

Does a short patch test training course have an impact?

Wolfgang Uter¹, Olaf Gefeller¹ and Vera Mahler²

¹Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-University Erlangen/Nürnberg, 91054 Erlangen, Germany and ²Allergy Unit, Department of Dermatology, University Hospital Erlangen, 91054 Erlangen, Germany

doi:10.1111/cod.12521

Key words: patch testing; standardization; training.

Patch test training schools are held for young professionals with the aim of teaching various aspects of patch testing, following respective guidelines [e.g. (1)]. In this context, live readings of patch test reactions may be performed. As a substitute for these, or also as a stand-alone measure, rating of digital images of patch test reactions is being used (2). In this setting, the trainee usually rates an image blind to the correct, consensus rating [and possibly also initially blind to the substance tested (3)], before the correct rating is displayed and discussed. A combined teaching approach was used in an EU-sponsored training school (see Acknowledgements), including live readings of patch tested patients and two sessions of rating digital photographs of patch test reactions; the results of the latter exercise are presented in this brief communication.

Methods

At the beginning of the 1-week training school, which comprised various aspects of patch testing, including several live readings later in the course, all participants took part in a first reading of 42 digital images of patch test reactions. These were provided by the online training tool of the German Contact Dermatitis Research Group (DKG) (<http://www.smartpracticeeurope.com/trainingtool/php/login.php>; last accessed 13 November 2015). All images had been previously viewed, discussed and agreed on regarding (i) their suitability, and (ii) the reaction

grade during the scientific meetings of the DKG. Images were projected with a standard digital projector with subdued ambient light. The categories used comprised +, ++ and +++ reactions, and doubtful (?+), irritant or negative reactions, in accordance with ICDRG (4) and European Society of Contact Dermatitis guidelines (1). If palpable infiltration was present, this was indicated on the slide. The individual reading results were collected by use of an electronic voting system (PowerVote SMART; PowerVote GmbH, Düsseldorf, Germany), and discussed immediately after the display of each image; correction of the recorded rating was not possible. Owing to this method of data acquisition emphasizing anonymization and data protection, it was not possible to link information from different image ratings of a given participant.

The audience consisted of 23 participants, including 9 consultant dermatologists (trainers), 4 dermatologists in training, 4 occupational physicians (one in training), 2 pharmacists, and 4 medical students. Experience in patch testing was < 3 months in 10, > 5 years in 7, and in between in the remaining 6. On the final day, a second round of readings was offered, with 50 images, and the aggregated results (frequencies of choosing the categories) were again collected as above for later analysis. The distribution of different grades of reactions in the two sessions is shown in Table 1.

To examine the hypothesis regarding whether there is a short-term effect of this patch test training programme on the (i) correctness (as regards agreement with the gold standard rating) and (ii) homogeneity of ratings, a number of suitable measures were derived from the data for each image:

- The percentage of perfect agreement with the gold standard.
- The percentage of agreement of the dichotomized rating ('allergic', comprising +, ++ and +++ reactions, versus 'non-allergic', comprising the remaining three outcome categories) with the gold standard identically dichotomized.

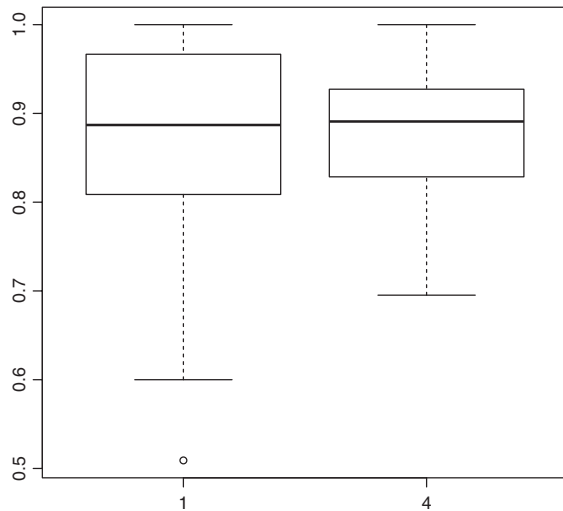
Correspondence: Wolfgang Uter, Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Waldstr. 4-6, D-91054 Erlangen, Germany. Tel: +49 9131 8522750; Fax: +49 9131 8522721. E-mail: wolfgang.uter@imbe.med.uni-erlangen.de

Conflicts of interest: W.U. has accepted honoraria for presentations or travel reimbursement from cosmetic industry associations for presentations given to these, and has received a lecture fee from Almirall Hermal. V.M. has received lecturing fees from SmartPractice, Almirall Hermal, GlaxoSmithKline, and Basilea. O.G. has no conflicts of interest to disclose. The online reading tool of the German Contact Dermatitis Research Society referred to is supported by SmartPractice, Germany.

Table 1. Distribution (absolute numbers) of reaction grades presented as digital images with a beamer on the first day (D1) and the final day (D4) of a patch test training school

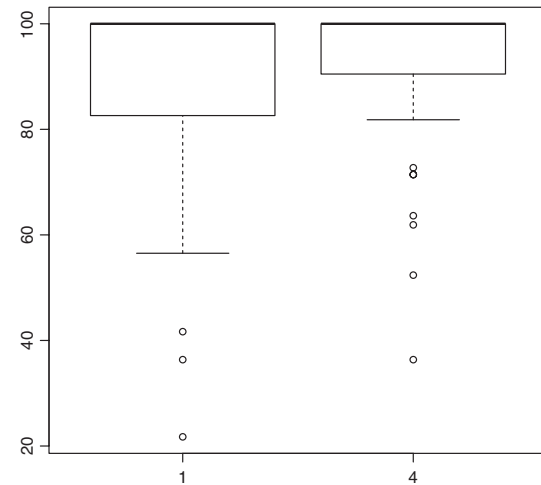
Day	+	++	+++	IR	? +	Negative	<i>p</i> -value
D1	14	10	14	0	3	1	0.004
D4	15	13	4	4	13	1	
	Positive			Non-positive			0.003
D1	38			4			
D4	32			18			

IR, irritant reaction.

p-value: results of an exact Fisher test.**Fig. 1.** Distribution of the Gini coefficient of ratings of digital images of patch test reactions on the first day (D1) and on the last day (D4) of a patch test training school.

- The Gini coefficient (5) of the distribution of outcome ratings over the six categories as a measure of concentration. The Gini coefficient ranges from 0 (entirely homogeneous distribution, i.e. the uniform distribution) to 1 (complete concentration in one category). To achieve the maximum value of 1 independently of the number of categories, the normalized version of the Gini coefficient was used.
- The absolute difference in proportions of the dichotomized outcome 'allergic' and 'non-allergic' reaction as a measure of heterogeneity, ranging from 0 (same proportion classified as 'allergic' and as 'non-allergic') to 100 (complete agreement).

In addition, for the dichotomized outcomes, chance-corrected agreement in terms of Cohen's kappa coefficient, aggregated over all images per day, was calculated. Data analysis was performed with the R statistical software package [version 3.1.3; <https://www.r-project.org/>, (6)], in particular, with the package 'ineq' for calculating the Gini coefficient. Besides descriptive statistics

**Fig. 2.** Distribution of the absolute difference between the proportions of images rated as 'allergic' (+, ++ or +++) and 'non-allergic' (? +, irritant reaction, negative) based on digital images of patch test reactions on the first day (D1) and the last day (D4) of a patch test training school.

and bivariate statistical tests, multifactorial analyses were employed in order to adjust for the degree of difficulty, which differed between the two sessions (see Table 1); to this end, images with irritant, doubtful or negative reactions were considered to be 'difficult'. Multiple linear regression modelling was used with the Gini coefficient as outcome, and negative binomial regression analysis addressing the absolute difference between proportions of dichotomized ratings. Confidence intervals (CIs) for estimates were calculated with the profile likelihood method.

Results

Perfect agreement with the gold standard in terms of exactly using the proposed reaction category was 65.7% on average on the first day and 64.6% on the final day ($p = 0.85$ for difference). For dichotomized ratings, the respective average percentages of agreement with the gold standard were 93.8% on the first day and 96.2% on the final day ($p = 0.20$ for difference). Cohen's kappa,

aggregated over all images and ratings of 1 day, was initially 0.71, and was 0.92 on the second day.

The average Gini coefficient was 0.873 on the first day and 0.884 on the final day (Fig. 1; $p = 0.56$ for difference). An adjusted analysis gave a very similar result.

Concerning the dichotomized ratings, the absolute difference between proportions was, on average, 87.6% on the first day and 92.3% on the final day (Fig. 2; $p = 0.35$ for difference). In this case, however, the adjusted analysis showed a significant impact of training on homogeneity in terms of reducing heterogeneity at the second reading by a factor of 0.379 (95%CI: 0.151–0.932, $p = 0.023$). The effect of 'difficulty' was of a similar magnitude and also significant ($p = 0.032$).

Discussion

The results of our small survey may seem disappointing: there were no significant improvements in (i) exact agreement with the gold standard rating and (ii) homogeneity of ratings among participants of a patch test training school after a 4-day period of intensive training. However, a closer look shows that there was some improvement at least concerning an aggregated outcome 'allergic' versus 'non-allergic', that is, the rating of major importance for the patient. This improvement, which is also mirrored by a distinctly higher kappa value on the second day, could even be identified as statistically significant after adjusting for the difficulty of assessment sessions. As patch test reading itself constituted only a small, if important, part of the course schedule, overwhelming effects can probably not be expected, and continuous training, supervision and 'calibration' (7) is necessary.

Moreover, it is clear that viewing a digital image of a patch test reaction – especially with a digital projector, as opposed to a well-calibrated computer screen – is a compromise for the sake of feasibility as compared with

the reading of patch test reactions on patients, as also performed at the training school, albeit in a necessarily limited format. Thereby, the 'gold standard' rating by the DKG, which had been based on similarly projected digital images, may not have been correct in each and every case, and also relied on the added information on 'palpable infiltration' by the one dermatologist contributing the image. In the discussion of single images, in particular, two aspects evolved that were considered to be difficult: (i) the distinction between light reflections on papules and vesicles, that is, differentiation between + and ++ reactions; and (ii) the distinction between vesicles (++ reaction) and confluent vesicles (+++ reaction). However, this reflects not only the visual shortcomings of digital images, which may be hard to overcome, but also the continuum of the biological allergic response in terms of a positive patch test reaction where there are transitions from papules to papules and vesicles, and from single to confluent vesicles. In routine patch testing, however, the grading of + to +++ strength of an allergic reaction is not of great importance for the final interpretation of patch test results and relevance assessment. Here, the distinction of an allergic versus a non-allergic reaction, that is, a positive reaction and a non-positive reaction, is crucial.

In conclusion, a positive effect of patch test training can be clearly recognized, as also observed in a similar study (7).

Acknowledgements

The study is based on data collected at a training school on patch testing 20–24 July 2015 in Erlangen, supported by the Horizon 2020 COST (European Cooperation in Science and Technology) action 'StanDerm' (TD-1206). We thank all trainees and trainers for their contribution.

References

- Johansen JD, Aalto-Korte K, Agner T et al. European Society of Contact Dermatitis guideline for diagnostic patch testing – recommendations on best practice. *Contact Dermatitis* 2015; **73**: 195–221.
- Uter W, Becker D, Schnuch A et al. The validity of rating patch test reactions based on digital images. *Contact Dermatitis* 2007; **57**: 337–342.
- Uter W, Frosch P J, Becker D et al. The importance of context information in the diagnostic rating of digital images of patch test reactions. *Br J Dermatol* 2009; **161**: 554–559.
- Fregert S (Ed). *Manual of Contact Dermatitis*: Copenhagen, Munksgaard, 1974.
- Brown M C. Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data. *Soc Sci Med* 1994; **38**: 1243–1256.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*: Vienna, R Foundation, 2014. ISBN 3-900051-07-0. Available at: <http://www.R-project.org>.
- Svedman C, Isaksson M, Björk J et al. 'Calibration' of our patch test reading technique is necessary. *Contact Dermatitis* 2012; **66**: 180–187.