Original Paper

International Archives of Allergy and Immunology

Int Arch Allergy Immunol 2014;163:59–68
DOI: 10.1159/000355333

Received: May 6, 2013
Accepted after revision: August 28, 2013
Published online: November 16, 2013

Objectifying the Conjunctival Provocation Test: Photography-Based Rating and Digital Analysis

S. Dogan[a] A. Astvatsatourov[a] T.M. Deserno[c] F. Bock[b] K. Shah-Hosseini[a]
A. Michels[a] R. Mösges[a]

[a]Institute of Medical Statistics, Informatics and Epidemiology and [b]Department of Ophthalmology, University Hospital Cologne, Cologne, and [c]Department of Medical Informatics, University Hospital, RWTH Aachen University, Aachen, Germany

Key Words
Conjunctival provocation test · Digital image analysis · Photography-based rating · Nasal provocation tests · Pollen allergens · Allergic rhinoconjunctivitis

Abstract
Background: Patients with allergic rhinoconjunctivitis are susceptible to both nasal and ocular symptoms. The conjunctival provocation test (CPT) is an established diagnostic procedure used in allergic rhinoconjunctivitis, particularly to document a patient's current reactivity to allergens. To date, there are no international guidelines defining the CPT. No approved evaluation method exists for interpreting CPT results. This paper aims to establish the digital analysis of macroimages as an objective, validated and standardized method for interpreting CPT results. ***Methods:*** In a clinical immunotherapy trial with 155 patients, treatment progress was documented based on the CPT. Local investigators used a symptom score to grade tearing, reddening and the patients' subjective perception of symptoms (mucosal irritation). A central observer rated conjunctival hyperemia via digital photography. Digital image analysis software was utilized to determine conjunctival hyperemia. ***Results:*** Spearman's correlation between the local investigators' and the central observer's ratings was $r = 0.729$ ($p < 0.001$); the percentage of total agreement was 48% (based on 739 photos). Digital image analysis (based on 48 photos) had a high percentage of total agreement with the central observer's ratings (69%) but a low percentage of total agreement with the investigators' ratings (38%). The corresponding correlations were $r = 0.264$ and 0.064, respectively. ***Conclusion:*** Photography-based rating by a central observer may represent a valuable supplement to the local investigator's assessment for making an objective evaluation of CPT results. Digital image analysis possesses the potential of being an objective evaluation method compared to the wide-spread subjective evaluation by the investigators.

Introduction

Allergic rhinoconjunctivitis is a widespread disease of great epidemiologic importance: about 500 million people worldwide are affected by allergic rhinitis [1]; its prevalence is over 20% in most countries [2], and the direct, indirect and intangible health costs associated with it continue to rise [3, 4]. The primary diagnosis of allergic rhinoconjunctivitis is usually made via skin prick testing or blood testing for specific IgE, methods which are approved and easy to apply in daily routine [5, 6]. Mucosal

KARGER

E-Mail karger@karger.com
www.karger.com/iaa

1018–2438/14/1631–0059$39.50/0

Correspondence to: Prof. Ralph Mösges
Institute of Medical Statistics, Informatics and Epidemiology
University Hospital Cologne, Lindenburger Allee 42
DE–50931 Cologne (Germany)
E-Mail ralph.moesges@uni-koeln.de

provocation is an established diagnostic procedure especially for documenting the strength of the allergic sensitization directly at the organ level under specific immunotherapy and the patient's current reactivity to allergens, respectively, whereas the skin prick test and blood testing for specific IgE are not suitable for this indication [5, 7]. Therefore, mucosal provocation tests are used in immunotherapy studies for documenting treatment effects and are accredited by regulatory agencies as a measure of outcome [5, 7–9]. Furthermore, direct mucosal allergen challenges, such as the conjunctival or nasal provocation test, are used as secondary diagnostic tools for a variety of purposes, for example: to confirm the diagnosis of occupational allergy; to confirm the diagnosis of seasonal allergic rhinitis and to exclude a clinical silent sensitization, if the result of a skin prick test or specific IgE measurement is contrary to the patient's clinical symptoms; assessing the symptoms directly at the organ level to evaluate the indication or efficacy of any antiallergic treatment, and to research the immunopathologic mechanisms in the target organs [7, 10, 11]. Nevertheless, different opinions exist in the literature as to whether the conjunctival provocation test (CPT) serves generally as a model of allergic rhinoconjunctivitis [7, 12–14] or not [14, 15]. The CPT has not yet been consistently defined by any international guidelines and is mentioned less often in the current literature [6, 10]. One of the reasons for this is that only little attention has been given to allergic conjunctivitis compared to other atopic disorders such as allergic rhinitis or asthma [16], although ocular symptoms occur in up to 70% of allergic rhinitis patients [17]. Different methods for evaluating CPT results are available: grading scales that take the local investigators' assessment and the patient's subjective perception of symptoms into account [7], photographic documentation of patients' eyes and rating of such images by an observer [18, 19], or rating by means of digital image analysis software [20]. However, none of these methods is a generally accepted tool for interpreting CPT results [10]. On the contrary, the nasal provocation test is mentioned in the literature as the approved mucosal provocation tool [21] and is consistently defined in detail by international guidelines [6, 11, 22, 23]. However, methods used to evaluate nasal provocation test results, e.g. anterior rhinomanometry, lack in sufficient reproducibility [24–26] and depend greatly on the examiner's experience and on the patient's compliance [7, 22].

The current literature describes several achievements in digital image analysis of conjunctival vessels, which are reported to be more sensitive than subjective ratings [19, 20, 27–36]. Combinations of different techniques are described, such as threshold-setting, edge-detection, color extraction, smoothing, fractal analysis or densitometry [28, 33].

Horak et al. [20] proved that digital imaging is a sensitive tool for measuring redness in conjunctival allergic reactions and claimed that this instrument may be able to replace subjective evaluations of CPT results. They demonstrated significant differences between the sensitivity of digital analysis and subjective assessment with respect to the differentiation of active treatment and placebo groups in the test results. Test reactions were documented using a special slit-lamp construction, and conjunctival redness was calculated by measuring optical density of the red fraction of the whole photo (technique of densitometry). Their publication is the only existing work validating digital image analysis for the use in CPTs.

Fukushima and Tomita [27] described a digital image analysis method capable of quantifying histamine-induced conjunctival hyperemia in guinea pigs. It was their objective to overcome the insufficient subjective grading with scoring systems and the need for automatic digital image analysis. The authors used ImageJ software to select vessel pixels in a predefined region of interest (ROI) after image processing with threshold setting and binarization. In a consecutive work, Yoneda et al. [28] investigated the reproducibility and reliability of automated software in analyzing conjunctival hyperemia in humans. They figured that current subjective evaluation methods using various grading scales needed to be replaced by more objective procedures. Their work also included the calculation of the percentage of red pixels in a predefined ROI and threshold setting. The authors asserted that it is possible to perform a simple and prompt analysis of photos taken with a slit-lamp using digital image analysis software. As a result the applied method was reproducible, subjects suffering from allergic conjunctivitis and subjects treated with Bimatoprost showed a significant higher percentage of detected pixels than the healthy control group. However, they described limitations of their analysis because the predefined ROI was very small and dependent on photographic conditions. Although no CPTs were conducted here, the patient collective included a few subjects with allergic conjunctivitis. The methods described [28] seem to possess the applicability in hyperemia measurement of the CPT.

Owen et al. [29] compared the measurement of conjunctival vessel width by an automated computer algorithm with the measurement by manual methods on digital photographs. The automatic algorithm was based on

the technique of smoothing that uses the Gaussian distribution of intensities in an image to detect certain structures, such as edges of vessels. The manual method is described as a hand-operated digital measuring of single conjunctival vessels and defined as the gold standard. The automatic method showed high intrasession repeatability and high intermethod repeatability with the manual method, although small over- and underestimations of the vessel width arose. The authors recommend the use of automated computer algorithms to accelerate and simplify digital measurements, in preference to the very time consuming and error-prone manual methods.

Fieguth and Simpson [30] compared the measurement of bulbar hyperemia by digital image analysis with ratings of bulbar hyperemia by clinicians using a grading scale with the aid of an internet-based survey. The digital image analysis consisted of the detection of vessel edges (Canny-edge algorithm) and the measuring of relative redness in each pixel (RGB color space) and was based on thirty sample photos of bulbar hyperemia. Results of the digital image analysis and the clinicians' judgments had a linear relationship to each other, but the variability between each measurement was definitely smaller than between the raters. In the end they came to the conclusion that this kind of analysis is able to replace subjective grading of bulbar redness.

Wolffsohn [33] applied a digital image analysis with color extraction and edge detection on sample photos of different subjective grading scales for conjunctival hyperemia. His work showed that the digital image analysis reflects exactly the values of the grading scales and presented highly reproducible results, in contrast to the results of the clinicians which show a high variability when using these scales. A clinical implementation was not carried out. Prior to this, he had also tested a digital analysis with threshold setting, color extraction and edge detection on a single subjective grading scale [34]. Color extraction and edge detection turned out to be the most stable and sensitive methods in detecting bulbar hyperemia.

Peterson and Wolffsohn [19] analyzed sensitivity and reliability of the digital image analysis with color extraction and edge detection in comparison with subjective rating using a grading scale. Pharmaceutical vasodilatation was induced in patients and afterwards captured with a digital camera and a slit-lamp biomicroscope. Sensitivity and reliability of the digital analysis were significantly higher than of the subjective rating. The authors concluded that this kind of objective analysis might serve as a new gold standard in examining the anterior eye. In a consecutive work [31] they converted the measurements of the validated digital image analysis into clinical grades to attain higher comparability with subjective grading scales. They showed that a conversion into objective grades allows for a more effective interpretation and more objective description of findings for clinicians. Beyond that, Peterson and Wolffsohn [32] did some research on the optimal resolution and the maximum compression of digitally analyzed photographs. They summarized that a medium resolution of 767 × 569 pixels and a strong compression of 50% of the original image do not cause a loss of quality of the evaluation.

Schulze et al. [35] were the first to apply fractal analysis to assess the degree of vascular branching in conjunctival hyperemia. They compared fractal analysis with measurements of chromaticity and vessel detection by implementation on sample photos of common grading scales for bulbar hyperemia, like Wolffsohn did before [33]. With all the techniques of digital image analysis that they applied [35], changings of redness in every scale could be detected. However, insufficient image quality limited the results of the fractal analysis.

Papas [36] researched eligible objective measured parameters of the conjunctival vascularization that reflect the ratings of clinicians as accurately as possible. Images from patients were captured with a digital camera and a slit-lamp biomicroscope. Thereafter, digital analysis using colorimetric and morphometric techniques and subjective rating using a grading scale were conducted. Morphometric measurements of the percentage of vessel-depicting pixels and of the number of vessels corresponded highly with the subjective ratings, whereas colorimetric methods had a lower degree of accordance. Therefore, the author concludes that subjective rating is focused primarily on the density of vessels rather than on the color.

As described above, ophthalmology is one of the major disciplines in medicine that uses digital photography and digital image analysis. Apart from the analysis of conjunctival hyperemia there are a lot of applications in ophthalmology and other specializations in medicine using such techniques. In ophthalmology, digital photography and analysis are used for diagnosis of the anterior eye segment [37], for vessel detection in corneal transplants [38], for analyzing the corneal involvement of pterygium conjunctivae [39, 40], for noninvasive calculation of hemoglobin in the wide field of emergency medicine [41], for examining rotational stability of intraocular lenses [42], in diagnosis of dry eye [43] and in the observation of side effects of antiglaucomatous treatment [44]. In dermatology they are implemented for measuring erythema and edema of the skin prick test [45–47] or intradermal skin

tests [48], for analyzing repigmentation during treatment of vitiligo [49], for quantification of skin lesions [50], and in diagnosis and treatment of atopic dermatitis [51, 52]. Otorhinolaryngology makes use of this technique during endoscopic transsphenoidal pituitary surgery [53] and for video-laryngoscopic-assisted surgery [54], while it is used for analysis of body posture in patients with asthma [55, 56] in pediatrics. In general it is a suitable method for documentation, communication and education on, for example, orthopedic problems and treatment [57] or photographic documentation in facial plastic surgery [58].

The work described in this publication aims to compare different methods for evaluating the results of conjunctival allergen challenge. Its purpose is also to establish in future the digital image analysis of macroimages as an objective, validated and standardized method for evaluating CPT results. This method, which should be fully automated, easy to apply and inexpensive, may thereby improve the scientific reputation of CPT in allergy research. Achieving the objectives would help CPT evolve to being an equal or even better alternative to the nasal provocation test for documenting a patient's current allergen reaction under immunotherapy treatment, or for being used as an outcome parameter in clinical hyposensitization trials. Since Blackley's [59] first report of applying pollen on his conjunctiva and documenting the reaction in 1873, CPT has been frequently used to diagnose allergic rhinoconjunctivitis, especially in scientific work. It is generally easy to perform, inexpensive, rarely associated with systemic reactions, and largely independent of the patient's compliance and examiner's experience [7, 20, 22, 60]. It is considered as an approved tool for analyzing pathomechanisms of allergic diseases and assessment of pharmaceutical effects on the allergic reaction [61].

Materials and Methods

Data Collection

Patient data were obtained from a prospective, double-blind, randomized, controlled, multicenter, dose-finding sublingual specific immunotherapy study with 155 subjects conducted before the 2012 grass pollen season. Clinical outcome was documented via CPT. Skin prick test was conducted to confirm the diagnosis of allergic rhinoconjunctivitis in each patient at the first visit prior to intake of any immunotherapy drugs. The preliminary results of this trial have been published elsewhere [62]. For our work, we used the patient data collected in the above-mentioned study, but our research only focuses on the methods applied in this study; the study treatment itself was not evaluated here.

Table 1. Gronemeyer scale for evaluating CPT test results [7]

Stage	Findings
0	No subjective or visible reaction
I	Itching, redness, foreign body sensation
II	Stage I plus tearing, vasodilation of the conjunctiva bulbi
III	Stage II plus vasodilation and erythema of the conjunctiva tarsi, blepharospasm
IV	Stage III plus chemosis, lid swelling

In the study, conjunctival allergen challenge was conducted three times with the standardized allergen extract ALK-lyophilized SQ (ALK-Abelló A/S, Hørsholm, Denmark): at visit 1 (prior to sublingual immunotherapy; SLIT), at visit 3 (after 4 weeks of treatment) and at visit 4 (after 12 weeks of treatment). Grass pollen allergen was applied to the patients' eyes in consecutive dosages of 100, 1,000 and 10,000 standard quality units (SQ-U)/ml at each visit. The testing procedure is derived from the CPT protocol by Riechelmann et al. [7] (whose trial did not include a further administration of an additional lower CPT dosage of 100 SQ-U/ml to the conjunctiva). The local investigator documented the patient's conjunctival reaction 10 min after each application. In the case of a clearly negative reaction, the investigator proceeded with the next higher dosage. If the reaction was definitely positive, allergen challenge was discontinued immediately. To rule out false-positive reactions, the allergen-free control solution ALK-diluent was applied as a blank value to the patients' eyes at each visit. By definition, the control solution was administered to the right eye and the allergen to the left eye (unilateral CPT).

Local Investigator Rating

In total, eleven local investigators were involved in the evaluation of the test results after each applied allergen dosage. Their rating comprised the assessment of the visible objective parameters of tearing and reddening as well as the documentation of the subjective symptoms of mucosal irritation (e.g. itching or foreign body sensation) reported by the patient. For systematic grading of symptoms, the investigators used the scale by Gronemeyer (table 1) [7].

The investigators rated the test reaction as positive when findings in the eye were at least stage II. Assessments at stage 0 and stage I were both considered negative and CPT was continued at the next higher dosage. A clear positive conjunctival reaction needed to consist of symptoms recognized both subjectively and objectively; the domination of subjective symptoms such as those in stage I was not sufficient.

Photographic Documentation

During the evaluation of CPT results, investigators made photographic documentation of each treated eye. For recording macroimages, all centers were equipped with a digital camera appropriate for macrophotography (Ricoh CX4, Ricoh Company Ltd., Tokyo, Japan). One study center was equipped with a digital single-lens reflex camera (Olympus E-3, Olympus Corporation, Shinjuku, Tokyo, Japan) combined with a macro lens (Olympus ZUIKO DIGITAL ED 50 mm MAKRO 1:2.0), a macro LED light

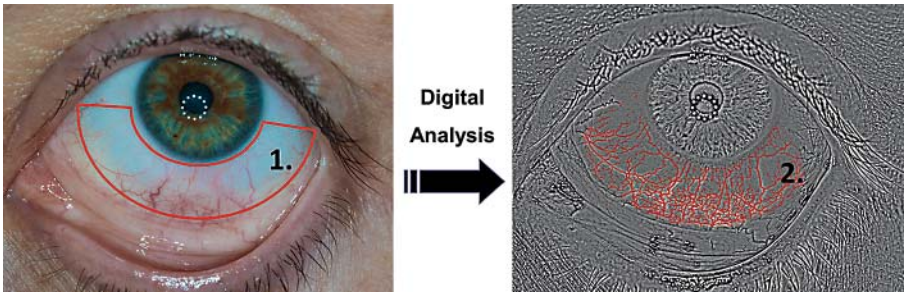

Fig. 1. Relevant steps in semiautomatic digital analysis using the software cell^F. 1. Manual segmentation of the ROI (area outlined in red). 2. Automatic vessel detection in gray value range.

and a remote-control release. The investigators were instructed to always use the following camera settings: autofocus, focal point positioning on the conjunctival area below iris, white balance of 4500 K, flashlight off, photosensitivity of ISO 400, a small focal aperture of f/10 and an automatic shutter speed based on the focal aperture. Furthermore, patients were requested to pull down their lower eyelid to uncover a maximum area of the bulbar conjunctiva below the iris and should focus on a colored marker on the camera's front side. A special stand ensured the steady position of patient and camera. It allows for positioning the camera in all directions; furthermore, it is light-weight, portable and inexpensive. The stand was developed and constructed by one of the authors (A. Astvatsatourov).

Central Observer Rating

On the basis of macroimages, a central observer rated the test reaction displayed on the computer screen. The observer only graded the visible objective parameter of conjunctival reddening and was blinded to treatment dosages, patient information, health conditions and local investigators' ratings. Table 2 shows the scale used for the systematic grading of redness. The status before CPT served as a reference point for the rating; during the rating process, the observer was able to make a side-by-side comparison of the patient's eyes before and after each allergen dosage.

The central observer rated the digital image of the test reaction as positive when findings were at least stage I. In this grading scale, the ROI is defined as a conjunctival area along the course of a single prominent visible vessel and its sprouts to be comparable and to mimic clinical subjective ratings [35]. The above classification is derived from the approved Efron scale for grading conjunctival hyperemia as one complication associated with contact lens wear [35].

Digital Analysis

Digital image analysis was performed thereafter to determine the visible objective parameter of conjunctival redness in the taken macrophotographs. It has been proven to be a sensitive tool for measuring conjunctival allergic reaction [20]. The digital image analysis program cell^F (Olympus Soft Imaging Solutions GmbH, Münster, Germany) is an appropriate instrument for detecting thin objects such as thread in a more or less uniform background. Bock et al. [38] used this software for detecting neoangiogenesis in corneal transplants. In our work cell^F was utilized to detect vessels in photographs of conjunctival allergic reactions.

This digital analysis procedure comprised the following steps (fig. 1): first, ROIs were segmented manually to direct the software to search for vessels and measure redness. To guarantee a consistent rating, all photos of a patient's visit were presented simultaneously on the computer screen to modify the ROI to fit to all of these images. Thus, the ROI may have differed from patient to patient and from visit to visit, but not between photos of a patient's visit and of a CPT session, respectively. All photos were taken so as to capture the whole visible bulbar conjunctival space, so that the segmented ROI could be very large and cover as much as possible of the conjunctiva, unlike other methods using small ROIs of constant size and form [28]. Second, all images were modified by a four-filter system, which was integrated in the software and used to enhance color contrast, depict vessel edges and rims, and reduce background noise. This filter sequence was adopted from the detailed description of Bock et al. [38] who were the first to implement cell^F software on vessel detection. Third, thresholds were set manually to differentiate bright vessels from the dark background, so that vessel detection was adjusted to preexisting redness and noise. Former analyses have shown that a threshold at the gray value of 110 in a value range between 0 and 255 (8-bit images) is, on the one hand, most accurate in detecting as many vessels as possible, and on the other hand able to prevent the detection of background noise (optimal signal-to-noise ratio). Fourth, it was possible to measure redness based on a gray value range in which blood vessels are best presented [38]. The software detected vessels within the predefined ROI in a single automatic step (fig. 1), unlike other software that analyzes each single vessel sprout and its vasodilatation manually [38, 63]. Finally, the software used in our investigation identified vessels inside the ROI and presented them as shown in figure 1. Redness was calculated as a percentage value of the vessel pixels recognized within the ROI. If the increase in redness after CPT was 10% greater than the initial percentage value, the test reaction was evaluated as clearly positive. In addition, the

Table 2. Rating scale for conjunctival redness in digital images

Stage	Findings
0	No visible reaction
I	Mild reddening in a single ROI
II	Stage I plus severe reddening in a single region of interest or in several ROIs
III	Stage II plus severe ubiquitous reddening, particularly in the conjunctival region close to the limbus corneae

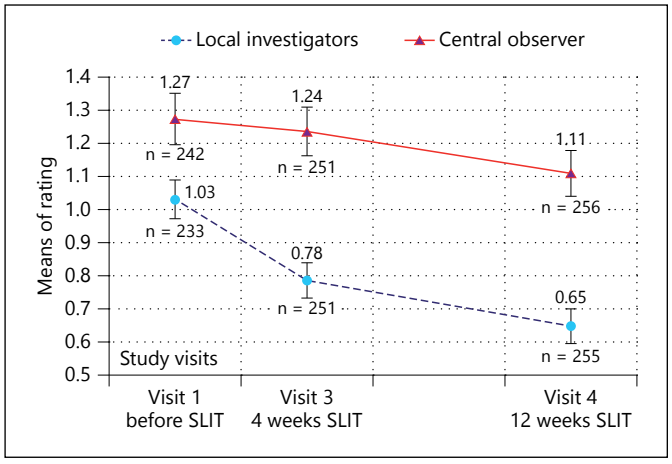

Fig. 2. Means of ratings of the local investigators (dashed line) and the central observer (solid line) across all study visits. Mean values and number of subjects are listed with each point. Error bars represent the SEM (standard error of the mean).

increase in redness was classified into different stages: stage 0 (increase <10%), stage I (increase ≥10%), stage II (increase ≥20%) and stage III (increase ≥30%), making this digital rating method easier to compare with the other rating methods. This digital procedure was applied in 1 of the 11 study centers for 16 patients.

Statistical Methods

To compare all methods with one another and to examine the equality of both the central observer's rating and the digital image analysis to the investigators' ratings, Spearman's rank correlation coefficient *r* and the percentage of total agreement were calculated. Spearman's rank correlation test was chosen because the rating results only have limited integer values from 0 to 4 (with regard to the above-mentioned grading scales). The percentage of total agreement was calculated in addition to Spearman's correlation as an alternative value of concordance. Simple deltas (Δ) between the results of all methods were calculated (Δ = CPT $stage_{Rating\ A}$ minus CPT $stage_{Rating\ B}$), i.e. local investigator rating *minus* central observer rating, local investigator rating *minus* digital image analysis, and central observer rating *minus* digital image analysis, respectively. A Δ of '0' defines a concordance of 100% between the compared ratings and, therefore, the percentage of this Δ was utilized to show a total agreement between the applied methods. In addition, the mean rating values of the utilized methods were calculated across all study visits to analyze the course of the ratings during the trial. Calculations were performed using the program SPSS (IBM Corp., New York, N.Y., United States).

Results

The correlation between the central observer's and the local investigators' was high with a Spearman's rank correlation coefficient of $r = 0.729$, corresponding to a statistical significance of $p < 0.001$ (739 photos from 155 subjects). The percentage of total agreement between both ratings was 48.4%. Moreover, analysis of the mean values of the ratings across all study visits showed that both the course of the central observer's rating and the course of the local investigators' ratings are characterized by decreasing rating stages throughout the study period (fig. 2; table 3). Decreasing rating stages indicated a lower CPT stage and therefore a reduced allergic reaction (see also table 1, 2). While the mean central observer's rating remained nearly constant for the first 4 weeks, the mean investigators' ratings showed a large decrease during the same period. However, starting at visit 3 (after 4 weeks), the central observer's rating decreased parallel to the investigators' ratings for the last 12 weeks of the study (visit 4). It must be noted that compared to the central observer's rating, the local investigators' ratings show lower absolute mean rating values across all study visits (fig. 2; table 3). So the judgments of the local investigators and the central observer tend to veer away from each other during the study period. In a similar manner, the Spearman's rank correlation coefficients decrease from $r = 0.763$ at visit 1 to $r = 0.725$ at visit 3 and $r = 0.723$ at visit 4 (table 4).

The digital image analysis (48 photos from 16 subjects) was characterized by a high percentage of total agreement of 68.75% with the central observer's rating, having similar over- and under-rankings, but it had a lower degree of total agreement of 37.5% with the local investigators' ratings. Similarly, Spearman's correlation between the digital image analysis and central observer was $r = 0.264$, while between the digital image analysis and the local investigators the Spearman's rank correlation coefficient was $r = 0.064$; both figures were not statistically significant (table 5).

Discussion

Three methods for evaluating CPT, all having different objective and subjective parameters, were compared to one another in this study: the first method involved symptom rating conducted by local investigators using a score comprising the subjective assessment of tearing and reddening, and subjective patients' information (mucosal irritation, e.g. itching or foreign body sensation); the second method was the subjective assessment of the single parameter conjunctival reddening on the basis of digital photos by an independent and trained central observer using a modified grading scale of conjunctival hyperemia,

Table 3. Mean ratings (bold) of the local investigators and the central observer across all study visits

	Investigators V1	Observer V1	Investigators V3	Observer V3	Investigators V4	Observer V4
Valid, n	233	242	251	251	255	256
Missing, n	40	31	22	22	18	17
Mean	**1.03**	**1.27**	**0.78**	**1.24**	**0.65**	**1.11**
Median	1.00	1.00	1.00	1.00	0.00	1.00
SEM	0.0588	0.0773	0.0539	0.0728	0.0525	0.0697
SD	0.897	1.202	0.854	1.154	0.838	1.115
Minimum	0	0	0	0	0	0
Maximum	3	3	3	3	3	3

Table 4. Spearman's correlation and percent agreement between the central observer and the local investigator

		Observer
Spearman's correlations: local investigators/central observer		
Total: investigators	Correlation coefficient	0.729
	Significance (two-tailed)	0.000
	n	739
Visit 1: investigators	Correlation coefficient	0.763
	Significance (two-tailed)	0.000
	n	233
Visit 3: investigators	Correlation coefficient	0.725
	Significance (two-tailed)	0.000
	n	251
Visit 4: investigators	Correlation coefficient	0.723
	Significance (two-tailed)	0.000
	n	255
	Frequency	%
Percent agreement: local investigators/central observer		
Investigators negative; observer positive	305	41.3
Agreement	358	48.4
Investigators positive; observer negative	76	10.3
n	739	100.0

Correlation is significant at the 0.01 level (two-tailed).

Table 5. Spearman's correlation and percent agreement between the digital image analysis software cell^F and the central observer or the local investigators

	Observer	Investigators
Spearman's correlations		
cell^F software – local investigators/central observer		
Correlation coefficient	0.264	0.064
Significance (two-tailed)	0.070	0.664
n	48	48
	Frequency	%
Percent agreement		
cell^F software – local investigators		
Investigators negative; software positive	27	56.3
Agreement	18	37.5
Investigators positive; software negative	3	6.3
n	48	100.0
cell^F software – central observer		
Observer negative; software positive	4	8.3
Agreement	33	68.75
Observer positive; software negative	11	22.9
n	48	100.0

Correlation is significant at the 0.01 level (two-tailed).

and the third method entailed rating conjunctival redness as shown in photos of the patients' eyes using the potentially most objective digital image analysis software Olympus cell^F.

The photography-based rating by a central observer is described in Results as having a high correlation to local investigators' ratings, and it can be used as a stable method to analyze CPT results, e.g. as an outcome parameter in clinical immunotherapy studies [64]. However, the percentage of total agreement is 48.4%. Photography-based observer rating and its suitability are not new concepts; other authors have already described the existing advantages. Kjaergaard et al. [18] pointed out in their research that a photographic rating of CPT results seems to be more sensitive than local rating with the naked eye.

More objectivity in evaluating CPT results is necessary [9], and a central rating is capable of improving this objectivity. Finally, the observer rating relies on the observer's subjective evaluation of the test reaction as seen in the photos, meaning that it is not the most objective method [35]. The observer, in our investigation, is only able to analyze the single parameter of redness; other important information is unavailable. Therefore, the local investigator rating should be the evaluation method of choice, supplemented by the assessment of a skilled examiner and the patient's subjective information [8]. We also conclude that a more objective photography-based rating should be carried out by a central observer, especially to improve the evidence of treatment effects in clinical trials [8, 18, 20]. Our findings support this conclusion, particularly the observations that the central observer's rating declined parallel to the investigators' ratings with a delay of 4 weeks and that the observer exhibited higher mean ratings across all study visits compared to the investigators. Central ratings may reveal subjective effects that appear mainly during the first few weeks of treatment. Such effects may lead to contrasting results, because the investigators' ratings take into account subjective feelings and symptoms (e.g. itching) of patients expecting a rapid onset of successful therapy.

A limitation to the comparability of the central observer rating with the investigator rating is the difference between the grading scales used. The investigator defines a clear positive reaction of the CPT starting at stage II, i.e. when subjective patient perception of symptoms (e.g. itching) occur simultaneously with symptoms visible to the investigator (e.g. redness or lacrimation; table 1). On the other hand, the central observer defines a clear positive reaction starting at stage I, i.e. even small increases of redness visible to the observer lead to positive test reactions (table 2). The blinded observer, however, possesses no information about what the patient perceives and is thus not able to make such fine differentiations as the local investigator can. To avoid this systematic error, the statistical calculation of concordance was classified as 100% agreement, negative deviation and positive deviation (table 4). A higher concordance between observer and investigators could possibly occur in statistics without that systematic error. This issue will be analyzed in subsequent scientific work.

Another limitation to objectivity is that the central observer is not blinded to the prior test reactions. Each photo is rated based on the photo of the lower CPT dosage. Otherwise, such a blinded observer rating is more suitable for evaluation of the test reaction in patients sensitized to multiple pollen allergens, although its objectivity is limited. Sometimes patients have preexisting conjunctival redness, so that rating photos after each CPT dosage based on the prior reaction or reaction before CPT is very beneficial here.

The Gronemeyer symptom scale [7] we used is not the only grading tool available. Nunez et al. [65] utilized the very detailed score by Abelson [66], in which the severity of each of the symptoms of reddening, chemosis, tearing and itching is assessed. A similar grading system of 0–4 for the severity of each symptom is described elsewhere [9]. In these systems, however, interindividual differences between investigators' ratings are not negligible and, without slit-lamp examination, are difficult to assess. We decided to use the Gronemeyer scale as a very current classification and as a good compromise for conducting a simple yet sufficient investigator evaluation. Furthermore, it conforms to the *Guideline on the Clinical Development of Medicinal Products for the Treatment of Allergic Rhinoconjunctivitis* by the European Medicines Agency [8].

The digital image analysis applied here has a greater percentage of total agreement with the central observer's ratings than with the local investigators' ratings. Similarly, the Spearman's correlation between the digital image analysis and the central observer's ratings is higher than between the digital image analysis and the local investigators' ratings, but the values are not statistically significant. Digital image analysis and the central observer rating were both based on the single objective parameter of conjunctival redness, whereas the local investigator rating relied on different, also subjective, parameters. This circumstance will be examined, however, in a consecutive ongoing study involving more patients.

The imaging software cell^F was applied for the first time in 16 patients to analyze conjunctival redness. The aim was to assess the digital image analysis software, originally applied to neoangiogenesis detection in corneal transplants of mice and using images taken under microscopic magnification with a professional SLR camera. Therefore, in this dose-finding study one center was equipped with a special stand and a professional SLR camera, so that the photographs could be taken under standardized and controlled conditions by one of the authors (S. Dogan). The photographs taken in other centers were used for documentation purposes. It was unclear whether this kind of analysis would have been suitable for detecting vessels in the photographs of the human conjunctiva because of anatomical differences between the human conjunctiva and mice cornea, and the different photographic magnification factors. Nevertheless, we decided to test this analysis software and felt confident that

adjustments to filter modification and threshold settings would remedy this issue. Moreover, other authors have proven that a method primarily employed for examining the cornea can also be used for the conjunctiva [35].

With cell^F software, a few very promising results arose (cf. fig. 1), although others were error prone. Sources of errors continue to be low image quality (especially light reflections and focus variations due to the wet and spherical background of the human conjunctiva), inadequate opening of patients' eyes or different lines of sight, and deviations in manual threshold settings and ROI segmentations.

Our research was not yet able to achieve its objective of developing a fully automated, objective, validated and standardized digital image method for analyzing CPT results. However, our research did show that photographic documentation is possible using affordable equipment (price of the self-constructed stand is EUR 250, prices of the camera device vary between EUR 200 for a standard digital camera and around EUR 2,000 for a suitable digital SLR camera including a suitable macro lens) and does not require large and expensive slit-lamp and microscope camera constructions, as demonstrated by the good correlation between the central observer and the local investigators' ratings. Digital analysis with cell^F as a semiautomatic method is easy to apply and works with the reasonably priced equipment. Such an inexpensive system could be attractive for everyday clinical use by allergists, who are often otolaryngologists, dermatologists, general practitioners or pulmonologists [4, 21] rather than ophthalmologists for whom slit-lamps are standard examination equipment.

We are in the process of developing a fully automated digital image analysis to replace manual segmentation by a single automated step. A first attempt has yielded very promising results with segmentation based on Hough transform for circles. In addition, color model transforms have been used for further objectivity and robustness [67]. There are lots of expectations that a higher validity can be attained using digital CPT analysis in the future, since digital analysis possesses the greatest objectivity [19, 20], for contributing to the revival of the CPT in allergy research.

References

1 Lightman S, Scadding GK: Should intranasal corticosteroids be used for the treatment of ocular symptoms of allergic rhinoconjunctivitis? A review of their efficacy and safety profile. Int Arch Allergy Immunol 2012;158:317–325.

2 Bousquet J, Bodez T, Gehano P, Klossek JM, Liard F, Neukirch F, Le Gal M, Janin N, Allaf B: Implementation of guidelines for allergic rhinitis in specialist practices: a randomized pragmatic controlled trial. Int Arch Allergy Immunol 2009;150:75–82.

3 Han DH, Rhee CS: Sublingual immunotherapy in allergic rhinitis. Asia Pac Allergy 2011;1:123–129.

4 Shah A: Current trends in upper airways and ocular allergic inflammation. Asia Pac Allergy 2011;1:105–107.

5 van Hage-Hamsten M, Pauli G: Provocation testing with recombinant allergens. Methods 2004;32:281–291.

6 Bousquet J, Khaltaev N, Cruz AA, et al: Allergic rhinitis and its impact on asthma (ARIA) 2008. Allergy 2008;63:8–160.

7 Riechelmann H, Epple B, Gropper G: Comparison of conjunctival and nasal provocation test in allergic rhinitis to house dust mite. Int Arch Allergy Immunol 2003;130:51–59.

8 European Medicines Agency: Guideline on the clinical development of medicinal products for the treatment of allergic rhinoconjunctivitis. London, EMA, 2004.

9 Friedlaender MH: Conjunctival provocative tests: a model of human ocular allergy. Trans Am Ophthalmol Soc 1989;87:577–597.

10 Melillo G, Bonini S, Cocco G, Davies RJ, de Monchy JG, Frolund L, Pelikan Z: EAACI provocation tests with allergens: report prepared by the European Academy of Allergology and Clinical Immunology Subcommittee on provocation tests with allergens. Allergy 1997;52:1–35.

11 Bachert C, Borchard U, Wedi B, Klimek L, Rasp G, Riechelmann H, Schultze-Werninghaus G, Wahn U, Ring J: Allergic rhinoconjunctivitis. Guidelines of the dgai in association with the ddg (in German). J Dtsch Dermatol Ges 2006;4:264–275.

12 Leonardi A, Battista MC, Gismondi M, Fregona IA, Secchi AG: Antigen sensitivity evaluated by tear-specific and serum-specific IgE, skin tests, and conjunctival and nasal provocation tests in patients with ocular allergic disease. Eye 1993;7:461–464.

13 Leonardi A, Fregona IA, Gismondi M, Daniotti E, Carniel G, Secchi AG: Correlation between conjunctival provocation test (CPT) and systemic allergometric tests in allergic conjunctivitis. Eye 1990;4:760–764.

14 Naclerio RM, Philpot EE: Is conjunctival allergen challenge a model of seasonal rhinoconjunctivitis? Allergy Asthma Proc 2009;30:212–214.

15 Radcliffe MJ, Lewith GT, Prescott P, Church MK, Holgate ST: Do skin prick and conjunctival provocation tests predict symptom severity in seasonal allergic rhinoconjunctivitis? Clin Exp Allergy 2006;36:1488–1493.

16 Katelaris CH: Ocular allergy in the Asia Pacific region. Asia Pac Allergy 2011;1:108–114.

17 Bousquet PJ, Demoly P, Devillier P, Mesbah K, Bousquet J: Impact of allergic rhinitis symptoms on quality of life in primary care. Int Arch Allergy Immunol 2013;160:393–400.

18 Kjaergaard SK, Pedersen OF, Taudorf E, Molhave L: Assessment of changes in eye redness by a photographic method and the relation to sensory eye irritation. Int Arch Occup Environ Health 1990;62:133–137.

19 Peterson RC, Wolffsohn JS: Sensitivity and reliability of objective image analysis compared to subjective grading of bulbar hyperaemia. Br J Ophthalmol 2007;91:1464–1466.

20 Horak F, Berger U, Menapace R, Schuster N: Quantification of conjunctival vascular reaction by digital imaging. J Allergy Clin Immunol 1996;98:495–500.

21 Wang de Y: Management of allergic rhinitis in general practitioners. Asia Pac Allergy 2012;2:233–236.

22 Riechelmann H, Bachert C, Goldschmidt O, Hauswald B, Klimek L, Schlenter WW, Tasman AJ, Wagenmann M: Application of the nasal provocation test on diseases of the upper airways. Position paper of the German Society for Allergology and Clinical Immunology (ENT section) in cooperation with the Working Team for Clinical Immunology (in German). Laryngorhinootologie 2003;82:183–188.

23 Malm L, Gerth van Wijk R, Bachert C: Guidelines for nasal provocations with aspects on nasal patency, airflow, and airflow resistance. International Committee on Objective Assessment of the Nasal Airways, International Rhinologic Society. Rhinology 2000;38:1–6.
24 Carney A, Bateman N, Jones N: Reliable and reproducible anterior active rhinomanometry for the assessment of unilateral nasal resistance. Clin Otolaryngol Allied Sci 2001;25:499–503.
25 Sipila J, Suonpaa J, Laippala P: Evaluation of nasal resistance data in active anterior rhinomanometry with special reference to clinical usefulness and test-retest analysis. Clinical Otolaryngol Allied Sci 1992;17:170–177.
26 Thulesius HL, Cervin A, Jessen M: Can we always trust rhinomanometry? Rhinology 2011;49:46–52.
27 Fukushima A, Tomita T: Image analyses of the kinetic changes of conjunctival hyperemia in histamine-induced conjunctivitis in guinea pigs. Cornea 2009;28:694–698.
28 Yoneda T, Sumi T, Takahashi A, Hoshikawa Y, Kobayashi M, Fukushima A: Automated hyperemia analysis software: reliability and reproducibility in healthy subjects. Jpn J Ophthalmol 2012;56:1–7.
29 Owen CG, Ellis TJ, Woodward EG: A comparison of manual and automated methods of measuring conjunctival vessel widths from photographic and digital images. Ophthalmic Physiol Opt 2004;24:74–81.
30 Fieguth P, Simpson T: Automated measurement of bulbar redness. Invest Ophthalmol Vis Sci 2002;43:340–347.
31 Peterson RC, Wolffsohn JS: Objective grading of the anterior eye. Optom Vis Sci 2009;86:273–278.
32 Peterson RC, Wolffsohn JS: The effect of digital image resolution and compression on anterior eye imaging. Br J Ophthalmol 2005;89:828–830.
33 Wolffsohn J: Incremental nature of anterior eye grading scales determined by objective image analysis. Br J Ophthalmol 2004;88:1434–1438.
34 Wolffsohn JS, Purslow C: Clinical monitoring of ocular physiology using digital image analysis. Cont Lens Anterior Eye 2003;26:27–35.
35 Schulze MM, Hutchings N, Simpson TL: The use of fractal analysis and photometry to estimate the accuracy of bulbar redness grading scales. Invest Ophthalmol Vis Sci 2008;49:1398–1406.
36 Papas EB: Key factors in the subjective and objective assessment of conjunctival erythema. Invest Ophthalmol Vis Sci 2000;41:687–691.
37 Kumar S, Yogesan K, Constable IJ: Telemedical diagnosis of anterior segment eye diseases: validation of digital slit-lamp still images. Eye 2008;23:652–660.
38 Bock F, Onderka J, Hos D, Horn F, Martus P, Cursiefen C: Improved semiautomatic method for morphometry of angiogenesis and lymphangiogenesis in corneal flatmounts. Exp Eye Res 2008;87:462–470.
39 Fallah Tafti MR, Khosravifard K, Mohammadpour M, Hashemian MN, Kiarudi MY: Efficacy of intralesional bevacizumab injection in decreasing pterygium size. Cornea 2011;30:127–129.
40 Mandalos A, Tsakpinis D, Karayannopoulou G, Tsinopoulos I, Karkavelas G, Chalvatzis N, Dimitrakos S: The effect of subconjunctival ranibizumab on primary pterygium: a pilot study. Cornea 2010;29:1373–1379.
41 Suner S, Crawford G, McMurdy J, Jay G: Non-invasive determination of hemoglobin by digital photography of palpebral conjunctiva. J Emerg Med 2007;33:105–111.
42 Weinand F, Jung A, Stein A, Pfützner A, Becker R, Pavlovic S: Rotational stability of a single-piece hydrophobic acrylic intraocular lens: new method for high-precision rotation control. J Cataract Refract Surg 2007;33:800–803.
43 Kawai M, Yamada M, Kawashima M, Inoue M, Goto E, Mashima Y, Tsubota K: Quantitative evaluation of tear meniscus height from fluorescein photographs. Cornea 2007;26:403–406.
44 Manni G, Centofanti M, Parravano M, Oddone F, Bucci MG: A 6-month randomized clinical trial of bimatoprost 0.03% versus the association of timolol 0.5% and latanoprost 0.005% in glaucomatous patients. Graefes Arch Clin Exp Ophthalmol 2004;242:767–770.
45 Setaro M, Sparavigna A: Quantification of erythema using digital camera and computer-based colour image analysis: a multicentre study. Skin Res Technol 2002;8:84–88.
46 dos Santos RV, Titus RG, Lima HC: Objective evaluation of skin prick test reactions using digital photography. Skin Res Technol 2007;13:148–153.
47 Holm O, Lindell E, Malm L: Colour change in the human histamine wheal made visible by LYYN: a technique to enhance colour differences. Skin Res Technol 2010;16:385–389.
48 Scherer K, Grize L, Schindler C, Surber C, Bircher AJ: Reaction pattern to histamine and codeine in a human intradermal skin test model. Clin Exp Allergy 2007;37:39–46.
49 Xing C, Xu A: The effect of combined calcipotriol and betamethasone dipropionate ointment in the treatment of vitiligo: an open, uncontrolled trial. J Drugs Dermatol 2012;11:e52–e54.
50 Skvara H, Burnett P, Jones J, Duschek N, Plassmann P, Thirion J-P: Quantification of skin lesions with a 3D stereovision camera system: validation and clinical applications. Skin Res Technol 2013;19:e182–e190.
51 Tremp M, Knafla I, Burg G, Wuethrich B, Schmid-Grendelmeier P: 'EASIdig' – a digital tool to document disease activity in atopic dermatitis. Dermatology 2011;223:68–73.
52 Emer JJ, Frankel A, Sohn A, Lebwohl M: A bilateral comparison study of pimecrolimus cream 1% and a topical medical device cream in the treatment of patients with atopic dermatitis. J Drugs Dermatol 2011;10:735–743.
53 Conrad J, Philipps M, Oertel J: High-definition imaging in endoscopic transsphenoidal pituitary surgery. Am J Rhinol Allergy 2011;25:e13–e17.
54 Klimek L, Mösges R: Computer-assistierte chirurgie (CAS) in der HNO-heilkunde. Laryngorhinootologie 1998;77:275–282.
55 Belli J, Chaves T, Oliveira A, Grossi D: Analysis of body posture in children with mild to moderate asthma. Eur J Pediatr 2009;168:1207–1216.
56 Ricieri DDV, Rosário NA, Costa JR: Chest diameter ratios for detecting static hyperinflation in children using photogrammetry. J Pediatr 2008;84:410–415.
57 Elbeshbeshy B, Trepman E: Digital photography in orthopaedic surgery. Foot Ankle Int 2001;22:67–74.
58 Thomas JR, Tardy ME Jr, Przekop H: Uniform photographic documentation in facial plastic surgery. Otolaryngol Clin North Am 1980;13:367–381.
59 Waite KJ: Blackley and the development of hay fever as a disease of civilization in the nineteenth century. Med Hist 1995;39:186–196.
60 Chelminska M, Niedoszytko M, Jassem E: Clinical value of conjunctival allergen challenge in diagnosing allergic conjunctivitis related to latex. Int Arch Allergy Immunol 2011;154:149–154.
61 Ciprandi G, Buscaglia S, Marchesi E, Danzig M, Cuss F, Canonica GW: Protective effect of loratadine on late phase reaction induced by conjunctival provocation test. Int Arch Allergy Immunology 1993;100:185–189.
62 Eichel A, Rohdenburg C, Shah-Hosseini K, Mösges R: Dose-finding study for the efficacy of lais grass pollen tabletts on patients with grass pollen induced allergic rhinoconjunctivits. Allergo J 2012;21:69.
63 Nagel E, Vilser W, Lanzl I: Online human conjunctival vessel diameter analysis: a clinical-methodical study. Clin Hemorheol Microcirc 2003;28:221–227.
64 European Medicines Agency: Guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases. London, EMA, 2008.
65 Nunez R, Carballada F, Lombardero M, Jimeno L, Boquete M: Profilin as an aeroallergen by means of conjunctival allergen challenge with purified date palm profilin. Int Arch Allergy Immunol 2012;158:115–119.
66 Abelson MB, Chambers WA, Smith LM: Conjunctival allergen challenge: a clinical approach to studying allergic conjunctivitis. Arch Ophthalmol 1990;108:84–88.
67 Bista SR, Sárándi I, Dogan S, Astvatsatourov A, Mösges R, Deserno TM: Automatic conjunctival provocation test combining hough circle transform and self-calibrated color measurements. Proc SPIE 2013;8670:86702J.