.
- 13 Herian AM, Bush RK, Taylor SL. Protein and allergen content of commercial skin test extracts for soybeans. *Clin Exp Allergy*, 1992; 22:461–8.
- 14 Wheeler AW, Jessberger B, Drachenberg KJ, Rakoski J. Design of an optimally-diagnostic skin test solution for diagnosis of sensitivity to timothy grass (*Phleum pratense*) pollen. *Clin Exp Allergy*, 1996; 26:897–902.
- 15 Hourihane JOB, Warner JO, Roberts SA. Resolution of peanut allergy: case control study. *BMJ*, 1998; 316:1271–5.
- 16 Jaeschke R, Guyatt GH, Sackett DL. User's guides to medical literature III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *J Am Med Assoc*, 1994; 271:703–7.
- 17 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. 1st edn. London. Churchill Livingstone, 1997.
- 18 Fagan TJ. Nomogram for Bayes's theorem. *New Eng J Med*, 1975; 293:257.

G. ROBERTS

G. LACK

Imperial College at St Mary's
Ground Floor Salton House
St Mary's Hospital
Praed Street
London W2 1NY
UK