Validation of AI-based software for objectification of conjunctival provocation test results in routine examinations and clinical studies

Yury Yarin[1]*, Alexandra Kalaitzidi[1], Kira Bodrova[1], Yannis Kalaidzidis[2]*

[1] Clinic for ENT and Allergology Dr. Yarin, Dresden, Germany
[2] Max Plank Institute for Cell Biology and Genetics, Dresden, Germany
*Correspondence to Yury Yarin dr.yarin@gmail.com and Yannis Kalaidzidis kalaidzi@mpi-cbg.de

Key words: Allergy, Conjunctival Provocation Test, Deep Learning, AI

Abstracts: (*250 words*)

1. Background:

Provocation tests are widely used in allergology to objectively reveal patient's sensitivity to specific allergens. The objective quantification of an allergic reaction is a crucial characteristic of the tests. Due to the absence of objective quantitative measurements the conjunctival provocation test (CPT) is a less frequently used method despite its sensitivity and simplicity. We developed a new method AllergoEye based on AI for quantitative evaluation of conjunctival allergic reactions and validated it in a clinical study.

2. Methods:

AllergoEye was implemented as a two component system. The first component is based on an Android smartphone camera for imaging the patient's eye and the second one is PC based for image analysis and quantification. For AllergoEye validation an open-label, prospective, monocentric study was carried out on 41 patients. Standardized CPT was performed with sequential titration of grass allergen in 4 dilutions with the reaction evaluated by subjective/qualitative symptom scores and by quantitative AllergoEye scores.

3. Results:

AllergoEye demonstrated high sensitivity (98%) and specificity (90%) as compared with human-estimated allergic reaction. Tuning cut-off threshold allowed to increase the specificity of AllergoEye to 97%, where correlation between detected sensitivity to allergen and sIgE CAP-class becomes obvious. Strikingly, such correlation was not found with sensitivity to allergen detected by subjective and qualitative symptom scores.

4. Conclusion:

Clinical validation demonstrated that AllergoEye is a sensitive and efficient instrument for objective measurement of allergic reactions in CPT for clinical studies as well as for routine therapy control.

1. Introduction

Provocation tests in allergology play an especially important role in the clinical cases of patients with allergic symptoms who have negative skin reactions and negative Carrier-Polymer-System (CAP) test results. The positive provocation reactions in this case are an objective proof of allergy. On the other hand, negative results of provocation in patients with positive skin tests and/or CAP without active allergic symptoms can reveal a sensitization without clinical relevance. Therefore, with both positive and negative outcomes, provocation tests support the decision of the necessity of therapy.

Among other provocation tests the conjunctival provocation test (CPT) is simpler than others without compromising accuracy and sensitivity [1,2]. In the last years the CPT is used for diagnosis as well as for therapy control and is becoming a crucial instrument of clinical studies [4,10,11]. However, despite a hundred years of CPT history [3,4], its role is still limited due to the main drawback - the subjective character of the evaluation of outcome. Nevertheless, the CPT has the same level of clinical importance as the nasal provocation test, which includes an objective instrumental evaluation of outcome [1,4,5]. Multiple clinical studies demonstrated a high level of concordance between these two methods [6, 7, 8, 9].
Up to date the main method of evaluating the CPT is a summation symptom score (SSS) [2]. SSS is the sum of categorically (none (0), mild (1), moderate (2), severe (3), intolerable(4)) estimated symptoms, part of which are subjective patient estimation of allergy symptoms (like eye itching, irritation and tearing) and eye redness subjectively estimated by medical staff (observers) carrying out the test [4,5,6].

Modern evidence-based medicine requires the quantitative reproducible evaluation of test results for diagnostics as well as for assessments of the effectiveness of therapy carried out. To overcome the CPT drawback, several approaches were proposed for the quantification of CPT-outcomes [10, 17, 19, 22]. The readout of quantitative CTP methods is based on eye redness measurement on digital images. The redness quantification can be divided in two approaches. First approach is based on calculation of apparent area of vessels in the sclera [17, 26]. Therefore, vessels were 1) contrasted, 2) segmented and then the relative area of sclera covered by segmented vessels was taken as a redness score. However, the partial observation of the sclera and different eye ball orientation toward the camera on sequential observations could degrade the accuracy of the method. Additionally, diffuse redness, which is related to the widening of capillary that could not be resolved on non-microscopic images,

is not captured by this method. The second approach calculates the redness for each pixel [10, 22]. The mean redness of the sclera [10] or histogram of redness distribution [22] was then used as a redness score. In this case, one has to note, that there are multiple definitions of redness [25, 10, 23] that could lead to different sensitivities of the method. The crucial problem of per-pixel redness estimation is its dependency on reproducible color balance of environment illumination and automated white balance of digital images. The limited application of all mentioned methods is either due to semi-automated image analysis [10, 16, 17] or due to absence of automated correction for illumination color changes (white balance) [22] or both of them [10, 16, 17, 22].

At the end of the last decade, the breakthrough in the Deep Learning approach brought Artificial Intelligence methods into wide practice. In medicine it was applied in diagnostics, image reconstruction in radiology modalities, emergency care devices etc. [12,13,14]. However, its usage in allergology is still limited. Recently we developed a new approach for qualitative evaluation of CPT results for diagnostics of allergic rhinoconjunctivitis and implemented it in high-throughput software platform AllergoEye, which comprises of Deep Neural Networks for automated image analysis and symptom evaluation [23]. For the implementation of AllergoEye in broad clinical practice the validation study for quantitative estimation of sensitivity and specificity of the method was required. In the presented work we demonstrate the results of the clinical validation and efficiency of AllergoEye on a cohort of patients with grass pollen allergy.

2. Methods

2.1 Approvals and ethics

The study was approved by the Ethics Committee (IEC) of the State Chamber of Physicians of Saxony, Germany (IEC number EK-BR-111/21-1). The study was conducted in accordance with local regulations, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (International Conference on Harmonization (ICH)–Good Clinical Practices (GCP)) and the Declaration of Helsinki [15].

106

107 *2.2. Study design*

108

109 This open-label, prospective, monocentric study was carried out in the Clinic of
110 Otorhinolaryngology and Allergology Dr. Yury Yarin (Dresden, Germany) in 2021. Female
111 and male patients between 18 and 75 years of age were included in this study. The study on
112 these grass pollen-allergic patients was performed outside of the grass pollen season and
113 consisted of 2 visits. Inclusion in the study took place after the patient signed the informed
114 consent prior to any study procedures being performed. For participation in the study the
115 patients were required to have a history of moderate or severe seasonal allergic
116 rhinoconjunctivitis (AR) with or without seasonal controlled asthma during the three grass
117 pollen seasons in the past. For objectivities of the allergy as criteria for the inclusion in a
118 study, a skin prick test and grass pollen–specific IgE (sIgE) antibodies were performed. A
119 wheal diameter ≥3 mm for a grass pollen allergen solution and sIgE antibodies > 0.01 kU/L
120 were a main necessary condition for study inclusion. Exclusion criteria were adapted from:
121 guidelines for daily practice [4] (see attachment 1).
122 During the first screening visit, the allergic anamneses was collected, the inclusion/exclusion
123 criteria were checked. After that the skin prick test was performed and a blood sample for the
124 antibody was taken. During the second visit, CPT was performed and estimated by summation
125 symptoms scores (SSS) [16] which were calculated at the end of the test by medical
126 personnel.
127 Simultaneously to the CPT procedure the results of patient's reactions on each applied
128 dilution were additionally recorded by AllergoEye. The results of the CPT and of AllergoEye
129 were joined into a merged database and statistical analysis was performed.

130

131 *2.3. Titrated CPT Procedure*

132 The CPTs and AllergoEye recording were conducted by the same investigator. Standardized
133 lyophilized allergen extracts for CPT with concentration in Histamine Equivalent Potency
134 units (HEP) 30 HEP/ml containing grass group supplied by Laboratorios LETI S.L. Madrid,
135 Spain were used. Reconstitution CPT stock solution was performed by drawing 4.8 mL from
136 a vial with 5.0 ± 0.2 mL of solvent and dispensing into a vial with grass extract. Further
137 dilution steps were 1:1000, 1:100, 1:10. A diluent will serve as negative control. The CPT was
138 performed by the following protocol: At the first step of the CPT, a negative control solution
139 (without allergens) is applied to the right eye. In case of a negative reaction after 5 minutes,
140 an allergen solution is applied stepwise to the left conjunctival sac in predefined increased
141 concentrations 1:1000, 1:100, 1:10 and stock solution. Between each application a 5 minutes
142 interval was held. The symptoms were documented in a standardized form [16] (see Table 1
143 in Suppl.Fig.3) in 5 minutes after allergen application. The four symptoms of tearing, itching,
144 irritations and redness were analyzed. Each of these symptoms was estimated and categorized

on a scale from 0 to 3, whereby, 0 = no reaction, 1 = mild reaction, 2 = moderate reaction, and 3 = strong reaction. Total score SSS was summarized from the meaning of all 4 symptoms subjective (itching, irritation and tearing) and objective (redness evaluated by medical staff) and built the value ranging from 0 to 12. The protocol was stopped when either medical personnel detected redness of the treated eye or maximum allergen concentration was achieved. Finally the allergen-provoked eye was treated with antihistaminic eye drops (Ketotifen).

2.4. AllergoEye - measurements and protocol

During the CPT we described above, both patient's eyes were imaged 5 min after each provocation, which was done on left eye, with sequentially increased concentration of the grass CPT solution. Control images were taken also before the test and after the treatment of the right eye with a control solution (test solution without allergen). The eyes were imaged at 3 positions (looking straight forward, to the right and to the left) for maximum coverage of the sclera. To control the illumination condition, the images were taken with a special mask (see Figure 1A) with continuous white LED illumination. At the end of the test the images were sent to a PC where the redness was evaluated by AllergoEye software.

2.5. AllergoEye - technical description

AllergoEye was implemented as a two component system. One part is based on Android smartphone HONOR 20 Pro (HUAWEI) and uses the build-in high-resolution photocamera for imaging the patient eye's reaction to the allergen (Figure 1a). The second part is a PC based software system, including the patient database, the communication module for exchanging data with multiple smartphones and a deep neural network for image analysis and redness evaluation. In short, patient's data and measurement details (allergen, dilution, exposure time) are transferred from PC to the mobile phone. The nurse acquires images of two eyes and wireless transfers them to the PC. There, the deep neural network is used to recognize and segment the iris and sclera. Additional control of coupling of iris and sclera as well as the presence of two open eyes was performed to exclude segmentation artifacts. We proposed a new method of per-pixel redness of sclera (R) evaluation as:

$$R = \frac{27}{4}\left(r - \sqrt{\frac{g^2 + b^2}{2}}\right)\sqrt{r^2 + g^2 + b^2}\left(1 - \sqrt{r^2 + g^2 + b^2}\right)^2 \quad (1)$$

where r, g, b are normalized red, green and blue intensities respectively. The mean redness and distribution of redness in the sclera are calculated for both treated and non-treated eyes

from all 3 eye positions. To take in account the inhomogeneity of the sclera reaction to the allergen (see Figure 1b), the distribution of squared gradients of redness with 4 spatial steps were calculated.

Despite of usage of a special mask (see Figure 1a), in some cases the changes of illumination in the room during the examination led to the shift of white color balance of images (compare right and left images on Figure 1b). To make our measurement robust to the illumination change, we used the sclera of the untreated (right) eye as an internal control. For this end, the redness distributions of both eyes were normalized (shifted) in such a way, that the redness distribution of the right eye on test and control images had a maximum possible overlap. After such normalization the difference between redness of the left eye in test and control conditions were used as AllergoEye scores.

2.6. Statistical Analysis

For significance estimation the Student-t test was used. * denotes $p_{value} < 0.05$, ** denotes $p_{value} < 0.01$, *** denotes $p_{value} < 0.001$. Statistical analysis and graph generation was performed by model analysis software FitModel (http://pluk.mpi-cbg.de/projects/fitmodel).

3. Results

First we characterized the distribution of patients in the study (N=41) by eye sensitivity to the grass allergen. As shown in Figure 2, most of the patients (n=16; 40%) exhibited a reaction at allergen dilution 1:10, whereas only 4 (10%) were sensitive to dilution 1:1000. It was found that 3 (8%) patients did not reveal any eye reaction for provocation at highest concentration (stock solution) despite anamneses record of moderate to severe rhinoconjunctivitis.

Next we characterized AllergoEye's sensitivity/specificity. For this end we built a scatterplot for 114 measurements (Figure 3a, blue points). Each point represents one measurement, where the x-value is human-estimated redness score (from 0 to 3) and the y-value is AllergoEye score (from 0 to 3.7). Then we gradually changed the threshold value from 0 to 3.7, and for each threshold value we labeled the measurements as positive (above threshold) or negative (below threshold). As the ground-truth labels we took human-assigned scores. Therefore, a positive value was labelled as true-positive if it corresponds to human-assigned non-zero score; otherwise it was labelled as false-positive. The true- and false-negative labels were assigned in a similar way. From these labels we constructed a ROC-curve for AllergoEye, where for each threshold value we calculated

$$Sensetivity = \frac{\text{true-positive}}{\text{true-positive} + \text{false-negative}} \tag{2}$$

$$Specificity = \frac{\text{true-positive}}{\text{true-positive} + \text{false-positive}} \tag{3}$$

The result is drawn as black dots on Figure 3b and presented in table on Figure 3d. The ROC-curve was smoothed (red line on Figure 3b) and the optimal threshold value was found to be 0.3 as denoted by red dot on graph Figure 3b, highlighted on table Figure 3d and is shown as a horizontal line on Figure 3a.

We found that 15 AllergoEye measurements were labelled as "false" at optimal threshold (red and green bars on Figure 3a, Suppl. Fig.1). The human-estimated scores of these measurements were re-evaluated by three independent experts. Re-evaluation changed human-estimated scores in 9 cases (marked green in Suppl.Fig.1) while 6 were confirmed (marked red in Suppl.Fig.1). In summary, AllergoEye was found to be more accurate than human operator (6 mistakes vs. 9 mistakes respectively). After re-evaluation a new ROC-curve was build (Figure 3c,e). From table on Figure 3e, we found that AllergoEye sensitivity toward eye redness is as high as 97.7% with 90% specificity.

229 Since usually allergic reaction is estimated as the sum of subjective patient's symptoms score
230 and objective redness score [16] (SSS), we decided to compare the AllergoEye score with
231 SSS as a ground-truth. The scatter-plot of 114 measurements with SSS scores (from 0 to 12)
232 on x-axis and AllergoEye score (from 0 to 3.7) on y-axis is drawn on Figure 4a by blue dots.
233 Similar to Figure 3a-c, the ROC-curve was plotted on Figure 4b,c. Interestingly, the optimal
234 threshold value was found to be the same (0.3) as in redness only based ground-truth.
235 However, the sensitivity and specificity were lower (~87%) due to the presence of non-
236 redness related subjective symptoms scores.
237 In order to find the objectiveness of AllergoEye we decided to compare the patient's
238 sensitivity as determined by the AllergoEye score with immunoglobulin sIgE concentration in
239 their blood. The sIgE concentration was scored in CAP-classes [27]. The characterization of
240 patients in the study by CAP-classes is shown on Figure 5a. Most of patients have CAP-class
241 3 and 4, however, 2 patients have CAP-class 6 and 2 patients have CAP-class 0 (sIgE is
242 slightly below the low boundary of CAP-class 1). We divided patients by sensitivity category
243 according to dilutions where the allergic reaction was detected according to criteria SSS ≥ 3
244 or human-evaluated redness score > 1. For each category, the mean CAP-classes were
245 calculated. In line with previous reports [28, 29], we did not find significant correlation
246 between mean CAP-classes and patient's allergic sensitivity as measured by SSS (Figure 5b).
247 Then we built graphs for mean CAP-classes vs. AllergoEye-evaluated allergy sensitivity for
248 the threshold value 0.3 (Figure 5c) and threshold value 0.5 (Figure 5d). Surprisingly, even
249 with AllergoEye threshold value 0.3 we found statistically significant difference between
250 CAP-classes of sensitive and insensitive patients (Figure 5c marked by stars). With
251 AllergoEye threshold value 0.5 (that corresponds to specificity 97% to SSS, Figure 4c) we
252 revealed a statistically significant dependency between CAP-classes and patient's allergy
253 sensitivity (Figure 5d, marked by stars).
254

255 **4. Discussion**

256 Recent breakthrough in deep learning resulted in wide application of artificial intelligence
257 (AI) methods in engineering, science and medicine. Most successful AI applications were
258 achieved in the field of image analysis (e.g. radiology, diagnostics etc.). However, AI
259 application in allergology is rather limited.
260 The quantitative analysis of CPT in clinical practices is a long standing challenge in
261 allergology. We addressed this challenge by AI-based methods for image analysis and a new
262 method of eye redness evaluation. Up to date, most attempts to quantify the results of the CPT

were based on the quantitative analysis of eye reaction on allergen in the manually segmented region of interest (ROI) in the sclera, either by measuring the apparent area of vessels [10, 17] or by per-pixel redness evaluation [10, 22]. Although succeeding in clinical studies, the methods based on manual ROI selection was too labor intensive for wide practices and methods based on apparent area of vessels was insensitive to the diffuse redness reaction of the sclera. Up to date, by best of our knowledge, only the work of Sirazitdinova with co-authors was published where automated sclera segmentation was implemented (by random forest method) [22]. However, similarly to previous works [10, 16, 17, 26] the method of Sirazitdinova et al. [22] requires matching automatically selected keypoints on sclera before and after provocation, which makes it too complicated for general practice usage. Indeed, such matching requires an accurate repetitive eye ball orientation from the patient during image acquisition, which significantly complicates the process of measurement and in a case of operator carelessness could lead to a significant degradation of CPT quantification accuracy.

Further, the measuring of the sclera redness is also impeded by the variation in illumination and white balance in digital cameras. During the allergen titration time that can take tens of minutes, the environment illumination conditions could change. Additional devices like a facemask were very useful, but unfortunately do not allow to totally suppress the white balance fluctuations. None of the previously proposed methods handle the changes in environment illumination conditions in automated manner [10, 17, 22, 26].

We developed a method that uses 1) a deep neuronal network (33 layers) for automated segmentation of iris and sclera on the image, which exclude manual ROI selection and as such increases the throughput of method; 2) a new formula (1) for brightness-independent redness per-pixel evaluation; 3) sclera redness score based on pixels redness distribution which does not require keypoints matching before and after provocation; 4) an automated white balance correction based on the untreated eye as an inner in-image control that makes method robust toward changes in environment illumination conditions. It is worth repeating, that our method of allergic reaction evaluation is not based on widening macroscopically recognizable vessels only, but also captures changes in a capillary network that manifest itself in the diffuse redness of sclera.

The image processing and redness evaluation takes ~15 sec. on PC (i7, Windows 10). In rare cases, when the AI fails in an accurate segmentation of sclera, manual correction of ROI with recalculation can be performed in a few seconds. The software stores subjective symptoms

297 and human-based redness evaluation of redness in the relational database. Cases with
298 contradiction between human-based redness evaluation and AllergoeEye scores are easy to
299 find by database query and re-evaluate if it is necessary.
300 Nowadays, the gold standard of rating the results of CPTs is the so-called summation
301 symptom score (SSS), although there are more than one of score definitions that are used
302 today [16]. However, the SSS includes subjective feelings of patients and fails to reveal a
303 correlation with titer of sIgE as it was previously reported in literature [28, 29], where
304 correlation was found for very high sIgE titers and high SSS only. Such absence of
305 correlation between CAP-classes and SSS was confirmed in our study (Figure 5b). One of
306 possible explanations of absence of correlation between SSS and CAP-classes for middle
307 sensitive patients may be the personal ability to tolerate unpleasant feelings which differ
308 between patients that result in wide variation in subjective score values. Therefore, SSS based
309 methods only partially satisfy the requirement for evidence-based medicine, where especially
310 accurate numerical values are essential in clinical studies. In contrast, the AllergoEye scores
311 revealed an evident correlation between sIgE CAP-classes and patient's sensitivity to allergen
312 provocation (Figure 5d). Therefore, AllergoEye scores alone as well as incorporated into SSS
313 [16] could provide and objective method for clinical studies and control therapy in general
314 allergological practice.
315 Validating diagnostic methods in medicine is an essential task in new development. The main
316 objective of this study was the validation of the AllergoEye system by determining the
317 sensitivity and specificity curve, the so-called ROC curve, and the selection of optimal and
318 strict thresholds for allergic response evaluation. This validation was based on two
319 considerations. First, we compared the dilution of the CPT solution that triggered a reaction as
320 detected by the redness determined by medical staff vs. redness detected by AllergoEye
321 (Figure 3b,c). Second, we compared the dilution of the CPT solution that triggered reaction as
322 detected by the redness determined by SSS vs. redness detected by AllergoEye (Figure 4b).
323 Both comparisons showed a high sensitivity and specificity of the AllergoEye (Figure 4b-d;
324 Figure 5 b,c) as an instrument for redness and allergy assessment. In addition, the superiority
325 of the AllergoEye over human and SSS estimation in terms of accuracy of obtained results in
326 terms of false positive and negative results was clearly demonstrated.
327 Selecting the right patients for the clinical trials in allergology overall and also using CPT is
328 an outstanding problem. The challenge is the large variability of the subjective symptoms of
329 CPT and absence of a clear correlation between these symptoms and objective parameters, in
330 particular the sIgE titer. Based on the AllergoEye ROC curves, the required specificity can be

chosen from tables presented in Figures 4 and 5. The strict specificity allows to objectively select high sensitivity patients. At the same time the sensitivity value allows to estimate the number of patients that have to be screened that is essential for clinical study design. In a clinical trial AllergoEye could be used for quantitative monitoring of the efficiency of allergy therapy, e.g. hyposensitization.

5. Conclusion

AllergoEye, an artificial intelligence-based platform for quantification of conjunctival provocation tests, showed high sensitivity and specificity compared to both the RMS, the redness of eye as measured by medical personnel (> 95%) and the SSS score, which represents subjective patient reports (> 86 %). It showed a clear and statistically significant correlation of the invoked allergic reaction in the eye with the IgE-CAP classes in the blood. Clinical validation demonstrated that AllergoEye is a sensitive and efficient instrument for objective evaluation of allergic reactions. It could be used for patient selection and controlling the treatment efficiency in clinical studies as well as for diagnostic and therapy control in routine allergological practice.

6. Acknowledgments

We thank Prof. Dr. Ralf Mösges for critical review of our study, Laboratorios LETI S.L. Madrid for CPT reagents, Dr. Andreas Bilstein for useful discussions and help in preparing clinical study and Alexander Kalaidzidis for software development.

The authors declare no competing financial interests.

Author Contributions: Conceptualization, Y.Yarin and Y.Kalaidzidis; Clinical study Y.Yarin and K. Bodrova; Documentation A.Kalaidzidi; Statistical analysis Y.Kalaidzidis.

Literature

1. Hauswald B, Yarin YM. Moderne Diagnostik und Therapie der Rhinitis allergica [Modern Diagnosis and Therapy of the rhinitis allergica]. Laryngorhinootologie. 2015 May;94(5):331-345. German. doi: 10.1055/s-0035-1548900. Epub 2015 May 7. PMID: 25951448.

2. Schröder J, Mösges R. Conjunctival provocation tests: prediction of seasonal allergy. Curr Opin Allergy Clin Immunol. 2018 Oct;18(5):393-397. doi: 10.1097/ACI.0000000000000470. PMID: 30020256.

3. Fineman A.H. (1926) Studies in hypersensitiveness. XXIII. A comparative study of the intradermal, scratch, and conjunctival tests in determining the degree of pollen sensitivity. Journat of Immunotogy, 11, 465.

4. Fauquert JL, Jedrzejczak-Czechowicz M, Rondon C, Calder V, Silva D, Kvenshagen BK, Callebaut I, Allegri P, Santos N, Doan S, Perez Formigo D, Chiambaretta F, Delgado L, Leonardi A; Interest Group on Ocular Allergy (IGOA) from the European Academy of Allergy and Clinical Immunology. Conjunctival allergen provocation test : guidelines for daily practice. Allergy. 2017 Jan;72(1):43-54. doi: 10.1111/all.12986. Epub 2016 Aug 18. PMID: 27430124.

5. Shamji MH, Kappen JH, Akdis M, Jensen-Jarolim E, Knol EF, Kleine-Tebbe J, Bohle B, Chaker AM, Till SJ, Valenta R, Poulsen LK, Calderon MA, Demoly P, Pfaar O, Jacobsen L, Durham SR, Schmidt-Weber CB. Biomarkers for monitoring clinical efficacy of allergen immunotherapy for allergic rhinoconjunctivitis and allergic asthma: an EAACI Position Paper. Allergy. 2017 Aug;72(8):1156-1173. doi: 10.1111/all.13138. Epub 2017 Apr 6. PMID: 28152201.

6. 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 Jan;130(1):51-9. doi: 10.1159/000068369. PMID: 12576735.

7. 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 (Lond). 1993;7 (Pt 3):461-4. doi: 10.1038/eye.1993.93. PMID: 8224308.

8. Mosbech H, Dirksen A, Madsen F, Stahl Skov P, Weeke B. House dust mite asthma. Correlation between allergen sensitivity in various organs. Allergy. 1987 Aug;42(6):456-63. doi: 10.1111/j.1398-9995.1987.tb00363.x. PMID: 2444129.

9. Petersson G, Dreborg S, Ingestad R. Clinical history, skin prick test and RAST in the diagnosis of birch and timothy pollinosis. Allergy. 1986 Aug;41(6):398-407. doi: 10.1111/j.1398-9995.1986.tb00319.x. PMID: 3789325.

10. Sárándi I, Claßen DP, Astvatsatourov A, Pfaar O, Klimek L, Mösges R, Deserno TM. Quantitative conjunctival provocation test for controlled clinical trials. Methods Inf Med. 2014;53(4):238-44. doi: 10.3414/ME13-12-0142. Epub 2014 Jun 27. PMID: 24970698.

11. Ansotegui IJ, Melioli G, Canonica GW, Caraballo L, Villa E, Ebisawa M, Passalacqua G, Savi E, Ebo D, Gómez RM, Luengo Sánchez O, Oppenheimer JJ, Jensen-Jarolim E, Fischer DA, Haahtela T, Antila M, Bousquet JJ, Cardona V, Chiang WC, Demoly PM, DuBuske LM, Ferrer Puga M, Gerth van Wijk R, González Díaz SN, Gonzalez-Estrada A, Jares E, Kalpaklioğlu AF, Kase Tanno L, Kowalski ML, Ledford DK, Monge Ortega OP, Morais-Almeida M, Pfaar O, Poulsen LK, Pawankar R, Renz HE, Romano AG, Rosário Filho NA, Rosenwasser L, Sánchez Borges MA, Scala E, Senna GE, Sisul JC, Tang MLK, Yu-Hor Thong B, Valenta R, Wood RA, Zuberbier T. Erratum to "IgE allergy diagnostics and other relevant tests in allergy, a World Allergy Organization position paper" [World Allergy Organ J 13/2 (2020) 100080]. World Allergy Organ J. 2021 Jun 17;14(7):100557. doi: 10.1016/j.waojou.2021.100557. Erratum for: World Allergy Organ J. 2020 Feb 25;13(2):100080. PMID: 34221218; PMCID: PMC8225965.

12. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, Kim N. Deep Learning in Medical Imaging: General Overview. Korean J Radiol. 2017 Jul-Aug;18(4):570-584. doi: 10.3348/kjr.2017.18.4.570. Epub 2017 May 19. PMID: 28670152; PMCID: PMC5447633.

13. Choi J, Shin K, Jung J, Bae HJ, Kim DH, Byeon JS, Kim N. Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy. Clin Endosc. 2020 Mar;53(2):117-126. doi: 10.5946/ce.2020.054. Epub 2020 Mar 30. PMID: 32252504; PMCID: PMC7137563.

14. Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW. Application of Artificial Intelligence in Medicine: An Overview. Curr Med Sci. 2021 Dec 6. doi: 10.1007/s11596-021-2474-3. Epub ahead of print. PMID: 34874486.

15. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310:2191-2194.

16. Pfaar O, Claßen D, P, Astvatsatourov A, Klimek L, Mösges R: Reliability of a New Symptom Score in a Titrated Quantitative Conjunctival Provocation Test Supported by an Objective Photodocumentation. Int Arch Allergy Immunol 2018;176:215-224. doi: 10.1159/000487884.

17. Astvatsatourov A, Mösges R: Image-based assessment of allergic inflammation under conjunctival provocation. Stud Health Technol. Inform 2015; 213: 15–18.

18. Kjaergaard SK, Pedersen OF, Taudorf E, Mølhave 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(2):133-7. doi: 10.1007/BF00383590. PMID: 2323832.

19. Horak F, Berger U, Menapace R, Schuster N. Quantification of conjunctival vascular reaction by digital imaging. J Allergy Clin Immunol. 1996 Sep;98(3):495-500. doi: 10.1016/s0091-6749(96)70081-7. PMID: 8828525.

20. Peterson RC, Wolffsohn JS. Sensitivity and reliability of objective image analysis compared to subjective grading of bulbar hyperaemia. Br J Ophthalmol. 2007 Nov;91(11):1464-6. doi: 10.1136/bjo.2006.112680. Epub 2007 May 2. PMID: 17475716; PMCID: PMC2095410.

21. 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 Jan;56(1):1-7. doi: 10.1007/s10384-011-0107-2. Epub 2011 Dec 1. PMID: 22130522.

22. Sirazitdinova E, Gijs M, Bertens CJF, Berendschot TTJM, Nuijts RMMA, Deserno TM. Validation of Computerized Quantification of Ocular Redness. Transl Vis Sci Technol. 2019 Dec 12;8(6):31. doi: 10.1167/tvst.8.6.31. PMID: 31853426; PMCID: PMC6908135.

23. Y. Yarin, A. Kalaitzidou, K. Bodrova, I. Kalaitzidis. Die Effektivität von „AllergoEye“, eines vollautomatischen Systems basiert auf künstlicher Intelligenz, bei der Objektivierung des konjunktivalen Provokationstestes. Abstracts, 16. Deutscher Allergiekongress, Dresden, 30. September-2. Oktober 2021. *Allergo J*. 2021;30(6): 61. doi:10.1007/s15007-021-4862-1

24. D. G. Altman and J. M. Bland. Measurement in Medicine: The Analysis of Method Comparison Studies, Journal of the Royal Statistical Society. Series D (The Statistician), Sep., 1983, Vol. 32, No. 3 (Sep., 1983), pp. 307-317

25. Amparo F, Wang H, Emami-Naeini P, Karimian P, Dana R. The Ocular Redness Index: a novel automated method for measuring ocular injection. Invest Ophthalmol Vis Sci. 2013 Jul 18;54(7):4821-6. doi: 10.1167/iovs.13-12217. PMID: 23766472; PMCID: PMC3720148.

26. Dogan S, Astvatsatourov A, Deserno TM, Bock F, Shah-Hosseini K, Michels A, Mösges R. Objectifying the conjunctival provocation test: photography-based rating and digital analysis. Int Arch Allergy Immunol. 2014;163(1):59-68. doi: 10.1159/000355333. Epub 2013 Nov 16. PMID: 24248100.

27. Kleine-Tebbe J, Eickholt M, Gätjen M, Brunnée T, O'Connor A, Kunkel G. Comparison between MAGIC LITE- and CAP-system: two automated specific IgE antibody assays. Clin Exp Allergy. 1992 Apr;22(4):475-84. doi: 10.1111/j.1365-2222.1992.tb00150.x. PMID: 1611547.

28. Corsico AG, De Amici M, Ronzoni V, Giunta V, Mennitti MC, Viscardi A, Marseglia GL, Ciprandi G. Allergen-specific immunoglobulin E and allergic rhinitis severity. Allergy Rhinol (Providence). 2017 Mar 1;8(1):1-4. doi: 10.2500/ar.2017.8.0187. PMID: 28381320; PMCID: PMC5380446.

29. Lichtenstein LM, Ishizaka K, Norman PS, Sobotka AK, Hill BM. IgE antibody measurements in ragweed hay fever. Relationship to clinical severity and the results of immunotherapy. J Clin Invest. 1973 Feb;52(2):472-82. doi: 10.1172/JCI107204. PMID: 4119162; PMCID: PMC302277.

Legends

Figure 1. Workflow of PCT evaluation by AllergoEye. **A.** (left) Nurse makes image of patient's eyes by smartphone using AllergoEye mobile app (right), then image (top) is wireless transferred to PC-based AllergoEye application (bottom right), the result of analysis is transferred to mobile application for control. **B.** Sclera and iris are segmented by neural network (contoured in yellow). Sclera redness before (left image) and after (right image) provocation is shown. **Note:** real photographs were replaced by sketches in accordance to medRxiv regulation for potentially patient identifying information.

Figure 2. Distribution of 41 patients by sensitivity defined as allergen dilution, where allergic reaction was detected as either SSS ≥ 3. Left axis is number of patients; right axis is of percent patients.

Figure 3. A. Scatter plot of 114 measurements (blue dots), which were performed on 41 patients to reveal dependency between redness score estimated by medical staff (horizontal axis) and redness evaluated by AllergoEye (vertical axis). Horizontal cyan line denotes threshold equal 0.3 (see text). Red and green rectangles highlight measurements that "false negative" and "false positive" respectively. **B, D**. ROC curve (**B**) (Sensitivity vs. Specificity) as defined by threshold value gradually changing from 0.0 to 3.7 and its table presentation (**D**). Experimental data are presented by black dots; interpolation curve is drawn by red line. Optimal threshold (0.3) is marked by red dot. **C, E**. ROC curve and its table presentation similar to panels B and D, but measurements that are highlighted by red and green rectangles on panel A were re-evaluated by two independent observers and Redness score were corrected by majority voting (see Suppl. Figure 1). Sensitivity and specificity at optimal threshold are highlighted on panels **D** and **E** by cyan color.

Figure 4. A. Scatter plot of 114 measurements (blue dots), which were performed on 41 patients to reveal dependency between sum symptoms score (horizontal axis) and redness evaluated by AllergoEye (vertical axis). Horizontal cyan line denotes threshold equal 0.3. Red and green rectangles highlight measurements that "false negative" and "false positive" respectively. **B, C**. ROC curve (**B**) (Sensitivity vs. Specificity) as defined by threshold value gradually changing from 0.0 to 3.7 and its table presentation (**C**). Experimental data are presented by black dots; interpolation curve is drawn by red line. Optimal threshold (0.3) is marked by red dot. Sensitivity and specificity at optimal and stringent thresholds are highlighted on panels **C** by cyan color.

Figure 5. A. Distribution of 41 patients by sIgE CAP classes. Two patients although have CAP-class equal zero (sIgE was slightly below low boundary of CAP-class 1), but have clinically relevant symptoms. **B.** Mean CAP-class vs. sensitivity to allergen dilutions, as determined by SSS. **C**. Mean CAP-class vs. sensitivity to allergen dilutions, as determined by AllergoEye with cut-off threshold equal 0.3. **D**. Mean CAP-class vs. sensitivity to allergen dilutions, as determined by AllergoEye with cut-off threshold equal 0.5. Asterisks denote statistically significant differences, where * < 0.05; ** < 0.01; *** < 0.001.

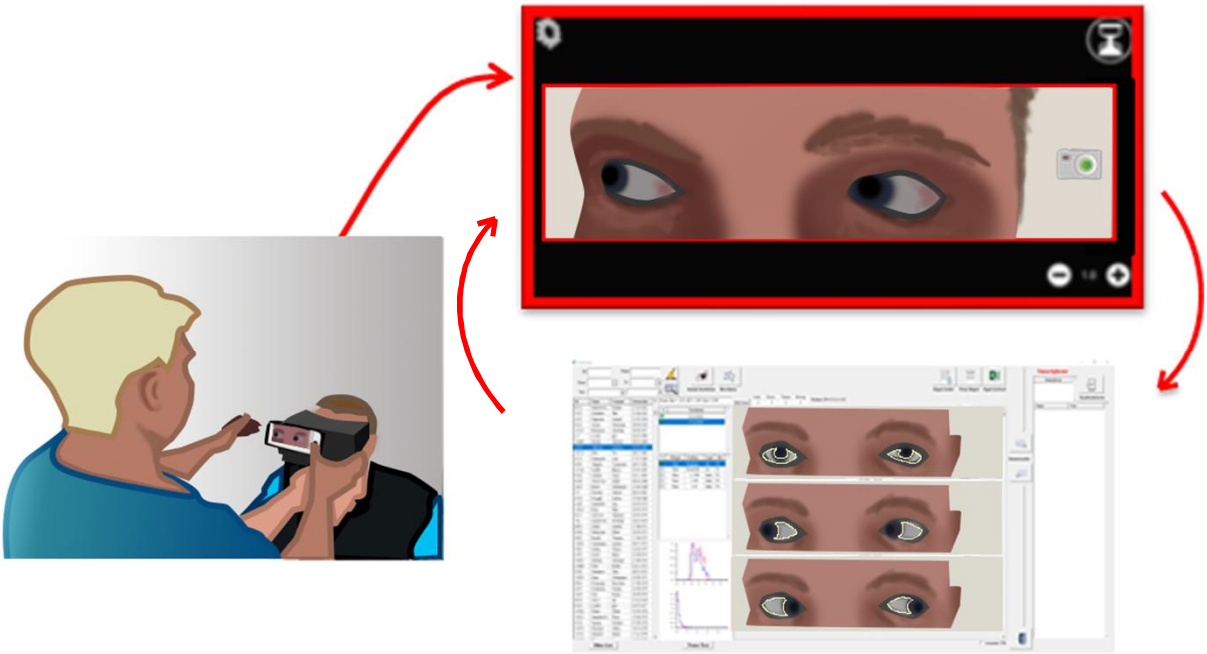
Suppl. Figure 1. Table of re-evaluation of eye redness as defined by medial staff during CPT for the measurements highlighted on Figure 3A. The redness scores that were kept unchanged is highlighted by red, those that were changed after re-evaluation is highlighted by green. The re-evaluation was performed by 2 independent observers and final decision was done by majority votes.

Suppl. Figure 2. CAP-classes distribution among patients.

Suppl. Figure 3. Example of symptom score evaluation protocol.

Figure 1

A



B

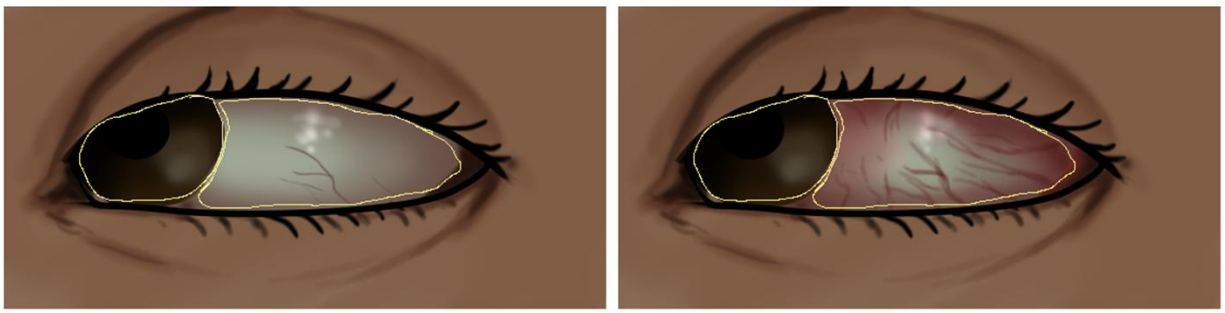


Figure 2

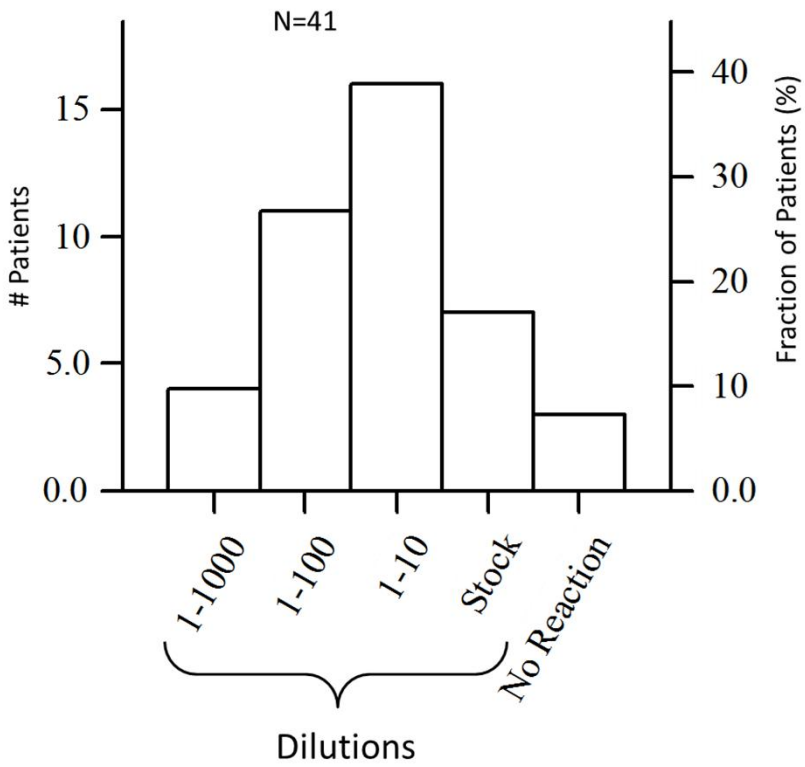
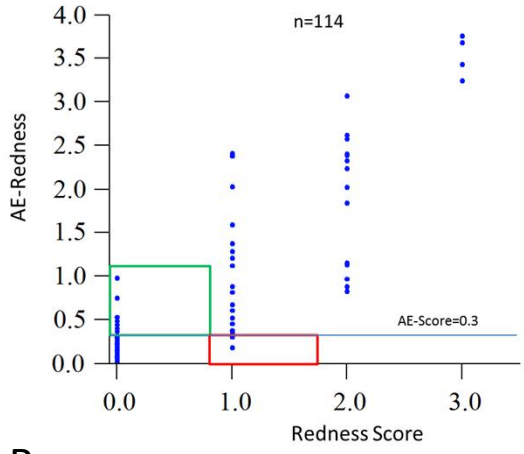
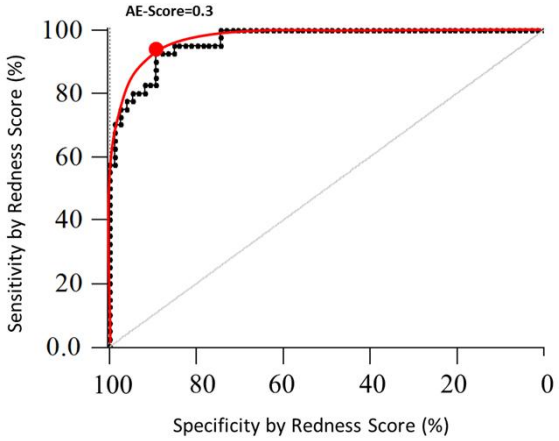

N=41
# Patients
15
10
5.0
0.0
40
30
20
10
0.0
Fraction of Patients (%)
1-1000
1-100
1-10
Stock
No Reaction
Dilutions

Figure 3

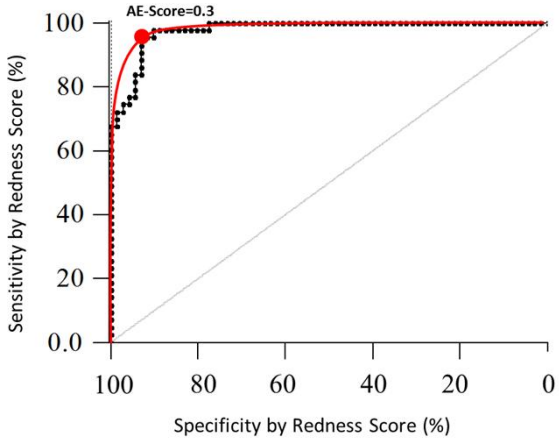
A



B



C



D

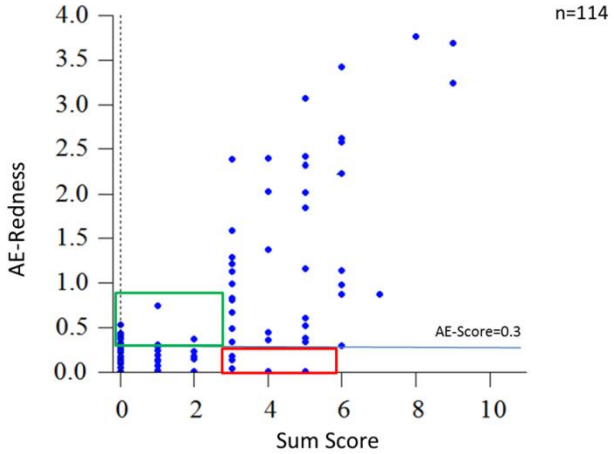
Specificity (%)	Sensitivity (%)	AE-Score threshold
0.00	100.00	0.00
60.82	100.00	0.10
75.68	95.00	0.20
82.35	96.70	0.25
85.32	95.00	0.30
91.18	87.20	0.35
93.10	84.20	0.40
94.60	77.50	0.45
95.90	75.00	0.50
98.70	70.00	0.75
100.00	57.50	1.00
100.00	35.00	2.00
100.00	10.00	3.00
100.00	1.50	3.70

E

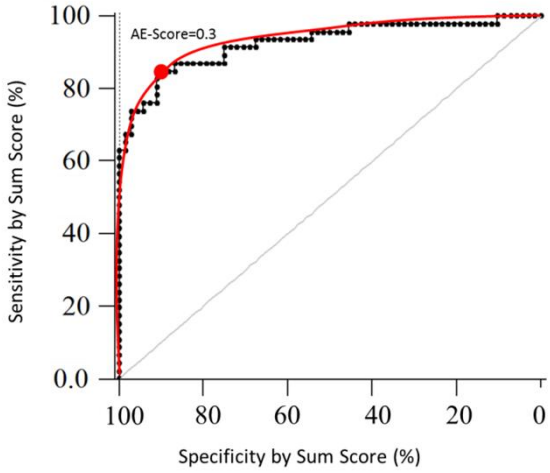
Specificity (%)	Sensitivity (%)	AE-Score threshold
0.00	100.00	0.00
74.65	100.00	0.15
76.06	100.00	0.20
86.00	98.50	0.25
90.14	97.67	0.30
93.07	83.72	0.35
94.37	78.80	0.40
95.77	76.74	0.45
97.18	74.42	0.50
99.90	65.00	0.75
100.00	53.20	1.00
100.00	32.50	2.00
100.00	9.30	3.00
100.00	2.30	3.70

Figure 4

A



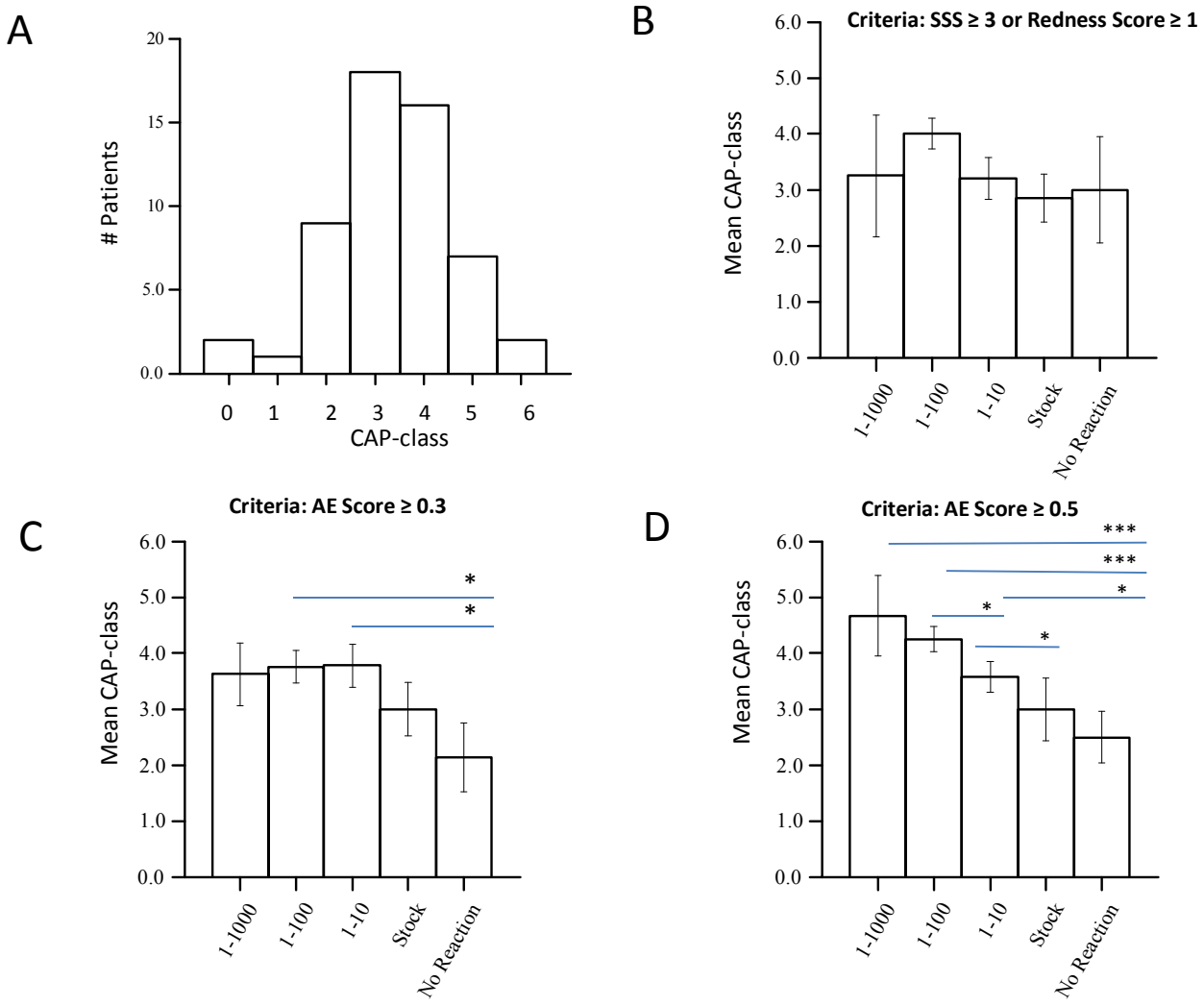
B



C

Specificity (%)	Sensitivity (%)	AE-Score threshold
0.00	100.00	0.00
61.76	93.48	0.10
76.47	86.96	0.20
82.35	86.96	0.25
86.76	86.96	0.30
91.18	76.09	0.35
95.59	73.91	0.40
97.06	71.74	0.45
97.06	67.39	0.50
100.00	63.04	0.75
100.00	50.00	1.00
100.00	30.43	2.00
100.00	8.70	3.00
100.00	2.17	3.70

Figure 5


A
Patients
CAP-class
B
Criteria: SSS ≥ 3 or Redness Score ≥ 1
Mean CAP-class
1-1000
1-100
1-10
Stock
No Reaction
C
Criteria: AE Score ≥ 0.3
Mean CAP-class
D
Criteria: AE Score ≥ 0.5
Mean CAP-class

Suppl. Fig.1

Gras 1-100	Redness was re-evaluated and found positive	TP
Gras 1-1 000	Redness was re-evaluated and found negative	FP
Gras 1-100	Redness was re-evaluated and found negative	FP
Gras 1-1 000	Redness was re-evaluated and found positive	TP
Gras 1-1 000	Redness was re-evaluated and found positive	TP
Gras 1-100	Redness was re-evaluated and found negative	TN
Gras 1-1 000	Redness was re-evaluated and found negative (red eye on start) (**CAP-class =1**)	FP
Gras 1-10	Redness was re-evaluated and found negative	TN
Gras 1-100	Redness was re-evaluated and found negative	TN
Gras 1-100	Redness was re-evaluated and found positive	TP
Gras 1-10	Redness was re-evaluated and found positive	FN
Gras Stocklösung	Redness was re-evaluated and found negative	TN
Gras 1-1 000	Redness was re-evaluated and found positive	TP
Gras 1-1 000	Redness was re-evaluated and found negative	FP
Gras 1-1 000	Redness was re-evaluated and found positive (**CAP-class = 0**)	FN

Suppl. Fig.2

	CAP Class						
	0	1	2	3	4	5	6
Male	1	1	3	8	10	1	1
Female	1	0	5	5	2	6	1

Suppl. Fig.3

Itching
- ☐ 0 = Not itchy
- ☐ 1 = Slightly itchy but not bothersome
- ☐ 2 = Itchy, bothersome and/or urge to rub eye at times
- ☐ 3 = Very itchy, very bothersome and/or urge to rub eye often

Irritation
- ☐ 0 = Not irritated
- ☐ 1 = Dry eye or other unusual feeling in eye, but not bothersome
- ☐ 2 = Gritty or foreign object feeling in eye, bothersome and/or urge to rub eye
- ☐ 3 = Burning or painful eye, very bothersome and/or difficulty keeping eye open and/or urge to rub eye often

Tearing
- ☐ 0 = Not watery
- ☐ 1 = Watery eye, but no tears and not bothersome
- ☐ 2 = Watery eye with a few tears, bothersome and/or urge to wipe eyes at times
- ☐ 3 = Watery eye with several tears, very bothersome and/or urge to wipe eye often

Redness
- ☐ 0 = No redness
- ☐ 1 = Redness in conjunctiva bulbi
- ☐ 2 = Redness in conjunctiva bulbi and tarsi
- ☐ 3 = Redness in conjunctiva bulbi and tarsi and edema