

A clinical index: a new method to assess efficacy of allergen immunotherapy

McHugh SM, Ewan PW. A clinical index: a new method to assess efficacy of allergen immunotherapy.
Allergy 1992; 47: 115–120.

S. M. McHugh, P. W. Ewan

Molecular Immunopathology Unit, MRC Centre,
Cambridge, UK

Assessment of efficacy in immunotherapy trials is difficult: a number of parameters can be measured; these may give conflicting results and it is not clear which best represents the clinical state. We therefore derived a Clinical Index (CI), by combining five individual parameters—diary card symptom score, visual analogue symptom score, drug score, nasal challenge and skin prick test—by rank analysis. The CI has been used to express the results in a trial of house dust mite (*D. pteronyssinus*) immunotherapy for allergic rhinitis, comparing two active treatments, Pharmedin® and Allpyral®, with placebo. The derived index correlated well with all parameters measured. Correlations between the five individual parameters were variable: with the exception of the skin test, good correlations were seen, each parameter correlating with two or three of the others, but the visual analogue showed the closest association with other parameters. Results from a single parameter taken alone could be misleading, e.g. in the Allpyral group, nasal challenge improved significantly, but symptoms assessed either by visual analogue or diary card score did not. The CI showed significant improvement in the Pharmedin group but not in the Allpyral group. The CI gave a more reliable measure of outcome than individual parameters.

Key words: house dust mite; immunotherapy; efficacy; rank analysis; rhinitis; assessment

Dr. P.W. Ewan
Molecular Immunopathology Unit
MRC Centre, Hills Road
Cambridge CB2 2QH
UK

Accepted for publication 29 August 1991

Although immunotherapy has been widely used for many years, it is of established value for only a limited number of allergic disorders, and may be dependent on the nature of the treatment extract used (6). One of the difficulties in immunotherapy trials is accurate assessment of outcome: a number of different parameters can be measured, e.g. symptoms or sensitivity in the target organ, and a single parameter, e.g. symptoms, can be measured in different ways. Earlier studies often relied on a simple assessment of symptoms alone, but more recent studies tend to have multiple assessments, often including target organ allergen challenge, e.g. nasal, bronchial or sting challenge, and skin prick tests, as a general measure of mast cell reactivity. Results from such a range of assessments can give conflicting information. Even symptom assessment can be difficult, e.g. in a study of house dust mite immunotherapy in patients with asthma and/or rhinitis which the authors concluded was of overall clinical benefit, diary cards showed no improvement in symptoms but questionnaires did (but only for 4 of 7 questions asked); nasal challenge improved significantly (3). How does one assess such a study? It would seem logical to require improvement in symptoms before

a treatment could be described as clinically effective. However, if a study in asthma showed significant improvement on bronchial challenge with allergen yet no improvement in diary card symptom scores, one might reasonably question the accuracy of the subjective measurement. Godfrey (7) has stated that methods used to assess symptoms is “an extremely difficult area in which there is a need for well-validated further work”.

A particular difficulty, especially where conflicting results occur, is that it is not known whether any one parameter gives a more reliable index of disease activity than the others. Interpretation and presentation of results of clinical assessments can therefore be difficult. Further, data from *in vitro* tests, such as specific IgE levels or histamine release assays, compounds the problem. We have therefore derived a Clinical Index (CI) from five parameters in an attempt to give an overall assessment of treatment without bias towards any one method. This has been used to assess the effect of *D. pteronyssinus* immunotherapy in a study involving 80 patients with rhinitis with/without asthma, comparing two active preparations with placebo (5). This paper describes the derivation of the CI and evaluates it in the three

treatment groups. This approach could be adapted for other studies using different parameters as appropriate.

Methods

Eighty patients (44 males and 36 females, age range 17–52 years) with a history of rhinitis with or without asthma due to allergy to house dust mite entered this placebo-controlled study of immunotherapy. Thirty patients received Pharmedgen® *D. pteronyssinus* and 30 histamine dihydrochloride (placebo) double blind. A further 20 patients received Allpyral® *D. pteronyssinus* (Dome/Hollister-Stier), but this part of the study had to be conducted single blind because the manufacturer's dose schedule was different. Criteria for inclusion in the trial, full details of the extracts and treatment regimen have been described in detail previously (5, 9). All patients had a history of house dust mite rhinitis, and positive nasal challenge and skin prick test to Pharmedgen *D. pteronyssinus*.

Five parameters were measured before and after 3 and 12 months treatment: skin prick test, nasal challenge, rhinitis symptoms from visual analogue score, rhinitis symptoms from diary card scores and diary card drug scores (rhinitis).

Results for three parameters were expressed as percent of the maximum possible responses as follows. The maximum for visual analogue score was 100% (where 0% is asymptomatic and 100% is very severe). The maximum for diary card symptom score was 72. Sneezing, discharge and obstruction were graded on a scale 0–3 and recorded day and night to give a maximum score of 18 per day. From these recordings over a 4-week period, a mean daily score for each week was calculated (maximum = 18) and then a total score for the whole period was obtained (maximum = $4 \times 18 = 72$). Nasal challenge was expressed as cumulative score with a maximum of 38 (derived from the concentration of *D. pteronyssinus* to give a 20% or greater fall in nasal peak expiratory flow (PEF), degree of sneezing, degree of secretions and actual % fall in nasal PEF (5)).

The skin prick test (to 100 000 BU/ml Pharmedgen *D. pteronyssinus*) was expressed as the wheal diameter (mm) after subtraction of the wheal diameter to the diluent (saline – human albumin).

Drug score was calculated over a monthly period from daily records on diary cards. Each of the drugs allowed for rhinitis (beclomethasone nasal spray or the short-acting oral antihistamines, terfenadine or chlorpheniramine) was allocated a score so that doses arbitrarily "equivalent" clinically had the same score (beclomethasone 2 puffs to each nostril three times daily had the same score as terfenadine 60 mg

twice daily). A mean daily drug use and hence score was derived from 1 month's records (9).

Clinical Index

The derivation of the CI is illustrated below. The reasons for this approach are discussed later.

- 1) The magnitude (+ or –) of change from baseline response for five clinical measurements (visual analogue symptom score, diary card symptom score, nasal challenge, skin prick test and drug score) was calculated for each patient.
- 2) These changes were ranked for each parameter (at both 3 and 12 months) for all patients who entered the trial ($n = 80$).
- 3) The mean rank of the five parameters was calculated for each patient at both 3 and 12 months. This was termed the Clinical Index and had a range of 1 (worse) to 80 (improved) with no change lying somewhere between. NB. Where data were missing a rank equivalent to that for no change was allocated, but was not used to calculate the CI.
- 4) The median CI was calculated for each treatment group at 3 and 12 months.
- 5) The baseline for the CI was determined by calculating the median rank for those patients who showed no change in clinical response after 3 and/or 12 months treatment. (See below).

Baseline calculation

The value of the mid interval of the rank for no change from baseline levels for each clinical parameter at 3 and 12 months was determined, e.g. for visual analogue at 3 months:

Change	Rank
3	20
2	19
0	16
0	16
0	16.....mid interval = 16
0	16
0	16
-1	12.5
-1	12.5

From this figure and equivalent ones (mid intervals) derived for each of the clinical parameters at 3 and at 12 months, the median was determined

e.g.	Nasal chal.	Visual analog.	Skin test	Symptom score	Drug use	Median rank "0" (CI)
3 m	32.5	16	30	31	30.5	= 27.0
12 m	27	22.5	23	23.5	34	

All 10 values were used to calculate the baseline Clinical Index (median rank).

Statistical methods

Median ranks were analysed by the non-parametric Kruskal-Wallis test and group ranks by Spearman's rank correlation test (1). Where clinical follow-up data were not available, sample sizes (n) in the control groups were necessarily reduced. This is indicated by lower n values than those stated in the methods section above.

Results

Some data on clinical outcome has been reported. A preliminary report of the first 38 patients entering the study showed Pharmalgen to be significantly better than placebo after 3 month's treatment (visual analogue, nasal challenge and skin prick test all improved significantly) (5). In a paper examining the role of IgG subclasses during immunotherapy, some data was presented on clinical outcome, in all patients, showing Pharmalgen was effective at 12 months (9). However, in this larger study, where five parameters were measured, some parameters gave conflicting results.

Clinical Index

Fig. 1 shows the CI (median $\pm$ 95% confidence intervals) in the three treatment groups during immunotherapy. This parameter showed Pharmalgen to be of clear clinical benefit with CI values increasing from (median) 27.0 (see Methods) pretreatment, to

42.5 ($n = 28$) after 3 months and 49.5 ($n = 28$) after 12 months treatment, compared with the placebo control of 32.5 ($n = 30$) at 3 months and 32.0 ($n = 29$) at 12 months (Kruskal-Wallis 2-way analysis of variance $p = 0.002$ and $p < 0.001$, at 3 and 12 months, respectively). It was also significantly more effective than Allpyral treatment where the CI was 36.5 ($n = 20$) at 3 months and 38.0 ($n = 17$) at 12 months ($p = 0.006$ over 12 months). Treatment with Allpyral caused some improvement but this was not significantly different from placebo ($p = 0.15$ over 12 months).

Individual parameters

Visual analogue, skin prick test and nasal challenge in the three treatment groups during treatment are shown in Fig. 2. Significant improvement occurred in all three parameters in the Pharmalgen group. The visual analogue score improved considerably in the placebo group, but there was no significant change in nasal challenge or skin prick test in these patients. In the Allpyral group, nasal challenge improved significantly compared with placebo, but visual analogue and skin prick test did not. Drug score fell from (mean $\pm$ SE) 1.42 ± 0.42 to 0.19 ± 0.12 at 12 months in the Pharmalgen group, but remained steady in the Allpyral group (0.94 ± 0.29 and 1.05 ± 0.57 at 0 and 12 months) and placebo groups (1.28 ± 0.55 and 0.96 ± 0.37 at 0 and 12 months). Symptom score fell in the Pharmalgen group (30.9 ± 4.0 and 16.0 ± 2.7 at 0 and 12 months), remained unchanged in the Allpyral group (14.7 ± 4.1 and 14.5 ± 3.2 at 0 and 12 months) and fell, but not significantly, in the placebo group (28.8 ± 3.5 and

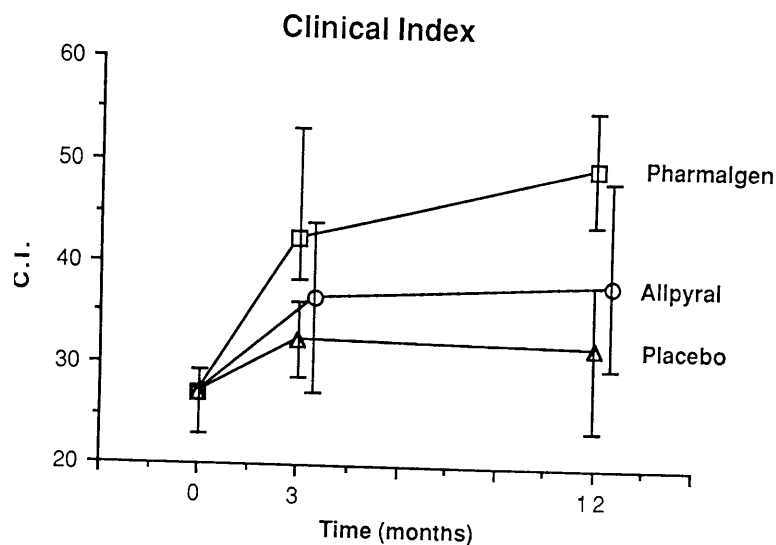


Fig. 1. The clinical index (CI) (median $\pm$ 95% confidence intervals) derived from five separate assessments to measure the efficacy of therapy in three groups treated with Pharmalgen *D. pteronyssinus*, Allpyral *D. pteronyssinus* or placebo for 12 months. Allpyral measurements at 3 and 12 months are displaced for clarity. The CI ranges from 1–80, and a rise equals improvement.

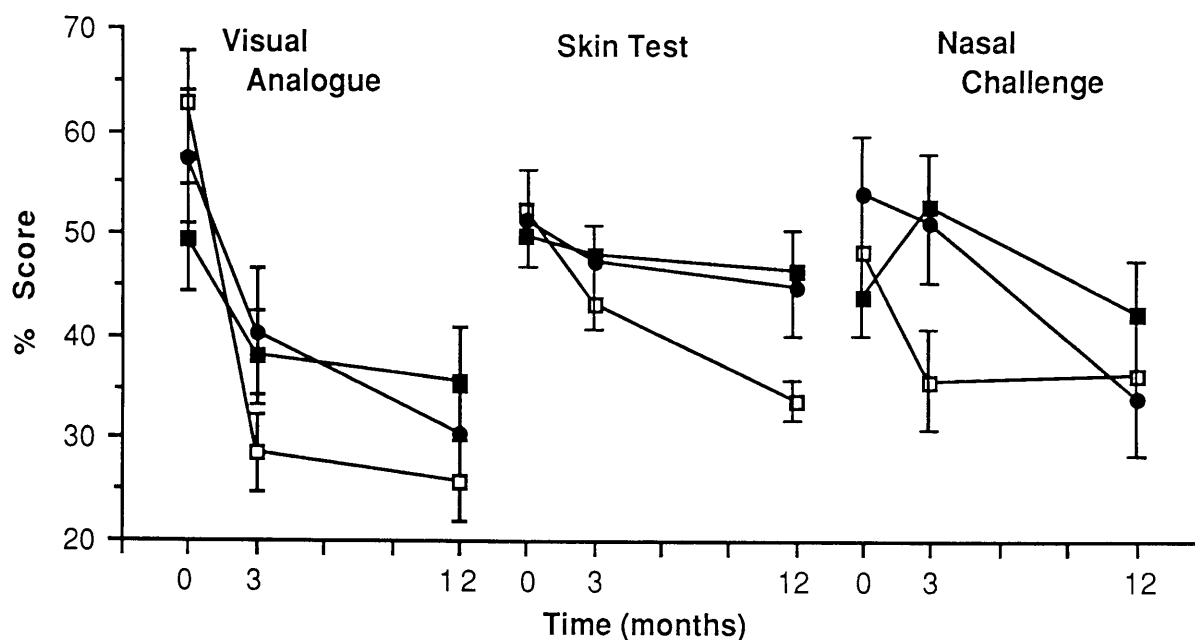


Fig. 2. Symptoms assessed by visual analogue score, skin prick test and nasal challenge test (mean $\pm$ SD) expressed as % maximum score in three groups during 12 months treatment. (□ = Pharmalgen *D. pteronyssinus*; ● = Allpyral *D. pteronyssinus*; ■ = placebo). For the derivation of visual analogue and nasal challenge test see text. For simplicity of graphical expression skin prick test is shown here as % maximum observed score.

22.4 $\pm$ 2.4 at 0 and 12 months). This data is shown fully elsewhere (9).

Comparison of clinical parameters

Table 1 shows Spearman's rank analysis of correlations of clinical parameters with each other and with the CI. Significant correlations between individual parameters and between each parameter and CI are summarized in Table 2.

When comparing the individual parameters, the visual analogue showed significant correlations with all other parameters except the skin test. More variable correlations were seen between the other parameters (Table 2), but drug score correlated with three parameters at one or both time points, and nasal challenge and symptom score with only two. Nasal challenge, drug score and symptom score correlated poorly with the other parameters, with the exception of visual analogue. Skin test results showed

Table 1. Comparison of clinical parameters using Spearman's rank correlation test. The numbers shown are the correlation coefficients (*r*)

		Visual analogue		Skin test		Nasal challenge		Drug score		Symptom score		Clinical index	
		3 mo	12 mo	3 mo	12 mo	3 mo	12 mo	3 mo	12 mo	3 mo	12 mo	3 mo	12 mo
Visual analogue	3 m												
	12 m		0.654**										
Skin test	3 m	0.055											
	12 m		0.146		0.482**								
Nasal challenge	3 m	0.233*		0.022			0.540**						
	12 m		0.207*		0.176								
Drug score	3 m	0.411**		0.124		0.266**			0.392**				
	12 m		0.307**		0.132		0.112						
Symptom score	3 m	0.282**		-0.096		0.038		0.049			0.520**		
	12 m		0.173		0.068		0.143		0.448**				
Clinical index	3 m	0.717**		0.397**		0.525**		0.580**		0.389**			0.571*
	12 m		0.583**		0.480**		0.599**		0.564**		0.548**		

The significance of the correlation is indicated at $p < 0.05$ (*) or $p < 0.01$ (**).

Table 2. Summary of significant correlations between individual parameters and between parameters and Clinical Index

	VA	NC	DS	SS	SPT
CI	++	++	++	++	++
VA		++*	++	+	-
NC	++*		+	-	-
DS	++	+		+	-
SS	+	-	+		-
SPT	-	-	-	-	

VA=visual analogue score; NC=nasal challenge; DS=diary card drug score; SS=diary card symptom score; SPT=skin prick test; CI=Clinical Index.

+ = Significant at either 3 or 12 mo (at 1% level, except for * at 5% level).

++ = Significant at both 3 and 12 mo. For p values see Table 1.

no significant correlations with any of the individual parameters at either 3 or 12 months.

The CI correlated well ($p < 0.01$) with all parameters at both 3 and 12 months. The lowest correlation coefficients (r values) were recorded between CI and diary card symptom scores and CI and skin test at 3 months, although these still gave p values of < 0.01 . High correlation coefficients were seen between CI and visual analogue, nasal challenge and drug score.

Comparison of single parameters between 3 and 12 months showed significant correlations throughout.

Discussion

Assessment of the effect of allergen immunotherapy is not straightforward, as a number of different parameters can be measured, either objectively or subjectively. Different assessments can show conflicting effects, making it difficult to know whether improvement has occurred (3). A system of combining results, without bias towards any individual parameter, is therefore proposed, and this offers a clear way of summarising outcome.

An alternative to such a multi-parameter analysis would be to consider a single parameter alone. In this case, one would have to select symptoms, since, to conclude a study was of benefit it would seem mandatory to show improvement in symptoms. This is still not without problems, as symptoms may be assessed in different ways, e.g. visual analogue or diary card symptom scores. In addition, a proper assessment of symptoms can only be made if changes in drug therapy are taken into consideration. For example, in a study where a group of patients showed little or no change in symptoms, but recorded a substantial decrease in concomitant drug therapy, analysis of symptoms alone would be misleading. When combining symptom and drug scores the weighting of each parameter is not straightforward.

There is no way of knowing which parameters

should be included in the Clinical Index. Of the parameters selected, symptom assessment (diary card symptom score and visual analogue score) and drug use are clearly relevant to disease activity. Although nasal challenge tests and skin prick tests with specific allergen might be considered less directly relevant to the clinical state, there is good evidence that these parameters improve with successful immunotherapy (2, 5). Indeed, the European Academy of Allergology and Clinical Immunology position papers on immunotherapy and on skin tests suggest that skin tests should be monitored during immunotherapy and show that good correlation exists between changes in sensitivity in the shock organ and the skin (8, 10).

In such a multi-parameter trial, each individual provides observations on several variables. The closeness of association between any two of these variables could be measured by the product-moment correlation coefficient r . However, this is not a satisfactory statistic for multi-parameter comparison for three reasons. First, the product-moment correlation is based on the concept of closeness to linear regression and its value could be affected by non-linear data transformations. Second, some of the measurements are qualitative and cannot be assigned comparable quantitative scales. Third, normality of sampling variation cannot be assumed for all parameters, particularly drug scores, and a distribution-free analysis is necessary. For these reasons a ranking analysis has been chosen where the correlation coefficient depends only on the ranks of the observations for each parameter. Therefore, the degree of improvement, termed the Clinical Index (CI), was assessed by rank sum and rank correlation analysis.

In the present study, there were significant correlations between all five parameters and the derived CI. That some correlation was found is perhaps not surprising, since the CI was derived from these parameters and because the parameters used were those likely to reflect disease activity. These parameters would, therefore, be likely to move in the same direction, but not necessarily to the same extent or at the same rate. That the correlation between the CI and each parameter was at the 1% level at both 3 and 12 months demonstrates that the CI reflects the various changes recorded well.

Significant correlations were not always found between individual parameters. Of these, visual analogue showed the best correlations with other parameters. The only parameter visual analogue did not correlate with was the skin prick test. In fact, the skin test was the only parameter which did not correlate with any other parameter. This was despite the fact that skin sensitivity in the Pharmed group improved significantly compared with placebo over 12 months (Fig. 2) as did the CI. The rate of im-

provement in skin sensitivity can be crudely judged by the time profile of response and can be seen to differ from the rate of improvement in nasal symptoms (assessed by visual analogue) and the rate of improvement in CI (Figs. 1 and 2). The latter show a similar profile over the 12-month treatment period.

In the Pharmedgen group all parameters showed improvement compared with placebo, as did the CI. In the Allpyral group only the nasal challenge improved significantly. The CI rose but not significantly. Thus, the CI distinguished between the treatments which were very effective (Pharmedgen) or at best produced a small but not significant improvement (Allpyral).

Of the two methods of assessing symptoms (diary cards and visual analogue), significant correlation was seen between these over the first 3 months ($p < 0.01$) but not over 12 months. Visual analogue correlates better with other parameters than diary card symptom score. In a study of hay fever, D'Souza et al. (4) compared three methods of measuring the effects of treatment: symptoms assessed by visual analogue, a post-treatment questionnaire and nasal air flow resistance. Four treatments (topical steroids, topical cromoglycate, oral antihistamine and oral placebo) were given, each for a 14-day period. They concluded the visual analogue score was the most valuable assessment as this, in addition to showing all active treatments were better than placebo, was sensitive enough to show differences between the active treatments. In contrast, post-treatment questionnaires to assess symptoms showed significant improvement in the actively treated groups, but did not differentiate these. Their conclusion that the visual analogue was more sensitive assumes that the active treatments did indeed cause differential improvement. It is of interest that in our study the visual analogue showed the best overall correlations, both with the derived CI and with other methods of assessment.

Bousquet (2) in a trial of immunotherapy with grass pollen, compared different parameters used to assess immunotherapy. In this study, in the actively treated group, significant improvement was found in symptom-medication scores, nasal challenge and skin tests, and a significant rise in grass pollen IgG occurred. Bousquet did not attempt to combine parameters, but did find significant correlations between nasal challenge and symptom score ($p < 0.005$ Wilcoxon W test) and skin prick test and symptom

score ($p < 0.001$). No significant correlation was found between symptom score and either specific IgG or specific IgE.

This approach using a derived CI, could be adapted for other studies, using different parameters for assessment if appropriate. This is likely to be of particular value in long studies, such as immunotherapy trials, where varying doses of concomitant therapy are used, and where spontaneous fluctuation in symptoms is likely to occur.

Acknowledgements

P.W.E. is supported by the Medical Research Council and S.M.McH. By the National Asthma Campaign. We are grateful to Dr Jane Wadsworth, Senior Lecturer in Medical Statistics, St Mary's Hospital Medical School, for help with statistical analysis.

References

1. ARMITAGE P, BERRY G. Statistical methods in medical research. 2nd ed. Oxford: Blackwell Scientific, 1987.
2. BOUSQUET J, MAASCH H, MARTINOT B, HEJJAOU A, WAHL R, MICHEL FB. Double-blind, placebo-controlled immunotherapy with mixed grass pollen allergoids. II. Comparison between parameters assessing the efficacy of immunotherapy. *J Allergy Clin Immunol* 1988; 82: 439-46.
3. D'SOUZA MF, PEPYS J, WELLS ID, et al. Hyposensitization with *Dermatophagoides pteronyssinus* in house dust allergy: a controlled study of clinical and immunological effects. *Clin Allergy* 1973; 3: 177-93.
4. D'SOUZA MF, EMANUEL MB, GREGG J, CHARLTON J, GOLDSCHMIDT J. A method for evaluating therapy for hay fever. A comparison of four treatments. *Clin Allergy* 1983; 13: 329-35.
5. EWAN PW, ALEXANDER MM, SNAPE C, IND PW, AGRELL B, DREBORG S. Effective hyposensitisation in allergic rhinitis using a potent partially purified extract of house dust mite. *Clin Allergy* 1988; 18: 501-8.
6. EWAN PW. Allergen immunotherapy. *Curr Opin Immunol* 1989; 1: 672-8.
7. GODFREY K. House dust mite avoidance - the way forward. *Clin Exp Allergy* 1991; 21: 1-2.
8. Immunotherapy subcommittee of the European Academy of Allergology and Clinical Immunology. Immunotherapy position paper. *Allergy* 1988; 43: (Suppl 6), 1-33.
9. MCHUGH SM, LAVALLE B, KEMENY DM, PATEL S, EWAN PW. A placebo controlled trial of immunotherapy with two extracts of *D. pteronyssinus* in allergic rhinitis, comparing clinical outcome with changes in specific IgE, IgG and IgG subclasses. *J Allergy Clin Immunol* 1990; 86: 521-32.
10. Subcommittee on skin tests of the European Academy of Allergology and Clinical Immunology. Skin tests used in type I allergy testing. Position paper. *Allergy* 1989; 44: (Suppl 10) 1-59.