Reliability Analysis of Subjectively Graded Real-Image Pterygium Based on its Translucence Appearance between Two Experienced Clinicians – Preliminary Report

Mohd Radzi Hilmi, Khairidzan Mohd Kamal, Mohd Hafidz Ithnin
Department of Optometry and Vision Sciences, Kulliyyah of Allied Health Sciences, International Islamic University Malaysia (IIUM)
Department of Ophthalmology, Kulliyyah of Medicine, International Islamic University Malaysia (IIUM),

Corresponding author:
khairidzan@gmail.com

Abstract
This study aimed compares the reliability of subjectively graded real-image pterygium based on its translucence appearance between experienced clinicians. Thirty (30) primary pterygium images from 30 pterygium patients were captured in a standardized magnification, illumination and formatting setting as previously described. All images were projected using PowerPoint presentation™ on liquid crystal display (LCD) monitor with standard resolution. Two experienced clinicians act as a grader and grade all images based on reference images provided. For reliability testing, intra-grader assessment was repeated twice with different sequence at least a month apart between each session. Both clinicians were given a set of 30 randomized pterygium images for all sessions. Reliability testing were evaluated using paired T-test and independent T-test. Descriptive analysis revealed observer 1 obtained mean grade of 2.33 (SD = 0.758) and 2.30 (SD = 0.837) for session 1 and 2 respectively. Observer 2 obtained 2.30 (SD = 0.702) and 2.17 (SD = 0.791) for session 1 and 2 respectively. Paired T-test results showed the difference for both observers were not statistically significant for both observer 1 and 2 (P = 0.662 and P = 0.293) respectively. Reproducibility testing using Independent T-test results showed the difference between observers was not statistically significant (P = 0.769). Subjectively graded pterygium clinical grading based on its translucence appearance was repeatable and reproducible. These findings could serve as basis for future work on to evaluate performance of pterygium clinical grading based on its morphology with different level of experience and larger number of samples.

Keywords:
Pterygium, Cornea, Conjunctiva, Grading scale, Subjective
Introduction

Pterygium is a common conjunctival disorder among ocular surface pathologies. Pterygium is defined as a fibrovascular encroachment of fibrous tissue of conjunctiva which originates from bulbar conjunctiva and progress towards the central cornea (Song et al., 2017). It is an established fact that pterygium are commonly found in tropical countries due to its high temperature and UV exposure. These areas also known as ‘Pterygium belt’ which define as a location positioned between 30° of north and south of the equator. Ultraviolet (UV) exposure has been hailed as the main risk factor for its development. It is also a well-known that pterygium cause significant reduction in vision due to changes made on the anterior corneal curvature as it progresses (Song et al., 2014; Chen et al., 2015).

In 1997, Donald Tan and his co-workers (Tan et al., 1997) had proposed a subjective clinical grading which classifies pterygium based on its translucency into three groups: Grade I - atrophy, Grade II - intermediate and Grade III - fleshy. This grading has been widely used since its inceptions. However, there is lack of literature that addresses the reproducibility of the clinical grading between clinicians, although it has been applied in routine clinic. Therefore, this study aims to determine the reproducibility of subjectively graded pterygium based on its translucency appearance using real-images of pterygium between two experienced clinicians.

Methods

A cross-sectional study was conducted to evaluate the reliability of subjectively graded real-image pterygium based on its translucence appearance between two experienced clinicians. Thirty (30) primary pterygium images from 30 eyes were captured in a standardized magnification, orientation and type of illumination (Azemin et al., 2014; Che Azemin et al., 2014; Che Azemin et al., 2015; Che Azemin et al., 2016; Che Azemin et al., 2016; Hilmi et al., 2017) by an individual operator. Participants were recruited from a University eye-specialist ophthalmic centre in order to obtain images which would represent a wide range of severity of pterygium patients. Ethical approval was obtained by the International Islamic University Malaysia (IIUM) research ethical committee (IERC) (IIUM/310/G13/4/4-125) and this study conformed to the tenets of the Declaration of Helsinki. All participants were given adequate information regarding methods and risks of this study. Written consent was obtained prior to acquisition of images and other procedures.

All images were attained using a standardized protocol at 10 – 16X magnification, under diffuse white-light illumination and oriented set at 30 - 35° to the corneal surface (Hilmi et al., 2017). A high definition (HD) digital camera was used to acquire these images through a SL-990 slit-lamp biomicroscope (Costruzione Strumenti Oftalmici (CSO), Firenze, Italy). All images were deposited and analysed using an image processing software, Phoenix™ version 1.2 (Costruzione Strumenti Oftalmici (CSO), Firenze, Italy) in form of Joint Photographic Experts Group (JPEG) format.

All primary pterygium images were carefully examined based on the Donald Tan’s grading by placed into a PowerPoint presentation (Microsoft Corporation, Redmond, WA, USA) and displayed on monitor with 17-inch 1280 X 1024 resolution (Hilmi et al., 2017). All images were arbitrarily organized using a research randomization software (Urbania & Plous, 2013). The images were matched and graded based on reference images as described in previous study (Tan et al., 1997) as shown in Figure 1 by two (2) experienced clinicians.

Data Analysis

Statistical analyses were performed using IBM SPSS (Predictive analytics software) (Version 24, SPSS Inc., Chicago, IL, USA). Descriptive statistics were used to determine the mean and standard deviation (SD) of image graded by both observers. Reliability analyses were evaluated via Paired and Independent T-test. Paired T-test was employed to assess the difference in measurement between two sessions. For intra-observer reliability testing, the process was repeated twice with different sequence at least a month apart between each session. The differences between two (2) observers were evaluated using Independent T-test. Both grading session was done simultaneously, but in different occasion. The alpha significance level was set at $P < 0.05$. 
Result

Based on thirty (30) primary pterygium, descriptive analysis revealed observer 1 (KMK) obtained mean grade of 2.33 (SD = 0.758) for session 1, and mean grade of 2.30 (SD = 0.837) for session 2. Observer 2 (MRH) obtained 2.30 (SD = 0.702) for session 1, and mean grade of 2.17 (SD = 0.791) for session 2. Descriptive analysis results were summarized in Table 1. Paired T-test results showed the difference for both observers were not statistically significant for observer 1 (P = 0.662, 95% CI: -0.121, 0.188) and observer 2 (P = 0.293, 95% CI: -0.388, 0.121). Reproducibility testing using Independent T-test results showed the difference between observers was not statistically significant (P = 0.769, 95% CI: -0.196, 0.263). Both repeatability and reproducibility findings were summarized in Table 2.

The purpose of this study was to evaluate the perceptual relationship between two (2) experienced clinicians in estimating the translucence appearance of pterygium based on references images by Tan's clinical grading (Tan et al., 1997). Previous works have shown that reproducibility for subjective clinical grading is somehow convincing, with careful consideration been applied (Efron et al., 2001; Efron et al., 2002; Murphy et al., 2007). This is due to variations between individuals which can lead to bias. Study by Fieguth and Simpson (2002) had proved that individual variations can extend to approximately 50%. In this study, we found that the variations could be deemed as acceptable as the mean and standard deviations were approximately similar.

We postulate that inconsistent evaluation of a subjective clinical grading could happen due various reasons such as lack of standardization of reference images, variations in methodology, technology and experience of individuals involved in assessing its reliability. However, with incorporations of reference image as the benchmark of a clinical grading and usage of a standardized protocol, inter-observer performance can be improved, hence reducing individual variations and bias. Recent work (Hilmi et al., 2017) had demonstrated by inclusion of reference image in subjectively grading pterygium redness, the clinical grading performance is promising. Previous works (McMonnies & Chapman-Davies, 1987; Efron, 2000; Hilmi et al., 2017) had demonstrated that subjective grading is permissible and reliable provided a standard protocol was employed. This study proves that using a static image as reference image with an approximate equal level of experience between observers, the repeatability and reproducibility are very good. However, it is worth to re-evaluate the reliability of this grading in a bigger scale in terms of more pterygium images, with inclusion of variations in experience/training factor involved. Proper training could further reduce individual variations.

Conclusion

Reproducibility of pterygium clinical grading based on its translucency using real-image is reproducible. This research work can serve as basis for future work on to evaluate performance of pterygium clinical grading based on its morphology in different level of experience with larger number of samples. Different level of clinical exposure could provide differences in reliability of the grading.

Acknowledgement

This research is financially supported by International Islamic University Malaysia (IIUM) under the Research Initiative Grant Scheme (RIGS) identification number RIGS17-148-0723 and RIGS17-070-064
References


Table and Figure

Table 1: Descriptive analysis results (n = 30)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Session 1 (mean ± SD)</th>
<th>Session 2 (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1</td>
<td>2.33 ± 0.758</td>
<td>2.30 ± 0.837</td>
</tr>
<tr>
<td>Observer 2</td>
<td>2.30 ± 0.702</td>
<td>2.17 ± 0.791</td>
</tr>
</tbody>
</table>

SD: standard deviation

Table 2: Repeatability and reproducibility assessment of Tan’s pterygium clinical grading (n = 30)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Session 1 (mean ± SD)</th>
<th>Session 2 (mean ± SD)</th>
<th>P value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1</td>
<td>2.33 ± 0.758</td>
<td>2.30 ± 0.837</td>
<td>0.662*, (-0.121, 0.188)</td>
</tr>
<tr>
<td>Observer 2</td>
<td>2.30 ± 0.702</td>
<td>2.17 ± 0.791</td>
<td>0.293*, (-0.388, 0.121)</td>
</tr>
</tbody>
</table>

*: Based on Paired T-test, P = 0.05 was set as level of significance
#: Based on Independent T-test, P = 0.05 was set as level of significance

Figure 1. Reference images of Tan’s pterygium clinical grading, (a) Grade I, (b) Grade II, (c) Grade III.