

The importance of context information in the diagnostic rating of digital images of patch test reactions

W. Uter, P.J. Frosch,* D. Becker,† A. Schnuch,‡ A. Pfahlberg and O. Gefeller

Department of Medical Informatics, Biometry and Epidemiology, University of Erlangen-Nürnberg, Waldstr. 6, D-91054 Erlangen, Germany

*Department of Dermatology, University of Witten/Herdecke and Klinikum Dortmund gGmbH, Dortmund, Germany

†Department of Dermatology, University Hospital of Mainz, Mainz, Germany

‡Information Network of Departments of Dermatology at the University of Göttingen, Göttingen, Germany

Summary

Correspondence

Wolfgang Uter.

E-mail: wolfgang.uter@imbe.med.uni-erlangen.de

Accepted for publication

19 January 2009

Key words

contact allergy, continuing medical education, digital image, patch testing, reliability

Conflicts of interest

None declared.

DOI 10.1111/j.1365-2133.2009.09167.x

Background Patch test (PT) guidelines recommend classifying PT reactions based on morphological criteria only, i.e. irrespective of context information such as substance, concentration and vehicle.

Objectives To analyse reclassification of PT reactions after revelation of context information subsequent to an initial blinded reading.

Methods One hundred and twenty-two participants (experienced dermatologists and researchers) rated 20 digital images of different PT reactions twice, first blinded, then unblinded regarding substance, concentration and vehicle. Agreement between both ratings was quantified with Cohen's kappa, and systematic differences statistically examined with tests for marginal homogeneity.

Results Mostly, ratings remained stable, e.g. in > 90% of cases of images showing typical strong or extreme positive PT reactions. Reclassification was comparatively often observed in images depicting irritant reactions. Conversely, 16 of 122 participants re-rated doubtful reactions to thiuram mix and dichromate, respectively, as 'allergic' (weak or strong positive) after knowing the substance.

Conclusions The considerable proportion of participants who choose an 'allergic' rating, despite a morphological presentation definitely not justifying this, points to a conceptual problem of the PT reading scale: the mixing up of morphological classification and interpretation. We therefore suggest amending the scale. Moreover, standardization of PT reading can be improved by continual PT training sessions.

The patch test (PT) is more than a century old, but is still an indispensable tool for the diagnosis of contact allergy. It has often been stated that while the application of a PT requires care and technical skill, the reading of PT reactions constitutes the crucial and most difficult step. In fact, the visual reading of a PT, strictly based on morphological criteria,¹ is only the first of three successive steps. It is followed by interpretation as allergic vs. nonallergic (irritant or unknown), and finally the evaluation of clinical relevance in those reactions deemed allergic.² Clear as this procedure may appear, it is often not applied in this sequential form in clinical practice. Unfortunately, the traditional PT reading scale itself hampers the distinction between, and adequate succession of, the first two steps by mixing morphological criteria with interpretation (Table 1).

The present analysis of data collected from participants at a standardized slide show of digital images of PT reactions

examines whether context information has any impact on the classification of PT reactions, compared with an initial blinded rating. Context information for PT reading in this study means knowing the substance tested and its concentration and vehicle. In more general terms, beyond the standardized, but artificial, study setting, context information may also include other pertinent factors, such as (past) atopic dermatitis or 'irritating' meteorological conditions important for some allergens.³ Specifically, the aim of the study is to yield evidence on whether or not knowledge of what is being tested may compromise the strict application of objective criteria mandatory in the first step.

Materials and methods

The images and the method of data collection have been published elsewhere.⁴ Additionally, the images are available

Table 1 Presently used patch test reading scale (according to Wahlberg and Lindberg¹)

Symbol	Definition	Comment
neg. or –	Negative reaction	
IR	Irritant reactions of different types	Denotes interpretation only, not morphology
? or ?+	Doubtful reaction; faint erythema only	Description of morphology without interpretation as allergic or otherwise
+	Weak positive reaction; erythema, infiltration, possibly papules	'Positive' implies allergic reaction (i.e. interpretation), the severity of which is morphologically classified in this ordinal part of the overall scale
++	Strong positive reaction; intense erythema, infiltration, papules, vesicles	
+++	Extreme positive reaction; intense erythema and infiltration and coalescing vesicles	
NT	Not tested	

as part of an online PT training tool (<http://www.ivdk.gwdg.de/dkg/>, go to 'Patch Test Reading Course') freely accessible to dermatologists after one-time registration. Briefly, participants at the 8th Congress of the European Society of Contact Dermatitis 2006 in Berlin (<http://www.escd.org>) and the subsequent German satellite meeting were invited to take part in a quiz with 20 digital images of PT reactions compiled by the European Environmental and Contact Dermatitis Research Group. Each image was first rated blind to the test preparation, with only information on presence or absence of palpable infiltration given. After the participant had entered his or her rating, the same image was presented again, but with information on substance, concentration and vehicle, to be rated a second time, before the quiz proceeded to the next image. Backwards navigation, to review or change the initial rating, was disabled. The pairs of ratings were cross-tabulated and the following measures were calculated, stratified per image (Table 2): (i) the distribution of unblinded ratings, along with the number of participants who had chosen the same respective rating at the blinded reading; (ii) statistical assessment of systematic changes between the two ratings, using the full reading scale (P-value of the test for marginal homogeneity⁵); and (iii) quantification of agreement between the two ratings, using the full reading scale, by giving the crude percentage of overall agreement and, in addition, Cohen's kappa coefficient addressing chance-corrected agreement, accompanied by its 95% confidence interval (CI). The notion of chance-correction when quantifying the amount of agreement between ratings has been suggested to account for the situation frequently encountered in practice that agreement in discrete ratings is even then observed when ratings are completely independent. The amount of this chance agreement varies with the type of distribution over rating categories and can be precisely computed. Cohen's kappa measures the agreement beyond chance and yields values between 0 and 1. Kappa values above 0.8 can be interpreted as showing almost perfect agreement, values between 0.6 and 0.8 signal substantial and values between 0.4 and 0.6 moderate agreement, whereas lower values indicate poor agreement.⁶

After collapsing the full reading scale to a dichotomous outcome 'allergic' (+, ++, +++) vs. 'nonallergic' (neg., IR, ?) for both blinded and unblinded readings, the following statistics and measures, respectively, were calculated for all images (Table 3), based on the resulting 2×2 contingency tables: (i) the frequency of a change of the rating resulting in a different classification as allergic vs. nonallergic, and the frequency of unchanged ratings (unchanged in this context either means literally unchanged, or changed just within the 'allergic' or 'nonallergic' range); (ii) the exact McNemar test for symmetry (and, in this 2×2 contingency table setting, equivalently for marginal homogeneity) addressing the question whether systematic differences between re-ratings exist, i.e. whether significantly more (or fewer) participants re-rated from 'nonallergic' to 'allergic' than vice versa; and (iii) Cohen's kappa coefficient supplemented with its 95% CI, for agreement between blinded and unblinded reading.

To elucidate which individual characteristics were associated with reclassification both on the detailed level of the original scale and on the level of the dichotomized outcome, the following factors were included in subsequent bivariate analyses, their association with reclassification (yes vs. no) quantified by the prevalence ratio (PR), accompanied by its 95% CI: (i) years of experience in patch testing (categorized into three levels, according to the tertiles of the distribution of the original data: up to 5 years, 6–15 years and > 15 years); and (ii) being a dermatologist vs. a researcher (according to self-classification). For data analysis, the statistical software package SAS (version 9.1; SAS Institute, Cary, NC, U.S.A.) was used.

Results

Characteristics of the sample of 122 persons taking part in the PT slide show and results only regarding concordance between the blinded readings and the gold standard have been published in detail.⁴ Expanding these initial results, the present paper focuses on a comparison between blinded and subsequent unblinded rating.

Table 2 Rating^a of patch test reactions after test allergen has been revealed (number of observers with same rating at blinded assessment in parentheses). P-value of test for marginal homogeneity. Agreement quantified as % total agreement and with Cohen's kappa coefficient accompanied by its 95% confidence interval (CI)

Image	Allergen (gold standard)	Unblinded reading: n (n of same rating at blinded reading) [blinded reading: n [†]]						P-value	Agreement		
		Neg.	IR	?	+	++	+++		% total	κ	95% CI
1	PPD 1% (+++)	–	–	–	–	7 (4)	115 (108)	0.054	91.8	0.42	0.13–0.70
2	Fragrance mix 8% (+)	–	5 (1)	6 (5)	93 (87)	17 (12)	1 (0)	0.77	86.1	0.65	0.52–0.79
3	Kerosene (as is) (IR)	1 (0)	109 (90)	–	–	6 (5)	6 (6)	0.01	82.8	0.47	0.29–0.64
4	Shampoo 1% ^b (IR)	12 (11)	104 (94)	6 (5)	–	–	–	0.35	90.2	0.68	0.52–0.85
5	Lyrall 5% (++)	–	5 (3)	–	35 (34)	79 (66)	3 (2)	0.007	86.1	0.74	0.63–0.85
6	Detergent 10% ^b (IR)	1 (1)	111 (90)	2 (2)	5 (5)	2 (2)	1 (0)	< 0.001	82.0	0.44	0.26–0.62
7	Fragrance mix 8% (+)	–	4 (3)	1 (0)	65 (61)	50 (45)	2 (0)	0.10	89.3	0.80	0.71–0.90
8	Bufexamac 5% (++)	–	20 (19)	–	2 (2)	37 (30)	63 (46)	< 0.001	79.5	0.69	0.59–0.80
9	Cutting oil 4% ^b (+)	1 (0)	24 (6)	41 (35)	54 (42)	1 (1)	1 (0)	< 0.001	68.9	0.51	0.39–0.63
10	Tree moss 1% (+)	–	3 (3)	4 (2)	56 (53)	58 (44)	1 (0)	< 0.001	83.6	0.71	0.60–0.82
11	Disinfectant (as is) (IR)	–	110 (95)	1 (1)	3 (3)	2 (2)	6 (4)	0.038	86.1	0.50	0.30–0.70
12	NaOH 5% ^b (IR)	12 (9)	103 (68)	3 (3)	4 (4)	–	–	< 0.001	68.9	0.38	0.24–0.51
13	Potassium dichromate 0.5% (?)	22 (20)	17 (4)	57 (50)	23 (10)	3 (0)	–	< 0.001	68.9	0.51	0.39–0.64
14	MDBGN 0.3% (++)	–	1 (1)	1 (0)	23 (22)	86 (79)	11 (9)	0.16	91.0	0.81	0.70–0.92
15	Compositae mix 6% (++)	–	3 (2)	–	7 (5)	100 (98)	12 (7)	0.44	91.8	0.71	0.53–0.88
16	Thiuram mix 1% (?)	4 (0)	12 (7)	76 (57)	27 (14)	3 (0)	–	< 0.001	63.9	0.38	0.25–0.51
17	PPD 1% (++)	–	3 (2)	–	5 (4)	97 (92)	17 (14)	0.22	91.8	0.77	0.64–0.90
18	Potassium dichromate 0.5% (?)	16 (11)	34 (17)	59 (54)	10 (8)	3 (1)	–	0.018	74.6	0.60	0.48–0.72

Table 2 (Continued)

Image	Allergen (gold standard)	Unblinded reading: n (n of same rating at blinded reading) [blinded reading: n ⁴]						P-value	Agreement		
		Neg.	IR	?	+	++	+++		% total	κ	95% CI
19	Nickel sulphate 5% (+)	–	10 (6) [9]	1 (0) [2]	20 (18) [27]	82 (71) [77]	9 (6) [7]	0.23	82.8	0.67	0.55–0.80
20	Benzalkonium chloride 0.1% ^b (?)	37 (25) [26]	36 (5) [5]	42 (41) [82]	7 (5) [8]	– – [1]	–	< 0.001	62.3	0.45	0.34–0.56

^aSee Table 1. All allergens in petrolatum, except where otherwise indicated; ^bwater. PPD, p-phenylenediamine; MDBGN, methyldibromoglutaronitrile.

Table 3 Agreement regarding 'allergic' (A: +, ++, +++) vs. 'nonallergic' (nonA: neg., IR, ?) rating^a between blinded and unblinded rating. P-value of exact McNemar test. Cohen's kappa coefficient accompanied by its 95% confidence interval (CI)

Image	Allergen (gold standard)	Blinded rating → unblinded rating				Agreement		
		A → A	A → nonA	nonA → A	nonA → nonA	P-value	κ	95% CI
1	PPD 1% (+++)	115	—	7	—	0.016	n.c.	—
2	Fragrance mix 8% (+)	105	4	6	7	0.75	0.54	0.29–0.79
3	Kerosene (as is) (IR)	12	17	—	93	< 0.001	0.52	0.33–0.70
4	Shampoo 1% ^b (IR)	—	—	—	122	n.c.	n.c.	—
5	Lyrar 5% (++)	115	2	2	3	> 0.999	0.58	0.21–0.95
6	Detergent 10% ^b (IR)	8	15	—	99	< 0.001	0.46	0.25–0.68
7	Fragrance mix 8% (+)	112	1	5	4	0.22	0.55	0.23–0.87
8	Bufexamac 5% (++)	80	1	22	19	< 0.001	0.52	0.36–0.68
9	Cutting oil 4% ^b (+)	44	12	12	54	> 0.999	0.60	0.46–0.75
10	Tree moss 1% (+)	110	2	5	5	0.45	0.56	0.27–0.85
11	Disinfectant (as is) (IR)	9	11	2	100	0.022	0.53	0.30–0.75
12	NaOH 5% ^b (IR)	4	3	—	115	0.25	0.72	0.41–1.0
13	Potassium dichromate 0.5% (?)	10	1	16	95	< 0.001	0.47	0.27–0.67
14	MDBGN 0.3% (++)	120	1	—	1	> 0.999	0.66	0.04–1.0
15	Compositae mix 6% (++)	118	1	1	2	> 0.999	0.66	0.21–1.0
16	Thiuram mix 1% (?)	14	4	16	88	0.012	0.49	0.30–0.68
17	PPD 1% (++)	119	—	—	3	n.c.	1.0	n.c.
18	Potassium dichromate 0.5% (?)	11	1	2	108	> 0.999	0.87	0.72–1.0
19	Nickel sulphate 5% (+)	107	4	4	7	> 0.999	0.60	0.35–0.85
20	Benzalkonium chloride 0.1% ^b (?)	6	3	1	112	0.63	0.73	0.48–0.98

^aSee Table 1. All allergens in petrolatum, except where otherwise indicated; ^bwater. n.c., not calculable, because two (or three) of the cells in the fourfold table have zero counts (i.e. changes between blinded and unblinded ratings are focused on one or two of the four possible constellations); PPD, p-phenylenediamine; MDBGN, methyldibromoglutaronitrile.

Classification using the full patch test reading scale

As a first, descriptive approach, the distribution of both the unblinded and the blinded rating is shown in Table 2. To quantify the stability of these ratings, compared with the initial blinded assessment, the number of participants having attached the same rating to the corresponding image in the blinded reading is given in parentheses in the middle row in each cell.

Moreover, the overall proportion of agreement is given as a percentage of the total sample in a separate column. In case of clear-cut reactions such as the first image of reactions to p-phenylenediamine (PPD) or Compositae mix, the amount of correction is limited. In other cases, particularly regarding irritants, the amount of correction is less negligible, e.g. in images of kerosene, detergent or NaOH. Also in the case of an image depicting an erythematous reaction to benzalkonium

chloride, a considerable number preferred a rating as irritant after disclosure of the substance, probably prompted by the notion that this compound is basically an irritant, notwithstanding extremely rare allergic reactions, which are, however, expected more in the ++ or +++ range of PT morphology.⁷ This is reflected in significant results of the statistical test for marginal homogeneity in 12 of the 20 images, meaning that the rating changed significantly between the two readings after context information was given.

The amount of change on the full PT reading scale is, furthermore, illustrated by the percentage total agreement, which ranges from 92% in the case of the two PPD images (i.e. only 8% of the observers changed their rating) to a mere 62% for the image of benzalkonium chloride and 64% for the image showing an erythematous reaction to thiuram mix, i.e. 38% and 36% of observers changing their rating, respectively. While low values of percentage total agreement correspond well with low values of Cohen's kappa, the opposite is not true in our situation. The reason for this discrepancy is that in some cases a high proportion of agreement can be expected by chance alone due to the type of distribution over the rating categories. As Cohen's kappa is a measure of agreement beyond chance, in these cases some discrepancy may occur, most notably in the case of the first image of a reaction to PPD.

Classification as allergic vs. otherwise

As it may be argued that some reclassifications, e.g. from ++ to +++ or vice versa, are only of academic interest and are irrelevant to clinical practice, a second set of analyses focused on a dichotomized outcome: 'allergic' vs. otherwise. Three IR images were re-rated as 'nonallergic' by 11, 15 and 17 observers, respectively, after knowing the substance causing the reaction, while the other two IR images (shampoo and NaOH) did not substantially change (improve) in this regard. This kind of re-rating occurred significantly more often than the reverse, i.e. from 'nonallergic' to 'allergic'. The opposite was seen in the image of a ++ reaction to bufexamac with unusual appearance, which was (correctly) recategorized as an allergic reaction by 22 participants after revelation of the test substance. However, in two instances, namely, the images of erythematous reactions to dichromate (first image) and thiuram mix, respectively, lacking palpable infiltration, a reclassification as 'allergic' (at least +) by 16 raters was noted after they had learnt the nature of the allergen, while the reverse reclassification occurred significantly less often.

Influence of observer characteristics

Addressing the question of who will be influenced most by context information, the association between altering the original grading and the dichotomous classification, respectively, and being a dermatologist and experience in patch testing, respectively, was analysed. Mostly, nonsignificant variation without a consistent pattern was found, with the following

exceptions: dermatologists re-rated the image of a + reaction to fragrance mix significantly less often than nondermatologists (PR 0.41, 95% CI 0.17–0.97), while they were influenced by context information significantly more often in the case of the ? reaction to dichromate (first image) (PR 5.12, 95% CI 1.32–19.9) and benzalkonium chloride (PR 2.33, 95% CI 1.02–5.31). In the case of the thiuram mix image, a reclassification regarding 'allergic' vs. 'nonallergic' was significantly less often made by dermatologists (PR 0.35, 95% CI 0.16–0.75), also in the case of the image showing a reaction to the fragrance mix (PR 0.28, 95% CI 0.09–0.91). Comparing the 'most inexperienced' (in this highly selected group of participants) with the most experienced subgroup, one striking finding is a more than tenfold increased frequency of re-ratings regarding the kerosene image: PR 15.1 (95% CI 2.1–109) for full scale rating, and PR 12.9 (95% CI 1.76–95) for the dichotomous outcome. Conversely, the 'inexperienced' significantly less often re-rated the bufexamac image: PR 0.27 (95% CI 0.08–0.88) and 0.09 (95% CI 0.01–0.66), respectively.

Discussion

Using a straightforward study design we were able to quantify the impact of context information on the rating of PT reactions, along with personal characteristics partly modifying this impact. In this framework, exploratory statistical testing helped to distinguish chance variation in terms of general intraobserver variability (impairing reliability) from those systematic changes in which we were interested.

One of the most consistent findings is a substantial amount of reclassification of the rating in the case of the five images depicting irritant PT reactions. Although it could be argued that morphological criteria should have sufficed to lead observers to this rating in the first place, the revelation of the test substance (undiluted kerosene, adequately diluted shampoo, insufficiently diluted detergent and disinfectant, respectively, and weak sodium hydroxide solution) has apparently convinced a considerable number of participants to rethink their initial 'non-IR' classification – to the least extent in the image of a PT with shampoo, which is easily recognizable as irritant even without this information. Hence, as far as well-known (potential) irritants are concerned, context information seems to have a considerable impact on the rating, with experience or qualification exhibiting no overall, consistent impact. The underlying expectation may be that any reaction to a known irritant is an irritant reaction – which is not a rule without (very rare) exceptions (e.g. regarding ++ or +++ reactions to benzalkonium chloride, see above⁷).

However, as far as well-known allergens are concerned, such as the thiuram mix or potassium dichromate, a mirrored expectation seemed to have led a considerable number of observers to reclassify the reaction as 'allergic' (i.e. as + or ++), frankly ignoring the information that erythema, but no palpable infiltration, was present (thiuram mix) and that no homogeneous erythema was present, but only a few scattered papules (first image of dichromate). Both morphologies,

according to current standards, would not qualify as a + reaction. Apparently the wish to denote these reactions as allergic – whether justified or not is not an issue here – has prompted several colleagues to resort deliberately to a rating which is definitely wrong, regarding its morphological aspect (Table 1).

Indeed, two situations may not infrequently occur in practice in which the present reading scheme is not fully appropriate: (i) weak, only erythematous or follicular reactions (i.e. ?) which are, nevertheless, truly allergic, e.g. in the context of low test concentrations in threshold testing with an individually proven allergen – as exemplified by the image with a ? reaction to thiuram mix; and (ii) conversely, erythematous, infiltrated, possibly papular PT reactions morphologically qualifying as + reaction, which are, nevertheless, irritant, e.g. to the model irritant sodium lauryl sulphate.

In our set of images, the sharply demarcated, intensely erythematous, papulovesicular reaction to bufexamac, being a slightly atypical ++ reaction, was initially considered IR by 41 raters. Of these, 22 corrected their classification after knowing the allergen. While it is, again independently from this example, not incorrect to classify any morphological type of reaction as IR (Table 1), the aggregation of all grades of irritant reactions, ranging from soap effect to necrosis, unnecessarily hampers the evaluation of a given test preparation. From the clinical perspective on the PT as diagnostic tool for the detection of contact sensitization, irritant reactions are a mere unwanted side-effect. However, the scientific perspective of developing and improving PT preparations should not disregard differentiated information on the intensity of IR reactions.

To relieve the dilemma of the present reading scale in terms of a mixture of morphological criteria and interpretation we suggest separating both. In practice, this would mean the following steps – quite in accordance with the authoritative references quoted at the beginning.

1 Morphological classification during each PT reading, irrespective of the allergen. This would imply that a ? or + reaction is recorded as such, whether caused by a 'clear-cut' allergen such as colophony or a model irritant such as sodium lauryl sulphate;

2 At the final reading, an additional statement regarding interpretation as allergic, irritant or unknown aetiology is made, possibly further expanded (cross-reaction, active sensitization). This is based on all available information, including the allergen tested, concentration, vehicle, time course, previous PT results, previous exposure to a (cross-reacting) allergen, etc. Presently, some European centres already use this amendment (R. Jolanki, D. Orton, personal communication 2008), adding to the quality of both patient care and epidemiological data;

3 For those reactions considered allergic, clinical relevance is evaluated, preferably following well-founded suggestions.^{8,9}

The problems identified in our somewhat artificial study setting are relevant to clinical practice, among other problems not addressed here. One of these is the definition of a weak positive reaction. In a recent suggestion of a new scale for reading¹⁰ a + reaction is characterized as: 'homogeneous redness in the test area with scattered papules'. For a ++ reaction 'homogeneous redness and homogeneous infiltration' are required. Even though the definition of infiltration – palpable or visible? – is not without pitfalls, the suggestion may help to standardize patch testing further. However, if a revision of the PT scoring system is on the horizon, the separation of morphology and interpretation could also be addressed on this occasion. At any rate, regular training courses with blinded and unblinded photographs of PT reactions – or even better in vivo sessions on patients, where feasible – are necessary to improve the quality of PT readings, particularly before starting multicentre trials.

Acknowledgments

Three images were kindly provided by Thomas Fuchs, Göttingen;⁴ they are reprinted by permission from Dustri Verlag, Deisenhofen, Germany. Two images were kindly provided by Jürgen Kreusch, Lübeck, Germany.⁴ The study set-up was generously sponsored by Hermal/Trolab, Reinbek, Germany. The authors thank all quiz participants.

References

- 1 Wahlberg JE, Lindberg M. Patch testing. In: *Contact Dermatitis* (Frosch PJ, Menné T, Lepoittevin JP, eds), Berlin: Springer-Verlag, 2006; 365–90.
- 2 Fischer T. The art of patch testing. *Contact Dermatitis* 1990; **23**:221–3.
- 3 Uter W, Hegewald J, Kränke B *et al.* The impact of meteorological conditions on patch test results with 12 standard series allergens (fragrances, biocides, topical ingredients). *Br J Dermatol* 2008; **158**:734–9.
- 4 Uter W, Becker D, Schnuch A *et al.* The validity of rating patch test reactions based on digital images. *Contact Dermatitis* 2007; **57**:337–42.
- 5 Agresti A. *Categorical Data Analysis*. New York: John Wiley & Sons Inc., 1990.
- 6 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**:159–74.
- 7 Uter W, Lessmann H, Geier J *et al.* Is the irritant benzalkonium chloride a contact allergen? A contribution to the ongoing debate from a clinical perspective *Contact Dermatitis* 2008; **58**:359–63.
- 8 Bruze M. What is a relevant contact allergy? *Contact Dermatitis* 1990; **23**:224–5.
- 9 Ale SI, Maibach HI. Clinical relevance in allergic contact dermatitis. An algorithmic approach. *Derm Beruf Umwelt* 1995; **43**:119–21.
- 10 Menné T, White I. Standardization in contact dermatitis. *Contact Dermatitis* 2008; **58**:321.