

## ORIGINAL ARTICLE

# Development of a clinical score system for the diagnosis of photoallergic contact dermatitis using a consensus process: item selection and reliability

S. Cazzaniga,<sup>1</sup> S. Lecchi,<sup>1</sup> M. Bruze,<sup>2</sup> O. Chosidow,<sup>3,4,5,6</sup> T. Diepgen,<sup>7</sup> M. Gonçalves,<sup>8</sup> J. Hercogova,<sup>9</sup> P.D. Pigatto,<sup>10</sup> L. Naldi<sup>1,11,\*</sup>

<sup>1</sup>Centro Studi GISED - FROM, Bergamo, Italy

<sup>2</sup>Department of Occupational and Environmental Dermatology, Skåne University Hospital, Lund University, Malmö, Sweden

<sup>3</sup>Department of Dermatology, Hôpital Henri-Mondor, Créteil, France

<sup>4</sup>UPEC Université, Paris Est-Créteil Val-de-Marne, France

<sup>5</sup>French satellite of the Cochrane Skin Group, Créteil, France

<sup>6</sup>INSERM, Centre d'Investigation Clinique 006, Créteil, France

<sup>7</sup>Department of Social Medicine, Occupational and Environmental Dermatology, University of Heidelberg, Heidelberg, Germany

<sup>8</sup>Department of Dermatology, University Hospital and Faculty of Medicine, University of Coimbra, Coimbra, Portugal

<sup>9</sup>Department of Dermatology, 2nd Medical School, Charles University, Prague, Czech Republic

<sup>10</sup>Department of Bioscience for Health, Dermatological Clinic, IRCCS Galeazzi Hospital, University of Milan, Milan, Italy

<sup>11</sup>Department of Dermatology, Azienda Ospedaliera papa Giovanni XXIII, Bergamo, Italy

\*Correspondence: L. Naldi. E-mail: luigi.naldi@gised.it

## Abstract

**Background** Photoallergic contact dermatitis (PACD) is an uncommon condition, and there is a lack of validated criteria for its diagnosis.

**Objective** To identify a set of relevant criteria to be considered when suspecting a diagnosis of PACD and to assess the reproducibility of these criteria.

**Methods** This was a diagnostic item selection and reliability study performed between July 2012 and October 2012. A panel of seven recognized experts was invited to consecutive rounds of a Delphi survey and to a conclusive face-to-face meeting with the aim of obtaining an agreement on criteria for the diagnosis of PACD. The panel was also provided with a series of 16 reports of suspected PACDs to be classified according to a five-point likelihood scale. Identified criteria with the weights attributed by experts were used to develop a score system for the diagnosis of PACD. Consensus was measured by calculating the Intraclass Correlation Coefficient (ICC). The performance of the score system was evaluated in terms of overall classification accuracy.

**Results** Seven criteria were identified by experts as relevant for the diagnosis of PACD. The criteria were related to the type of skin lesions, accompanying symptoms, skin area involved, general medical history, modality of exposure to the culprit substance, history of exposure to the sun or other light sources and photopatch test results. Experts reached a moderate agreement on PACD cases classification, with ICC = 0.69 (95% Confidence Interval, CI, 0.50–0.86). The score system enabled discrimination of probable and definite PACD cases from possible and unlikely or excluded ones, with a nearly perfect agreement being observed between the score system classification and judgment by experts.

**Conclusion** A diagnostic score was proposed. The score should receive a comprehensive validation on a larger series of cases and with multiple evaluators.

Received: 4 August 2014; Accepted: 23 October 2014

## Conflicts of interest

None declared.

## Funding sources

This study was supported by a research grant obtained from the Ketoprofen Consortium, which consists of Menarini Industrie Farmaceutiche Riunite S.r.l., Bayer S.p.A., Cyathus Exquirere Pharmaforschungs GmbH, Dompe' S.p.A., EG S.p.A., JSC Grindeks, Hisamitsu UK Ltd., Istituto Biochimico Italiano G. Lorenzini S.p.A., Italfarmaco S.p.A., Pierre Fabre Ibérica S.A., Sandoz International GmbH, Sanofi-Aventis Groupe.

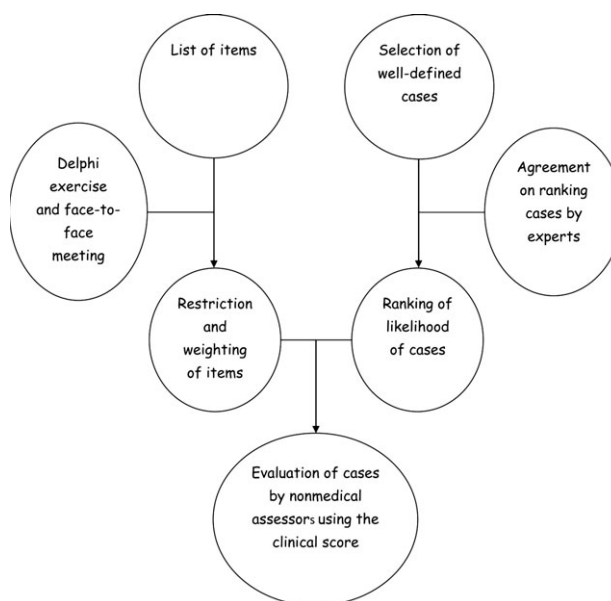
## Introduction

Photoallergic contact dermatitis (PACD) is an allergic reaction to a photoactivated antigen applied to the skin, which occurs in subjects previously sensitized to the same chemical, or to an immunologically similar chemical (cross-sensitization).<sup>1,2</sup> PACD has particular relevance in the area of drug and cosmetic safety, as well as in occupational medicine.<sup>3–6</sup> PACD is considered as an uncommon or rare condition, and limited data are available concerning incidence. In Italy, the observed rate of any reported photosensitivity reaction to drugs (including photoallergy and phototoxicity from topical and systemic agents) was 1.86 per million inhabitants per year (corresponding to 5.5 per 100 000 adverse reactions reported).<sup>7</sup> Similarly, in Norway, a retrospective analysis of adverse drug reaction reports for the years 1970–1995 revealed that approximately 5 per 100 000 reactions were classified as photosensitivity.<sup>8</sup>

The occurrence of PACD might have been underestimated to date. A specific evaluation of a potential PACD is often not fully undertaken, and many patients are diagnosed as suffering from contact dermatitis or drug rash without any further investigation or any direct specification of a potential photosensitizing substance.<sup>9–16</sup> This situation may also partly reflect the lack of standardized criteria for the clinical diagnosis of PACD. PACD is not always easily identifiable: the reaction may appear from a few days upon exposure to the culprit substance in previously sensitized patients, to several weeks when the sensitization process still needs to occur, and on other variables, such as the frequency and intensity of sun exposure in relationship with exposure to the chemical and the level of retention of the molecule in the epidermis.<sup>1,17</sup> In addition, the complexity of the diagnostic process is amplified by a wide variety of potential causal agents thus far identified. Thus, it is important to define common rules for suspecting PACD and to confirm it. The need of standardized diagnostic criteria is particularly high in the area of drug safety when assessing reports of suspected adverse drug reactions to perform the necessary measures to minimize risks, and, in general, when determining the incidence of PACD at the population level. Using a consensus process, we selected a set of relevant criteria for consideration in the diagnosis of PACD and developed a diagnostic score to improve the accuracy in the assessment of the suspected PACD cases, harmonizing diagnostic procedures and management decisions.

## Materials and methods

This was a diagnostic item selection and reliability study performed between July 2012 and October 2012. The general scheme of the study is presented in Fig. 1. The study was approved by the ethics committee of the main participating centres, namely Bergamo (Italy), Créteil (France) and Prague (Czech Republic).



**Figure 1** Scheme of the study. The figure shows the diagram of the study workflow.

### Delphi survey and selection of items for the diagnostic system

To generate a consensus for the diagnosis of PACD, seven selected recognized experts in drug allergy, photoallergy and/or contact dermatitis were invited to form a panel undergoing two consecutive rounds of a Delphi survey. The Delphi method is a procedure-tool used to obtain consensus among a group of people.<sup>18,19</sup> It relies on a systematic iterative approach for maximizing areas of agreement and exploring points of disagreement on questions that may be controversial or overlooked. The participants were asked to express their opinion on a web-based list of questions. The anonymity of the system enables the participants to express their personal view without face-to-face pressure with other peers. For this study, a list of 18 questions was prepared that focused on key issues of potential disagreement among the experts concerning diagnostic criteria for PACD. After the first round, the results were summarized and sent to all of the participants. The responses were used to develop a further questionnaire for the second Delphi round. After two Delphi rounds, a face-to-face meeting was held to allow a concerted reviewing of the entire process and to weigh the individual items.

### Expert classification of a series of reported cases

In addition to the participation in the Delphi exercise, the experts were sent a series of 16 potential cases of PACD. They represented the cases reported as suspected reactions to drugs in the Council for International Organisations of Medical Sciences

(CIOMS) Suspect Adverse Reactions Reporting forms. The reports included details on drug exposure, symptoms, medical history, risk factors and the results of photopatch tests when available. Experts were asked to independently classify these reports based on the following likelihood scale: excluded (i.e. definitely not a PACD), unlikely, possible, probable and definite case of PACD.

#### **Reproducibility of the diagnostic system among non-medical assessors**

The reproducibility of the judgment on the list of diagnostic criteria developed based on the Delphi exercise, was tested using an evaluation of the same set of 16 potential PACD cases, which were previously submitted to experts, by three independent and blinded non-medical assessors.

#### **Finalization of a diagnostic score**

The criteria identified in the Delphi rounds with the weights attributed by experts were used to develop a score system for the diagnosis of PACD. The total score was derived from the sum of weights of all of the characteristics shared by each case. Thresholds of the score system were defined, by relating scores with the average likelihood ranking of a series of 16 potential PACD cases made by experts, resulting in a classification of cases into a 5-point likelihood scale: excluded, unlikely, possible, probable and definite. If more than three main categories had missing data, then the score was considered as not assessable.

#### **Statistical analysis**

Two-way random single and average measure Intraclass Correlation Coefficient (ICC) [20] together with its 95% Confidence Interval (CI), was used to assess the absolute agreement among experts on the PACD assessment as well as their average rating. ICC can be interpreted as follows: 0–0.2 poor agreement, 0.3–0.4 fair agreement; 0.5–0.6 moderate agreement; 0.7–0.8 strong agreement and >0.8 almost perfect agreement.<sup>20</sup> Two-way random single measures ICC together with its 95% CI was also used to assess the absolute agreement among non-medical assessors.

The average (median) expert classification was used as a 'ground truth' evaluation to refine classification thresholds for the diagnostic score. Classification boundaries were first derived based on a uniform partition of the score ranges, and then refined by considering the points of maximum separation between adjacent categories obtained by averaging values between the maximum and minimum score of two consecutive categories.

## **Results**

#### **Items derived from the Delphi exercise**

The panel of experts reached a consensus on elements to be considered for diagnosis after two consecutive Delphi rounds and a

final consensus meeting. The diagnostic items selected were related to the type of skin lesions, accompanying symptoms, skin area involved, general medical history, modality of exposure to the culprit substance, history of exposure to the sun or other light sources and photopatch test results. The items were grouped into two groups: clinical history and diagnostic tests (Table 1).

#### **Expert classification of cases and reproducibility of assessment of items by non-medical assessors**

A moderate agreement among experts on case classification was found (ICC = 0.69, 0.50–0.86 95% CI). The classification for each case showed variations among experts of a maximum of two points on the likelihood scale. This enabled the use of their average rating as a 'ground truth' evaluation for classification purposes (average ICC = 0.91, 0.81–0.96 95% CI). The reproducibility of the judgment on cases based on the items selected during the Delphi exercise among the three independent non-medical evaluators was nearly perfect (ICC = 0.91, 0.81–0.97 95% CI). These results are shown in Fig. 2.

#### **Finalization of the diagnostic score**

The score system developed for the diagnosis of PACD based on the agreement exercise among experts is presented in Table 1.

On the basis of the case classification by experts, we derived the thresholds for the likelihood ranking of a PACD diagnosis as follows: excluded (–17 to –9), unlikely (–8 to +1), possible (+2 to +6), probable (+7 to +16) and definite (+17 to +23). The weighted criteria enabled the discrimination of probable and definite PACD cases from possible and unlikely or excluded cases, by adopting a threshold of ≤6 vs. >6 for the summary score, which yielded a nearly perfect agreement between the score system and judgment by experts (13 of 16 cases correctly classified using the score system). The accuracy of the overall score classification system was 11/16 cases (68.7%): 2/4 (50.0%) for unlikely, 2/4 (50.0%) for possible, 4/5 (80.0%) for probable and 3/3 (100.0%) for definite cases.

## **Discussion**

No validated criteria exist for the suspect diagnosis of PACD. The reaction can only be confirmed using photopatch testing, a procedure that most dermatologists, even those who patch test and provide phototherapy in their office, do not perform regularly.<sup>21</sup> The primary result of our study was the documentation that the development of diagnostic criteria for PACD is feasible. Validated diagnostic criteria may serve as a guide to make decisions regarding further investigations in patients with suspected PACD or may serve to confirm diagnoses of reported cases.

Diagnostic scores are procedures, in the form of formulas or scoring systems, used to guide and standardize healthcare decisions. The process of deriving a score is not straightforward and different approaches are possible. The first main approach is to

**Table 1** Clinical score for the diagnosis of PACD

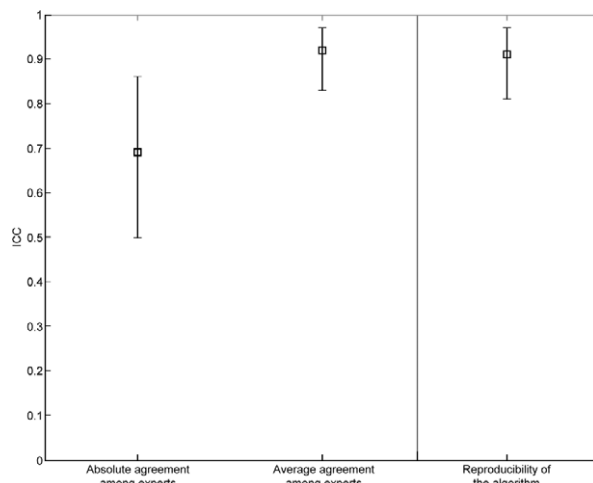
Criterion	Rule	Clinical importance (weighted score)*	Range
Clinical history			
Type of skin lesions	Vesicles and/or bullae on erythematous base	Suggestive (+2)	−2 to +4
	Erythema and oedema persisting for at least 48 h	Compatible (+1)	
	Papulae	Compatible (+1)	
	Necrosis	Unlikely (−2)	
Accompanying symptoms	Fever related to the dermatitis	Unlikely (−2)	−3 to 0
	Local symptoms other than itching	Unlikely (−1)	
Area of skin involvement	Primary involvement of shaded areas (areas not exposed to the sun/light sources	Very unlikely (−3)	−3 to +3
	Reaction limited to the areas exposed to the sun/light sources	Compatible (+1)	
	Reaction in the areas exposed to the sun/light sources with secondary spread to non-exposed areas	Suggestive (+3)	
Past medical history	History of diseases associated with photosensitivity (e.g. lupus erythematosus, polymorphic light eruption and xeroderma pigmentosum)	Very unlikely (−3)	−3 to +3
	History of previous photosensitivity reactions to the potential culprit substance	Suggestive (+3)	
Modality of exposure to the culprit photosensitizing substance	Last exposure to the culprit substance from more than 1 month to 2 months prior to the onset of the reaction	Compatible (+1)	0 to +2
	Last exposure to the culprit substance within 1 month prior to the onset of the reaction	Suggestive (+2)	
History of sun exposure	The last sun/light source exposure occurred within 48 h before the onset of the reaction	Suggestive (+2)	0 to +2
	The last sun/light source exposure occurred from more than 48 h to 2 weeks before the onset of the reaction	Compatible (+1)	
Diagnostic tests			
Photopatch test with the culprit substance	Doubtful (?) or irritant reaction (IR)	Neutral (0)	−6 to +9
	Positive reaction (+, ++, +++)	Highly suggestive (+9)	
	Negative reaction (−)	Very unlikely (−6)	

\*The final score is the sum of all weights related to applicable features. Based on this scoring system, PACD diagnosis is considered as: excluded (–17 to –9), unlikely (–8 to +1), possible (+2 to +6), probable (+7 to +16) and definite (+17 to +23). If more than three criteria have missing data, then the score is not assessable.

derive a set of rules on expert consensus using standardized procedures, such as the Delphi method and then to define a classification, in the form of decision trees or thresholds, based on a ‘ground truth’ evaluation of selected cases. This is particularly useful, as in our situation, when limited data are available or when it is difficult to gather experts to make independent evaluations. The second general approach is to initially have a ‘ground truth’ classification and then to derive a set of rules based on statistical methods (regression, classification trees, etc.). However, this method requires an adequate number of cases to have sufficient power in the analysis. Following its development, the score needs to be validated for reproducibility. This is an important point because any trouble in interpreting the score might result in an unpredictable behaviour in clinical practice. A further step indicates the assessment of the goodness of classification on a larger independent sample to further evaluate the accuracy of the score in practice. A final step is to introduce the score, using a monitoring system, in practice to determine whether it is really

productive and useful, or if it needs further adjustment or improvement. Within this long process, our study places in the second stage of development, i.e. assessment of reproducibility. We identified, using a consensus Delphi process, a set of relevant criteria for consideration when analysing suspected cases of PACD. The reproducibility of judgement on these criteria by independent assessors evaluating a series of carefully characterized historical cases was nearly perfect, and it was possible to associate weights to individual items.

Our scoring system had two components, in which one component focused on clinical history and the other component involved diagnostic procedures. When considering the type of lesions, the panel of experts agreed on giving the most relevance to the presence of vesicles and/or bullae on an erythematous base. The presence of erythema and oedema persisting for at least 48 h as well as the presence of papulae were judged as compatible with a diagnosis of PACD, although not necessarily indicative, whereas necrosis was not considered a feature of



**Figure 2** Measure of agreement among experts and reproducibility of the score. The figure shows ICC measurements together with their 95% CI

PACD. The presence of additional symptoms besides itching was regarded as suggestive of other diagnoses. Given the nature of the reaction, the initial lesions generally occurred on areas that come into contact with the culprit substance and are exposed to light. Sparing of the photoprotected area is characteristic of PACD. However, a secondary spread to non-light-exposed areas is notably common. Potential differential diagnoses include other diseases associated with photosensitivity. PACD is triggered by the photoactivation of a potential culprit substance upon exposure to the sun or other light sources. Most often, the reaction occurs within 1 month upon exposure to the photosensitizer; however, occasionally, the reaction can arise after a longer interval since the last local application of the drug. This finding is observed due to the potential accumulation and retention of the drug into the skin.<sup>17</sup> History of sun or other light source exposures and its relationship with skin lesions clearly suggests the possibility of a photoreaction, either mediated by an allergic or a phototoxic mechanism. A 48-h delay between light exposure and the onset of the reaction is generally suggestive of photoallergy, but our experts considered that in some cases, PACD could manifest even 2 weeks upon sun exposure. The method of choice for a complete diagnosis of PACD is the photopatch test, which is a fundamental piece of confirmatory evidence.<sup>15,16</sup> Unfortunately it is still underused in many countries.<sup>12</sup> Photopatch test results were given a high, but not exclusive importance by experts. Critical aspects in performing a photopatch test include the concentration of the substance, the vehicle used, the time and dose of irradiation, light source/UV spectrum, as well as the interpretation of the results. The patch is generally applied after the complete resolution of the reaction; however, some experts believe that it could be used even during

the active phase if an adequate region on the back is clear and unaffected. According to the consensus methodology, a duplicate set of patches should be applied for either 24 or 48 h before one of the sets is irradiated with 5 J/cm<sup>2</sup> of UVA, whereas the other is protected from UV light.<sup>12</sup> The irradiation dose (and UV spectrum) is also equally important as it should be sufficient for triggering a photoallergy response, but if excessive, it may trigger a false-positive phototoxic response. Any phototoxic substance will induce a positive photopatch test if sufficiently light irradiated. The reading should be recorded compared to irradiated vs. non-irradiated areas and preferably performed at consecutive intervals after irradiation (e.g. 2–4 days) to detect a suggestive crescendo pattern, typical of photoallergy.<sup>11</sup> Photopatch test results were given a strong importance in our score, however, in consideration of all the potential pitfalls and technical imperfections described thus far, it could not be considered as conclusive if taken alone in the diagnostic process.<sup>10,12</sup>

Our study had some limitations. It was based on a small number of potential PACD cases, and the predictive model should be further assessed and validated on a large series of cases and non-cases in different study settings. In addition the item selection was based on expert recommendations derived from the Delphi process. This is equivalent, in terms of grading of recommendations assessment (GRADE) system, to a level of evidence D.<sup>22</sup> However, we think that our study offers a starting point to get hints into the development of standardized diagnostic criteria for photosensitivity reactions.

The main result of our study was the documentation of the feasibility of the development of a diagnostic score for PACD to be further refined by considering subsequent steps in diagnostic criteria development.

## References

- 1 Drucker AM, Rosen CF. Drug-induced photosensitivity: culprit drugs, management and prevention. *Drug Saf* 2011; **34**: 821–837.
- 2 Glatz M, Hofbauer GF. Phototoxic and photoallergic cutaneous drug reactions. *Chem Immunol Allergy* 2012; **97**: 167–179.
- 3 Mark KA, Brancaccio RR, Soter NA, Cohen DE. Allergic contact and photoallergic contact dermatitis to plant and pesticide allergens. *Arch Dermatol* 1999; **135**: 67–70.
- 4 Shaw T, Simpson B, Wilson B, Oostman H, Rainey D, Storrs F. True photoallergy to sunscreens is rare despite popular belief. *Dermatitis* 2010; **21**: 185–198.
- 5 Victor FC, Cohen DE, Soter NA. A 20-year analysis of previous and emerging allergens that elicit photoallergic contact dermatitis. *J Am Acad Dermatol* 2010; **62**: 605–610.
- 6 Spiewak R. The frequency and causes of photoallergic contact dermatitis among dermatology outpatients. *Acta Dermatovenerol Croat* 2013; **21**: 230–235.
- 7 Noize P, Bénard-Larivière A, Aulois-Griot M, Moore N, Miremont-Salamé G, Haramburu F. Cutaneous adverse effects of ketoprofen for topical use: clinical patterns and costs. *Am J Clin Dermatol* 2010; **11**: 131–136.
- 8 Naldi L, Conforti A, Venegoni M et al. Cutaneous reactions to drugs. An analysis of spontaneous reports in four Italian regions. *Br J Clin Pharmacol* 1999; **48**: 839–846.

- 9 Selvaag E. Clinical drug photosensitivity. A retrospective analysis of reports to the Norwegian Adverse Drug Reactions Committee from the years 1970-1994. *Photodermatol Photoimmunol Photomed* 1997; **13**: 21-23.
- 10 DeLeo V, Gonzalez E, Kim J, Lim H. Phototesting and photopatch testing: when to do it and when not to do it. *Am J Contact Dermat* 2000; **11**: 57-61.
- 11 Neumann NJ, Hölzle E, Plewig G, Schwarz T, Panizzon RG, Breit R *et al*. Photopatch testing: the 12-year experience of the German, Austrian, and Swiss photopatch test group. *J Am Acad Dermatol* 2000; **42**: 183-192.
- 12 Bruynzeel DP, Ferguson J, Andersen K *et al*. European Taskforce for Photopatch Testing. Photopatch testing: a consensus methodology for Europe. *J Eur Acad Dermatol Venereol* 2004; **18**: 679-682.
- 13 Bryden AM, Moseley H, Ibbotson SH, Chowdhury MM, Beck MH, Bourke J *et al*. Photopatch testing of 1155 patients: results of the U.K. multicentre photopatch study group. *Br J Dermatol* 2006; **155**: 737-747.
- 14 Pigatto PD, Guzzi G, Schena D *et al*. Photopatch tests: an Italian multicentre study from 2004 to 2006. *Contact Derm* 2008; **59**: 103-108.
- 15 European Multicentre Photopatch Test Study. (EMCPPTS) Taskforce. A European multicentre photopatch test study. *Br J Dermatol* 2012; **166**: 1002-1009.
- 16 Gonçalo M, Ferguson J, Bonevalle A, Bruynzeel DP, Giménez-Arnau A, Goossens A *et al*. Photopatch testing: recommendations for a European photopatch test baseline series. *Contact Derm* 2013; **68**: 239-243.
- 17 Sugiura M, Hayakawa R, Kato Y, Sugiura K, Ueda H. 4 cases of photocontact dermatitis due to ketoprofen. *Contact Derm* 2000; **43**: 16-19.
- 18 Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; **311**: 376-380.
- 19 Mullen PM. Delphi: myths and reality. *J Health Organ Manag* 2003; **17**: 37-52.
- 20 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; **1**: 30-46.
- 21 DeLeo VA. Photocontact dermatitis. *Dermatol Ther* 2004; **17**: 279-288.
- 22 Terracciano L, Brozek J, Compalati E, Schünemann H. GRADE system: new paradigm. *Curr Opin Allergy Clin Immunol* 2010; **10**: 377-383.