

‘Calibration’ of our patch test reading technique is necessary

Cecilia Svedman¹, Marlene Isaksson¹, Jonas Björk², Martin Mowitz¹ and Magnus Bruze¹

¹Department of Occupational and Environmental Dermatology, Skåne University Hospital, Malmö, Lund University, 205 02 Malmö, Sweden and

²Competence Centre for Clinical Research, Skåne University Hospital, Lund, Lund University, 205 02 Malmö, Sweden

doi:10.1111/j.1600-0536.2011.02044.x

Summary

Background. Patch testing has been used for more than 100 years as the best means to diagnose contact allergy. The increased use of the technique makes standardization of the methodology most important. Progress in the standardization of the patch test technique has been made through guidelines.

Objectives. To investigate the intra-individual and inter-individual differences in reading of patch tests in order to validate the actual patch test reading when reading is performed according to the same guidelines, and to determine whether information on guidelines actually improves the outcome of the patch test readings

Patients/materials/methods. A patch test course for dermatologists was performed. Volunteers were patch tested with irritants and allergens in different concentrations. The patch test reader did not know what substances had been tested.

Results. Irritant reactions/doubtful reactions and weak allergic reactions are difficult to differentiate (accordance for these reactions 21%, 19%, and 44%). Information and education have a beneficial effect on results, as shown by an improved kappa value.

Conclusions. Standardization of patch test reading is important. Information on the classification system resulted in a significant improvement. Interpretation of guidelines should be continuously evaluated.

Key words: doubtful reactions; guidelines on patch test reading; irritant reactions; validation of patch test reading.

Patch testing has been used for more than 100 years as the best means to diagnose contact allergy. Several *in vitro* techniques have been tried, but the clinician is still almost totally dependent on patch testing for diagnosing contact allergy. The increased use of patch testing and the standardization of the methodology have actually been described as some of the major advances in clinical dermatology in the 20th century (1), ensuring that patients receive correct information about diagnosis and possible prevention. Regarding reproducibility (2–4), investigations have shown fairly good results, even

though these may differ (5). Also on comparison of accordance between dermatologists reading patch tests the results have been good, but show certain difficulties in discrimination between morphologies that are important for classification in patch testing (6).

However, the technique has certain weaknesses. Over the years, different patch test traditions and reading systems have evolved, and are used unequally in different countries (7). This makes results difficult to compare. Standardization is thus difficult to perform, a factor that is becoming important as multicentre studies are becoming increasingly common and important (8). The weaknesses of the patch testing technique can be divided into problems related to technical factors and human factors. Technical factors include the use of different chambers (9), vehicles (10), occlusion techniques, tapes, exposure times, reading days, and quality of patch test preparations. Human factors comprise dispensing

Correspondence: Cecilia Svedman, Department of Occupational and Environmental Dermatology, Skåne University Hospital, Malmö Lund University, 205 02 Malmö, Sweden. Tel: 0046-40-331000; Fax: 0046-40-331192. E-mail: Cecilia.Svedman@skane.se

Conflicts of interest: The authors have declared no conflicts.

Accepted for publication 14 November 2011

Table 1. Recording of patch test reactions according to the ICDRG^a

?+	Doubtful reaction; faint erythema only
+	Weak positive reaction; erythema, infiltration, possibly papules
++	Strong positive reaction; erythema, infiltration, papules, vesicles
+++	Extreme positive reaction; intense erythema and infiltration and coalescing vesicles
–	Negative reaction
IR	Irritant reaction of different types
NT	Not tested

ICDRG, International Contact Dermatitis Research Group.

^aFregert S (1981) *Manual of Contact Dermatitis*, 2nd edn. Munksgaard, Copenhagen.

of petrolatum preparations in a test chamber (11, 12) and reading the patch test as such.

In this study, we focus on the importance of validation of patch test readings, an issue that has been addressed before (2, 7, 13) but that may need to be readdressed (8, 14, 15).

The actual patch test reading is based on inspection and palpation of the test area, suggesting that the assessment is subject to the reader's knowledge and experience. The classification of the reaction can be made according to several models, where the grading of different reaction types differs (6).

In this study, all readings were performed according to the guidelines developed by the International Contact Dermatitis Research Group (ICDRG) (Table 1) (16), implying that one possible methodological problem would thus be ruled out. The goal when patch test reading is that the assessment should be performed strictly according to the morphology of the patch test reaction (10), and therefore the participants had no knowledge of the actual substances giving rise to the reaction, in order to rule out this possible source of bias. The aim of the present study was to investigate the intra-individual and inter-individual differences in the reading of patch tests in order to validate the actual patch test reading process. Another aim was to determine whether a lecture focusing on the necessary morphological features for each scale in the classification system would actually improve the outcome. The study was performed as a patch test reading course in which the participating dermatologists actually had the possibility of testing their abilities on real patients tested prior to the course.

Material and Methods

Study design

Eleven dermatologists from several European countries and six patients (I–VI) participated in the study (Fig. 1). The patients were patch tested 3 days prior to the test

reading. A senior dermatologist ('gold standard'), with more than 20 years of experience of routine patch test reading, in our department read all the patch test reactions of all the patients according to the 1981 ICDRG guidelines (16) 1 hr before the participants. The gold standard read the patch tests using ordinary protocols, that is, with the different preparations tested defined with concentrations and vehicle. Reactions were read according to the morphology; that is, a reaction from a substance known to be an irritant (see below) was read as allergic if the morphology was consistent with an allergic reaction.

The patch test course

The participating dermatologists were informed that readings had to be performed without knowledge of the actual substances tested, but that the substances tested were from the European baseline series, including well-known fragrances and preservatives, and test preparations representing the patient's own products. They were told to read the tests according to the same ICDRG guidelines (16). The patients were divided into two groups with 3 patients each. Each participating dermatologist read the same 3 patients on two different occasions, before and after a lecture on patch test reading.

Patients

Six patients (5 females, 1 male, mean age 43 years) with known contact allergy to several substances found in the European baseline series (Table 2) participated voluntarily in the study. The patients, all dermatitis patients, had been tested 2–18 months prior to the study. The patients were covered with a green cloth except for the test area on the back, so that the dermatologist would not recognize the patient.

Preparations used for patch testing

The preparations used, with contact sensitizers and irritants and the suppliers of these chemicals, are described in Table 2. For each patient, according to the most recent patch test result, different dilutions were used for each preparation tested in order to enhance the chance of obtaining different strengths of the patch test reactions, preferably from no reaction to a +++ reaction.

Commercial test preparations of fragrance mix I (FM I), fragrance mix II, cinnamal, cinnamyl alcohol, hydroxyisohexyl 3-cyclohexene carboxaldehyde, *Evernia prunastri* extract, *Myroxylon pereirae* resin, nickel

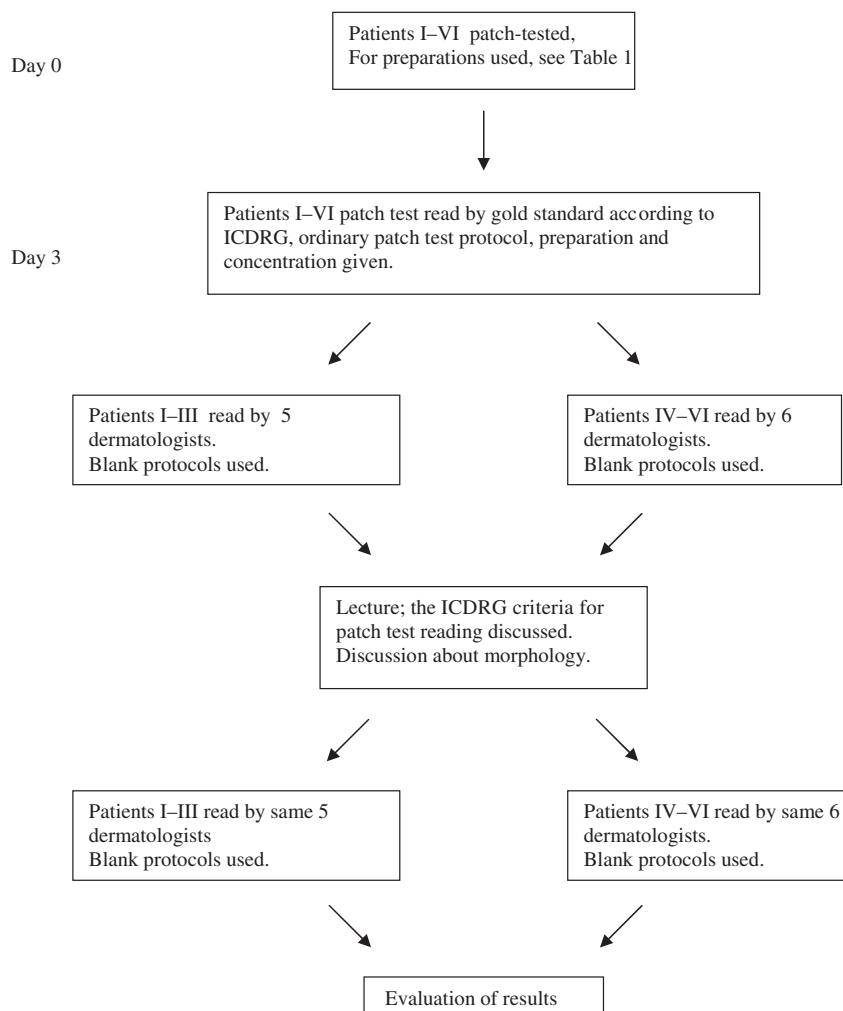


Fig. 1. The logistics of the patch test course.

sulfate, palladium chloride and methylchlorisothiazolinone/methylisothiazolinone (MCI/MI) (Chemotechnique Diagnostics®, Vellinge, Sweden) were used for the highest concentrations in the dilution series.

The preparations of FM I and palladium chloride were further diluted in pet., and the preparation of MCI/MI was further diluted in water. Dilutions corresponding to the lower concentrations of cinnamal (Bedoukian Research, Danbury, CT, USA), cinnamyl alcohol (DSM Special Products, Sittard, The Netherlands), hydroxyisohexyl 3-cyclohexene carboxaldehyde (International Flavors & Fragrances, Union Beach, NJ, USA) and *M. pereirae* resin (Apoteksbolaget, Stockholm, Sweden) were made in acetone, whereas *E. prunastri* extract (Robertet, Grasse, France) was diluted in ethanol, and nickel sulfate (Acros Organics, Geel, Belgium) was diluted in water.

Serial dilutions of formaldehyde (Acros Organics) and sodium lauryl sulfate (Sigma, St Louis, MO, USA) were made in water, and serial dilutions of nonanoic acid

(Sigma-Aldrich, Steinheim, Germany) were made in isopropanol.

All patients were patch tested with their own allergens in different concentrations, and also with the known irritant substances nonanoic acid and sodium lauryl sulfate. The latter two substances were tested in dilution series, based on previous testing on two controls, to ensure that the reactions would clearly represent irritant and doubtful reactions and where some reactions would be indistinguishable from true allergic reactions (Table 3). In all, each patient had 40 Finn Chambers® applied, but only 16–26 of the chambers in each patient had preparations with sensitizers or irritants in them, whereas the remaining chambers were applied empty.

Patch testing

Finn Chambers® (Epitest Ltd Oy, Tuusula Finland) attached to Scanpor® tape (Norgesplaster, Vennesla,

Table 2. Substances tested, test concentration interval, and the distribution of test preparations between patients

Substance concentration	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
Fragrance mix I						
8.0% pet.	X		X	X		
4.0% and 2.0% pet.	X		X			
Fragrance mix II						
14.0% pet.		X				X
Cinnamal						
1.0% pet.	X					
0.32%, 0.1%, 0.03% and 0.01% acetone	X					
Cinnamyl alcohol						
2.0% pet.	X					
0.63%, 0.2%, 0.06% and 0.02% acetone	X					
Hydroxyisohexyl 3-cyclohexene carboxaldehyde						
5.0% pet.		X				X
1.6%, 0.5%, 0.16% and 0.05% acetone		X				X
Evernia prunastri extract						
2.0% pet.			X	X		
0.63%, 0.2%, 0.06% and 0.02% ethanol			X	X		
Myroxylon pereirae resin						
25.0% pet.				X	X	X
12.5%, 6.3%, 3.1% and 1.6% acetone						X
2.5%, 0.25%, 0.025% and 0.0025% acetone				X	X	
Nickel sulfate						
5.0% pet.					X	
1.6%, 0.5%, 0.16% and 0.05% aqua					X	
Palladium chloride						
2.0%, 1.0%, 0.5%, 0.25% and 0.13% pet.					X	
Methylchloroisothiazolinone/methylisothiazolinone						
200, 100, 50, 25 and 12.5 ppm aqua		X				
Formaldehyde						
2.0%, 1.0%, 0.3%, 0.1% and 0.03% aqua		X		X		X
Nonanoic acid						
80.0%, 40.0%, 20.0%, 10.0% and 5.0% isopropanol	X	X	X	X	X	X
Sodium lauryl sulfate						
2.0%, 1.5%, 1.0%, 0.5% and 0.125% aqua	X	X	X	X	X	X

Norway) were used. For aqueous solutions, 15 µl was micropipetted on the Finn Chamber® disc (Ø = 8 mm) containing a filter paper, and for chemicals in pet., 20 mg was applied (11). The test preparations were randomly distributed on the test chambers, which were applied on the upper back. The Finn Chambers® were removed by the patients after 48 hr.

The patch test reader

All 11 patch test readers were dermatologists with previous experience of the ICDRG guidelines (8, 16). Their experience of patch test reading differed, but all but one had read patch tests on a regular basis for more than 1 year and had read patch tests on a regular basis within the last 3 years. Because the dermatologists differed in experience, even among the more experienced doctors, we also compared the less experienced with the

senior group (see Statistics), and found no significant differences. As the number of doctors was limited, and previous investigations have not shown major differences when comparing more experienced with less experienced readers (6), the doctors were regarded as equal in this respect, and were therefore not divided into groups. When performing the patch test reading, the dermatologist was provided with a pen, a plastic reading plate used for Finn Chambers®, a magnifying glass, and an empty protocol.

Test reading

Patch test reading was performed on day 3, that is, the day after removal of the patch tests. The 1981 ICDRG guidelines (Table 1) (16) were used.

The dermatologist was alone with the patient when reading, in a room equipped with 'daylight' lamps, giving the best possible environment for examining the skin.

Table 3. Accordance between the gold standard and the other dermatologists in the assessment of morphologically allergic reactions

Readings by the gold standard	Readings of the 11 dermatologists, %					
	+ reaction	++ reaction	+++ reaction	Irritant reaction	Doubtful reaction	No reaction
Test reading 1						
+ Reaction (36)	44 (16/36)	22 (8/36)	5 (2/36)	3 (1/36)	22 (8/36)	3 (1/36)
++ Reaction (48)	50 (24/48)	38 (18/48)	6 (3/48)	0	2 (1/48)	4 (2/48)
+++ Reaction (103)	17 (17/101)	41 (41/101)	39 (39/101)	0	0	6 (6/101)
No reaction (852)	0.2 (2/852)	0.2 (2/852)	0.3 (3/852)	0.3 (3/852)	0.1 (1/852)	99 (841/852)
Test reading 2						
+ Reaction (36)	44 (16/36)	28 (10/36)	0	8% (3/36)	14 (5/36)	5 (2/36)
++ Reaction (48)	58 (28/48)	27 (13/48)	2 (1/48)	2 (1/48)	8 (4/48)	2 (1/48)
+++ Reaction (103)	20 (20/101)	39 (39/101)	38 (38/101)	0	2 (2/101)	4 (4/101)
No reaction (852)	0.3 (3/852)	0	0.2 (2/852)	1.1 (9/852)	0.8 (7/852)	98 (831/852)

Numbers in parentheses are the total numbers of reactions found by the gold standard and the number for the other dermatologists where accordance for the given reaction type was the same.

An assistant was present in the beginning, so that the physician would be familiar with the equipment and how to find the correct order in which to read the patch tests. No communication between doctor and patient was allowed, and the dermatologists were not allowed to communicate between themselves during each patch test session.

After reading the first patient, the dermatologist circulated to the next. This meant that the dermatologist read 3 patients during each session. After the first session, the participants were given a lecture on patch test reading techniques, with a focus on the ICDRG guidelines (Table 1). Session 2 immediately followed; in this, the dermatologists again read the same patients in the same way, but in new rooms and in a different order, to minimize the risk of recognition.

Lecture

Between reading sessions, a 25 min lecture and additional 5 min for discussion with the lecturer was given. In the lecture, emphasis was given to the importance of objectivity and the fact that there are, when grading, certain important pitfalls: the difficulties in discriminating between a doubtful reaction and + reaction, between an irritant reaction and a + reaction, and between a doubtful reaction and an irritant reaction. Special focus was placed on the morphology of the allergic (+ reaction), irritant and doubtful reactions (Table 4).

Statistics

To assess the agreement between the gold standard and the patch test readers, the kappa coefficient for dichotomized categorical items (e.g. doubtful reaction versus not doubtful reaction or irritant reaction versus not irritant reaction) was calculated. As a rule

Table 4. The focus of the lecture was on morphology of reaction type

Morphological guidance when reading on day 3/4	
Allergic morphology	Erythema covering the whole test area Infiltration/oedema covering the whole test area Papules Vesicles Bullae (Erosion—if this constitutes the remains of previous vesicles)
?+ morphology	Erythema not covering the whole test area Infiltration not covering the whole test area Few papules but no erythema/infiltration covering the whole test area
Irritant morphology	Bullae Dry skin Scaling Pustules Erosion Petechiae Shiny skin Cigarette paper structure

of thumb, kappa values from 0.41 to 0.6 indicate moderate agreement, values from 0.61 to 0.8 good agreement, and values from 0.81 to 1.00 very good agreement (17).

The Wilcoxon signed rank test was used to assess the impact of the lecture on the agreement, by comparing kappa values pairwise before and after the lecture among the 11 dermatologists. To assess whether there was a systematic effect of choosing one person as the gold standard for the calculations, the kappa value was also calculated with the different doctors as gold standards, but without important improvement. We also analysed whether there was a possible difference in the doctor group between the less experienced and the senior group,

Table 5. Results of patch test readings 1 and 2 regarding irritant, doubtful, and + reactions; accordance

Gold standard	Irritant reaction, % (no.)	Doubtful reaction, % (no.)	Allergic reaction (+), % (no.)	++/+++, no.	No reaction, % (no.)
Test reading 1					
Irritant reaction (151)	21 (32/151)	10 (15/151)	14 (21/151)	2/151	54 (81/150)
Doubtful reaction (130)	7 (9/130)	20 (25/130)	34 (42/130)	10/130	34 (44/130)
Allergic reaction (+) (36)	3 (1/36)	23 (8/36)	44 (16/36)	10/36	3 (1/36)
Test reading 2					
Irritant reaction (151)	46 (70/151)	11 (16/151)	8.6 (13/151)	1/151	34 (51/151)
Doubtful reaction (130)	2 (3/130)	44 (57/130)	20 (26/130)	1/130	33 (43/130)
Allergic reaction (+) (36)	8 (3/36)	14 (5/36)	44 (16/36)	10/36	6 (2/36)

to find whether there was a marked difference in the kappa values, but this was not the case.

Ethics

The study was approved by the Ethics Committee of the Faculty of Medicine (2008 Dnr 235/2008), Lund University, Sweden. The patients gave written consent to participate in the study.

All of the patch test readers, after having finished the patch test course, agreed to the results being used for publication (11 of 11 agreed).

Results

The results of the patch test readings were assembled according to reaction type, independently of what test preparation had caused the patch test reaction. All of the results were compared with those of the gold standard.

Test reading 1

Regarding 'no reaction', that is, where there had been no test substance in the Finn Chamber® or when the concentration of the tested preparation proved too low to induce a reaction, there was 98.7% (841/852) accordance. In 0.8% (7/852) of the cases, the reaction had been regarded as allergic in nature.

When judging whether there was an allergic reaction, that is, a + to +++ reaction, the accordance was 89.8% (168/187), 4.8% (9/187) were read as doubtful reactions, and 0.5% (1/187) were read as irritant reactions (Table 3). On analysis of these data, with the sole aim of discriminating between an allergic reaction and a non-allergic reaction, the kappa value was 0.73 in the first reading session, and if all different reaction types were included, that is, where the goal was to discriminate between all six reaction types, the kappa value was 0.48.

Irritant reactions and doubtful reactions had the lowest accordance, at 21% (32/151) and 19% (25/130),

respectively. Regarding irritant reactions, these were read as doubtful in 10% (15/151) of cases, as 'no reaction' in 54% (81/150), and as allergic in 15% (23/151). Doubtful reactions were read as irritant reactions in 7% (9/130) of cases, as 'no reaction' in 34% (44/130), and as allergic in 40.0% (52/130) (Table 5).

Test reading 2

This reading was performed after the lecture on test reading. Regarding judging whether there was a reaction or not, the accordance was the same, at 97.5% (831/852). In 0.6% (5/852) of these cases, the 'no reaction' was read as allergic. When judging whether there was an allergic reaction, that is, a + to +++ reaction, the accordance was 88.2% (165/187), and 5.9% (11/187) were read as doubtful and 2.1% (4/187) as irritant reactions (Table 3). The kappa value after the lecture regarding discrimination solely between allergic/non-allergic reactions was 0.79, and when all reaction types were taken into account, it was 0.56.

Regarding doubtful and irritant reactions, the accordance in reading had improved (Table 5). For doubtful reactions, the accordance was 44% (57/130), and for irritant reactions it was 46% (70/151). On calculation of the kappa values for the dichotomized different reaction types, there was, as expected, a difference, with the best values being found regarding strong positive reactions and negative reactions. There was an improvement in test reading 2 regarding negative reactions (the median kappa value among 11 dermatologists in test reading 1 was 0.76, and that in test reading 2 was 0.80; $p = 0.04$), doubtful reactions (median kappa 0.11 versus 0.33; $p = 0.01$), and irritant reactions (median kappa 0.18 versus 0.54; $p = 0.02$).

In patch reading 2 regarding irritant reactions, 11% (16/151) were read as doubtful, 34% (51/151) as 'no reaction', and 9% (14/151) as allergic. For doubtful reactions, 2% (3/130) were read as irritant, 33% (43/130) as 'no reaction', and 21% (27/130) as allergic.

On comparison of the values for session 1 and session 2, there was a difference, and the kappa value based on all reaction types also significantly improved, from 0.48 to 0.56 ($p = 0.03$). An improvement was also seen when the doctors were discriminating solely between allergic and non-allergic reactions, although this improvement was not significant ($p = 0.11$).

Discussion

For the individual patient, negative and positive patch test results are of importance, because they may influence the patient's life in many ways, and the advice that their doctor will give. Therefore, patch test readings must be standardized. If the reading is not standardized, it becomes impossible to improve on the interpretation of the reading. In this study, we have focused on the actual readings, because these are important for other aspects as well, such as the statistics and epidemiological interpretations derived from the results. Multicentre studies, in which several investigators in different countries perform the same task, have become increasingly common, and there are known problems with these (18). In these studies, where the results are to be pooled together, not only is it necessary for the study to be performed in the same way, but it is, of course, also presumed that measurements/readings are performed with instruments being calibrated in the same way. Patch test reading has to be carried out with great care, as it is performed not with an instrument, but with the human eye and hand. It is essential that this is also performed in a standardized way, whereby our definitions are 'calibrated' in the same way. This study was performed in order to show the importance of how we read patch tests. As reading is performed by the dermatologist and not by a machine, one can argue that the standard used for the readings has not been 'properly calibrated'. In this study, the reading of the gold standard was taken as the 'calibrated value'; this was possible because the gold standard was a clinician who had read patch tests regularly for over 20 years, and when the other clinicians were used as the gold standard, there was no statistically significant difference in the kappa value on comparison with the analysis performed with the chosen gold standard.

This study showed good agreement in discriminating whether there is a reaction or not and whether the reaction is of a clearly allergic morphology, the latter having been shown previously (6). The fact that almost the same number of 'no reactions' were read as allergic in test reading sessions 1 and 2 may give an indication of the number of reactions that the test reader actually placed in the wrong category by mistake, that is,

'misreading'. There is a clear difficulty in discriminating between doubtful and irritant reactions and in whether the physician will grade a reaction as 'no reaction' or irritant/doubtful. There is also a difficulty in grading between the doubtful reaction/irritant reaction and allergic reaction, that is, the + reaction. Doubtful reactions constituted 40% of the cases in test reading session 1 described as allergic, and 21% of the cases in test reading session 2. Irritant reactions constituted 15% of the cases read as allergic in test reading session 1, and in test reading session 2, 9% were defined as allergic reactions. The allergic reactions (+) were read as doubtful in 23% of cases, and in test reading 2, 14% were read as doubtful. The results may seem in favour of *in vitro* analysis of contact allergy but, on the whole, sensitivity of *in vitro* analysis is not better, and it is only available for selected allergens (19, 20).

Although the average changes in the kappa values were generally small, the lecture seems to have improved agreement. Here, the only educational option was a lecture, which has proven to be a rather good instrument for the transfer of information, although perhaps even better results would have been achieved with a more interactive educational tool (21).

It is important to note that there are difficulties in reading doubtful/allergic and irritant reactions in an accordant manner (note the low kappa values for these reaction types). This may have implications when the test concentrations for new substances are being defined during the planning of larger studies and, especially, multicentre studies. This problem underlines several points: the importance of finding an optimal test concentration and test vehicle when testing new substances to obtain optimal reading conditions, and the importance of actually reading test reactions without judging, that is, without being influenced by the knowledge of which sensitizer is tested or by the patient's history. The clinician should read patch tests only by following the morphological criteria for the various defined reactions according to the ICDRG. Constant work regarding standardization is of the utmost importance. In this study, we did not use a manual whereby the physician could have graded the reactions according to morphology, but we suspect that this approach would have improved the results. Discussion and education on how we read also seem to improve the concordance in reading. We would like to suggest, especially in multicentre studies, the use of manuals of the type often used for repeated open application tests (6, 22), as these force the reader to define the reaction according to morphology. Perhaps, in multicentre studies, each centre should grade the reactions only according to morphology, and the actual defining of the reaction according

to the ICDRG should be performed afterwards, either by one centre or by a computerized method. This study may also indicate the need for practical educational sessions and continuous discussion in the clinic on how reactions should be read. It has been suggested that the guidelines for reading patch tests should be changed (8), or at least changed when used for research purposes (23); in the latter case, it is suggested that doubtful reactions should be read as negative, because of the fact that the problem seems to be in discriminating between doubtful and irritant reactions, and the fact that the doubtful reaction may be a true allergic one, but where the test concentration or

application time is too low/short, we do not think that a simplified version of the ICDRG may be the solution.

Acknowledgements

We would like to thank the participants and patients who made the course possible, and the Research Institute of Fragrance Materials (RIFM), who contributed financially to the study. This research was supported by the RIFM. Professor Magnus Bruze is a member of the Expert Panel of the RIFM, which evaluates the safety of fragrance materials and is supported by the manufacturers of fragrances and consumer products containing fragrances.

References

- 1 Storrs F J. Technical and ethical problems associated with patch testing. *Clin Rev Allergy Immunol* 1996; **14**: 185–198.
- 2 Lachapelle J M. A left versus right side comparative study of Epiquick patch test results in 100 consecutive patients. *Contact Dermatitis* 1989; **20**: 51–56.
- 3 Ale S I, Maibach H I. Reproducibility of patch test results: a concurrent right-versus-left study using TRUE Test. *Contact Dermatitis* 2004; **50**: 304–312.
- 4 Lachapelle J M, Bruynzeel D P, Ducombs G et al. European multicenter study of the TRUE Test. *Contact Dermatitis* 1988; **19**: 91–97.
- 5 Gollhausen R, Przybilla B, Ring J. Reproducibility of patch tests. *J Am Acad Dermatol* 1989; **21**: 1196–1202.
- 6 Bruze M, Isaksson M, Edman B et al. A study on expert reading of patch test reactions: inter-individual accordance. *Contact Dermatitis* 1995; **32**: 331–337.
- 7 Wilkinson D, Fregert S, Magnusson B et al. Terminology of contact dermatitis. *Acta Derm Venereol* 1970; **50**: 287–292.
- 8 Menné T, White I. Standardization in contact dermatitis. *Contact Dermatitis* 2008; **58**: 321 (Editorial).
- 9 Isaksson M, Gruvberger B, Frick-Engfeldt M et al. Which test chamber should be used for acetone, ethanol, and water solutions when patch testing? *Contact Dermatitis* 2007; **57**: 132–136.
- 10 Wahlberg J E, Lindberg M. Patch testing. In: *Contact Dermatitis*, Frosch P J, Menné T, Lepoittevin J P (eds): Berlin, 2006: pp. 365–390.
- 11 Bruze M, Frick-Engfeldt M, Gruvberger B et al. Recommendation in the amount of petrolatum preparation applied at patch testing. *Contact Dermatitis* 2007; **56**: 38–42.
- 12 Antoine J L, Lachapelle J M. Variations in the quantities of petrolatum applied in patch testing. *Derm Beruf Umwelt* 1988; **36**: 191–194.
- 13 Mowad C M. Patch testing: pitfalls and performance. *Curr Opin Allergy Clin Immunol* 2006; **6**: 340–344.
- 14 Uter W, Becker D, Schnuch A, Gefeller O, Frosch P J. The validity of rating patch test reactions based on digital images. *Contact Dermatitis* 2007; **57**: 337–342.
- 15 Uter W, Frosch P, Becker D et al. Are we biased when reading a doubtful patch test reaction to a clear-cut allergen such as the thiuram-mix. *Contact Dermatitis* 2009; **60**: 234–235.
- 16 Fregert S. *Manual of Contact Dermatitis*, 2nd edition: Copenhagen, Munksgaard, 1981.
- 17 Altman D G. *Practical Statistics for Medical Research*, London, Chapman & Hall, 1991.
- 18 Andersen K E. Multicenter patch test studies: are they worth the effort? *Contact Dermatitis* 1998; **38**: 222–223.
- 19 Jakobson E, Masjedi K, Ahlborg N et al. Cytokine production in nickel-sensitized individuals analysed with enzyme-linked immunospot assay; possible implication for diagnosis. *Br J Dermatol* 2002; **147**: 442–449.
- 20 Spiewak R, Moed N, von Blomberg B M. Allergic contact dermatitis to nickel modified in vitro protocols better detection of allergen-specific response. *Contact Dermatitis* 2007; **56**: 63–69.
- 21 de Winstanley P A, Bjork R A. Successful lecturing: presenting information in ways that engage effective processing. *New Directions for Teaching and Learning*, Wiley Periodicals, Inc., 2002; **89**: 19–31.
- 22 Johansen J D, Bruze M, Andersen K E et al. The repeat open application test: suggestion for a scale of evaluation. *Contact Dermatitis* 1997; **39**: 95–96.
- 23 Ivens U, Serup J, O'Goshi K. Allergy patch test reading from photographic images: disagreement on ICDRG grading but agreement on simplified tripartite reading. *Skin Res Technol* 2007; **13**: 110–113.