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

S. Cazzaniga,1 S. Lecchi,1 M. Bruze,2 O. Chosidow,3,4,5,6 T. Diepgen,7 M. Gonçalo,8 J. Hercogova,9 P.D. Pigatto,10 L. Naldi1,11,*

1Centro Studi GISED - FROM, Bergamo, Italy
2Department of Occupational and Environmental Dermatology, Skåne University Hospital, Lund University, Malmö, Sweden
3Department of Dermatology, Hôpital Henri-Mondor, Créteil, France
4UPEC Université, Paris Est-Créteil Val-de-Marne, France
5French satellite of the Cochrane Skin Group, Créteil, France
6INSERM, Centre d’Investigation Clinique 006, Créteil, France
7Department of Social Medicine, Occupational and Environmental Dermatology, University of Heidelberg, Heidelberg, Germany
8Department of Dermatology, University Hospital and Faculty of Medicine, University of Coimbra, Coimbra, Portugal
9Department of Dermatology, 2nd Medical School, Charles University, Prague, Czech Republic
10Department of Bioscience for Health, Dermatological Clinic, IRCCS Galeazzi Hospital, University of Milan, Milan, Italy
11Department 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
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).\(^1\) PACD has particular relevance in the area of drug and cosmetic safety, as well as in occupational medicine.\(^2\) 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).\(^7\) 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.\(^8\)

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.\(^9\)–\(^16\) 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.\(^1,17\) 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).

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.\(^18\)–\(^19\) 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.
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. 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.

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. 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
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

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

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

*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.
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/cm² of UVA, whereas the other is protected from UV light. 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. 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.

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. 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


