

Both subscales of the Keratoconus End-Points Assessment Questionnaire have excellent test-retest reliability

Kepa Balparda, David J Galarreta¹, Claudia M. Vanegas-Ramirez², Laura A Silva-Quintero³, Mariana Botero-Giraldo³, Maria I Maya-Naranjo², Paula A Pizarro-Marín⁴

Purpose: The keratoconus end-points assessment questionnaire (KEPAQ) is a disease-specific scale designed to evaluate the quality of life in keratoconus patients and provides the measurement of both functional and emotional compromise in keratoconus. It was previously developed, tested, and validated and now we want to evaluate the test-retest reliability of the KEPAQ, in an effort to contribute evidence on its internal consistency and capability of measuring clinical state with minimal inference of random chance. **Methods:** This is a prospective analytical study, designed to evaluate the test-retest reliability of the KEPAQ through the repeated application of the questionnaire to a group of clinically stable individuals. A number of patients with a confirmed diagnosis of keratoconus underwent double application of the KEPAQ, seven days apart. Mean KEPAQ score was obtained through Rasch analysis, while test-retest reliability was evaluated through Spearman rank-order correlation and intraclass correlation coefficient. Rasch analysis was performed in JMetrik version 4.1.1 (Psychomeasurement Systems LLC; Charlottesville, VA, USA) in a MacBook Air computer running macOS Catalina version 10.15.2 (Apple Inc.; Cupertino, CA, USA). **Results:** A total of 100 patients were included. For KEPAQ-E, Spearman correlation was $R = 0.963$ while ICC was 0.981 (95% confidence interval 0.972–0.987). For KEPAQ-F, Spearman correlation was $R = 0.921$ while ICC was 0.952 (95% confidence interval 0.929–0.968). **Conclusion:** The KEPAQ is a robust, well-developed, extremely reliable scale which can be confidently used for clinical and research endeavors.

Key words: Cornea, eyes, keratoconus, ophthalmology, vision

Keratoconus is the most common primary corneal ectasia worldwide. It is characterized by a progressive distortion of the corneal anatomy, associated with a significant decrease in visual quality. Although so far there are a considerable number of surgeries and optical aids aimed at improving the visual quality of patients with keratoconus, subjects with the disease tend to show significant alterations in their ability to perform their daily tasks normally.

At present, patient-reported outcome measurements (PROMs) have gained great significance as an effective and simple mechanism to collect information on the burden of disease from a patient's point of view. This kind of instruments allows for a reliable determination of how much subjective alteration patients feel on their quality of life (QoL), according to the disease they suffer. This approach is especially important considering that visual alteration is a highly subjective experience, and that visual disturbance referred by the patient is not necessarily associated with the anatomical alteration or

with other elements directly measurable by the staff physician. Although general PROMs may be useful in some situations, disease-specific PROMs are preferred for research, as they give much more information regarding the current state of the patient suffering from a determined complex disease.

Nowadays, only two keratoconus-specific scales have been validated worldwide. The first one is keratoconus outcomes research questionnaire (KORQ) designed by Khadka *et al.*^[1] and recently studied by Kandel *et al.*^[2] The KORQ has been recently validated in Colombian population by our group.^[3] Although the KORQ exhibits adequate psychometric characteristics, a great limitation of this instrument is that it completely ignores the emotional compromise that the disease causes to the patient. To keep matters clear, it must be mentioned that "psychometrics" refers to all matters of psychological measurement, while "emotional compromise" refers to just one of the aspects of this psychometric endeavor. Therefore, all subjective aspects evaluated by a PROM are part of the realm of psychometrics. In their recent literature review, Kandel *et al.*^[4] have stressed that ectatic diseases cause a marked effect in

Department of Cornea and Refractive Surgery, Black Mammoth Surgical, Medellín, Colombia, ¹Department of Cornea and Refractive Surgery, Instituto Oftalmológico Recoletas, Valladolid, Spain, ²Department of Ophthalmology, Universidad Pontificia Bolivariana, ³Department of General Medicine, Mediacarte, Medellín, ⁴Department of General Medicine, Hospital San Juan de Dios, Santa Fe de Antioquia, Colombia

Correspondence to: Dr. Kepa Balparda, Carrera 43 # 29 35, Office 712, Clínica de Oftalmología Sandiego, Medellín, Colombia. E-mail: kb@kepabalparda.com

Received: 06-Dec-2020

Revision: 21-Mar-2021

Accepted: 29-Mar-2021

Published: 25-Aug-2021

Access this article online

Website:

www.ijo.in

DOI:

10.4103/ijo.IJO_3569_20

Quick Response Code:



This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

Cite this article as: Balparda K, Galarreta DJ, Vanegas-Ramirez CM, Silva-Quintero LA, Botero-Giraldo M, Maya-Naranjo MI, *et al.* Both subscales of the Keratoconus End-Points Assessment Questionnaire have excellent test-retest reliability. Indian J Ophthalmol 2021;69:2430-5.

emotional elements, with keratoconus patients having a greater than normal prevalence of clinical depression and different personality coping mechanisms.

In an effort to provide a tool for adequate measurement of both functional and emotional compromise in keratoconus, our research group has previously developed, tested, and validated the keratoconus end-points assessment questionnaire (KEPAQ). Through extensive study, our group has demonstrated the KEPAQ to be a robust, well-constructed instrument^[5] with excellent unidimensionality and reliability^[6] and good clinical correlation.^[7] Nevertheless, no study so far has demonstrated whether the KEPAQ demonstrates good test-retest reliability. Results from a PROM designed for clinical use must remain stable when the clinical state of the patient remains stable, and should only change when clinical state does.^[8] This is a critical characteristic for a well-constructed instrument to have, as it demonstrates that results obtained in a point in time represent an actual value, instead of a random score generated by mere chance. This is part of the test internal consistency and ensures that data obtained actually represents a given clinical state from a patient.

Therefore, our group decided to evaluate the test-retest reliability of the KEPAQ, in an effort to contribute evidence on its internal consistency and capability of measuring clinical state with minimal inference of random chance.

Methods

This is a prospective analytical study designed to evaluate the test-retest reliability of the KEPAQ through the repeated application of the questionnaire to a group of clinically stable individuals.

A number of patients with a confirmed diagnosis of clinically and topographically stable keratoconus underwent application of the full version of the KEPAQ twice, seven days apart. Seven days was considered as an arbitrary but adequate period of time between repeated applications, as it is long enough for patients not to remember exactly the answers they gave the first time but short enough for their clinical state to remain stable. KEPAQ score was obtained through Rasch analysis, while test-retest reliability was evaluated through Spearman rank-order correlation and especially through an intraclass correlation coefficient as has been suggested previously.^[9]

Sample size and study population

A convenience sample of 100 subjects was included as was considered to be over the minimum sample size for this kind of study. The population studied for this paper consisted on patients with a confirmed diagnosis of keratoconus, whose main assigned cornea specialist was the main author at the Clínica de Oftalmología Sandiego (Medellín, Colombia). Inclusion criteria included a confirmed diagnosis of keratoconus from both clinical and tomographic standpoints, age over 15, and a desire to participate in the study as evidenced by the informed consent. Exclusion criteria included the presence of any other ophthalmological disease such as glaucoma or retinal compromise and the presence of a cognitive impairment of any level or origin.

KEPAQ scale

The KEPAQ is a keratoconus-specific scale, recently developed and validated by Balparda *et al.*^[6] It consists of a total of 16

questions divided into two subscales that measure different constructs. The first part of the scale consists of 7 questions and evaluates the *Emotional* compromise of the patients secondary to the disease [KEPAQ-E, Table 1]. The second subscale consists of 9 questions revolving around the *Functional* compromise secondary to ectasia [KEPAQ-F, Table 2]. All questions are written in a clear and concise manner, and ask the patient about how much they feel the disease handicaps them in a number of different situations. All questions use a Likert-Like response system with a corresponding scoring system as follows: "Not at all" = 3; "A little" = 2; "Quite a Bit" = 1; "A Lot" = 0. All patients are also given the possibility of marking "Not Applicable" if they feel the question does not correlate with any situation in their lives. Then, the sum score is converted to a *Rasch-derived score* using two tables developed by our group and the subject is given a total of two scores, one for the KEPAQ-E [Table 3] and one for the KEPAQ-F [Table 4], with a greater score meaning less disability caused by disease. When clinicians have a significant number of patients and want to get scores for them, they can also perform a Rasch analysis themselves to get an exact score for their given sample.

Rasch analysis

Rasch analysis has recently caused a great change in the way PROM scales are constructed, validated, and scored. It allows for a much better evaluation of the different psychometric properties of instruments, and many studies consider it to be well superior to classical test theory.

A Rasch analysis was performed for both subscales in order to obtain an *interval-level* kind of score expressed in an arbitrary unit called "Logits" as has already been published by our group.^[5] This converted score has been demonstrated to be much superior when compared with the mere sum score as is suggested in classical test theory. Important elements such as *Person Separation Index*, *Item Infit*, and *Item Outfit* were evaluated to make sure results complied with Rasch analysis expectations of a well-constructed instrument.^[5] Rasch analysis was performed in JMetrik version 4.1.1 (Psychomeasurement Systems LLC; Charlottesville, VA, USA) in a MacBook Air computer running macOS Catalina version 10.15.2 (Apple Inc.; Cupertino, CA, USA).

Test-Retest reliability

Test-retest reliability was calculated by comparing the mean Rasch score of every participant during every call for each of the two KEPAQ subscales separately. It was assessed in two ways. First, a Spearman rank-order correlation between both calls was calculated by obtaining both an *R* and a *P* value. The reason for selecting a Spearman rank order instead of a Pearson correlation was the non-normal nature of most results, as will be explained in the next section.

Second, ICC estimates and their 95% confidence intervals were calculated based on a mean rating ($k = 2$), absolute-agreement, 2-way mixed-effects model. This kind of model was selected based on Koo and Li's^[9] specific recommendation regarding test-retest studies. Test-retest reliability evaluation was performed through IBM SPSS Statistics version 23 (International Business Machines Corporation; Armonk, NY, USA) in a MacBook Air computer running macOS Catalina version 10.15.2 (Apple Inc.; Cupertino, CA, USA).

Table 1: Emotional compromise subscale of the keratoconus end-points assessment questionnaire (KEPAQ-E)

	Not at All	A Little	Quite a Bit	A Lot	N/A
1. Do you feel your eye disease has affected your confidence to perform your daily tasks?	3	2	1	0	X
2. Do you feel your eye disease has affected your confidence to leave your house alone?	3	2	1	0	X
3. Do you feel your eye disease has affected your happiness in general?	3	2	1	0	X
4. Do you feel your eye disease has affected your confidence to go from one place to another?	3	2	1	0	X
5. Do you feel your eye disease has affected your self-esteem?	3	2	1	0	X
6. Do you feel your eye disease has affected your confidence about the future?	3	2	1	0	X
7. Do you feel your eye disease has caused you to fear about the future?	3	2	1	0	X

Table 2: Functional compromise subscale of the keratoconus end-points assessment questionnaire (KEPAQ-F) + (KEPAQ-F)

	Not at All	A Little	Quite a Bit	A Lot	N/A
1. Do you feel your eye disease has affected your ability to play sports?	3	2	1	0	X
2. Do you feel your eye disease has affected your ability to see objects at near?	3	2	1	0	X
3. Do you feel your eye disease has affected your ability to perform your daily tasks?	3	2	1	0	X
4. Do you feel your eye disease has affected your ability to watch a movie?	3	2	1	0	X
5. Do you feel your eye disease has affected your ability to do your job?	3	2	1	0	X
6. Do you feel your eye disease has affected your ability to watch television?	3	2	1	0	X
7. Do you feel your eye disease has affected your ability to use the computer?	3	2	1	0	X
8. Do you feel your eye disease has affected your ability to read books?	3	2	1	0	X
9. Do you feel your eye disease has affected your ability to see objects that are faraway?	3	2	1	0	X

Table 3: Transforming KEPAQ-E raw score to Person Measure, which is the value that should be used for epidemiological and clinical applications according to Rasch analysis theory

Raw score	Person measure equivalent	Standard error
0	-5.47	1.89
1	-4.12	1.10
2	-3.22	0.83
3	-2.64	0.71
4	-2.20	0.63
5	-1.83	0.58
6	-1.51	0.55
7	-1.21	0.54
8	-0.93	0.53
9	-0.65	0.53
10	-0.37	0.53
11	-0.09	0.54
12	0.21	0.56
13	0.53	0.58
14	0.88	0.60
15	1.26	0.64
16	1.70	0.69
17	2.22	0.76
18	2.87	0.86
19	3.72	0.99
20	4.89	1.21
21	6.40	1.93

Compliance with ethical standards

There are no potential conflicts of interests related to this article. This research adhered to the tenets of the Helsinki's declaration and proper ethical approval was obtained at the Comité de Ética en Investigación – Clínica de Oftalmología Sandiego. All patients provided a telephone-based informed consent accepting their participation. When the subject was underage, one of the parents gave informed consent. No external funding was received for this research.

Results

A total of 100 patients with a confirmed diagnosis of keratoconus were included. Mean age of the patients was 34.90 ± 11.52 years (minimum 11–Maximum 64 years). Mean age at the diagnosis of ectasia was 25.00 ± 10.23 years (minimum 8–Maximum 62 years). 49 (49.00%) of the cohort were female.

Upon questioning about their current refractive status, 45 (45.00%) patients used only glasses, while 21 (21.00%) referred they did not regularly use any kind of refractive aid. The rest of the patients used only contact lenses or a combination of contact lenses and glasses.

Regarding prior keratoconus surgery, 28 (28.00%) patients had a history of keratoplasty in at least one of their eyes (either penetrating or deep anterior lamellar techniques). Fifty two (52.00%) and 31 (31.00%) patients had a history of corneal ring implantation and corneal collagen cross-linking, respectively.

Emotional compromise (KEPAQ-E)

All of the patients answered the KEPAQ-E in two occasions seven days apart, and none of them referred to have any

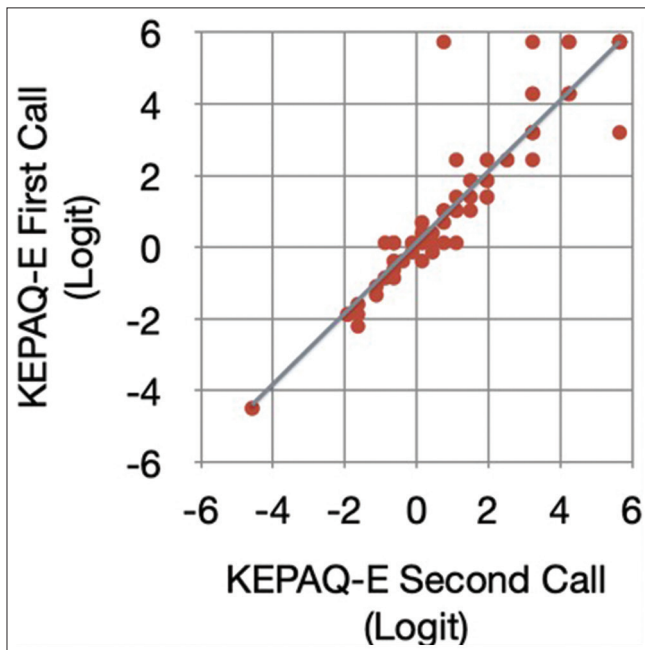


Figure 1: Spearman rank-order correlation between the score of the first and the second call for the emotional compromise subscale of the KEPAQ (KEPAQ-E)

problem in understanding or answering the questions. Regarding the first call, the mean score for the KEPAQ-E according to Rasch scoring was 2.19 ± 2.66 Logit (first quartile 0.14 Logit; median 1.95 Logit; third quartile 4.23 Logit). Skewness of the score was -0.21 (standard error 0.24) while Kurtosis was -0.83 (standard error 0.47). According to Kolmogorov-Smirnoff test, results were non-normal ($P < 0.001$). Regarding the second call, the mean score for the KEPAQ-E according to Rasch scoring was 2.28 ± 2.73 Logit (first quartile 0.12 Logit; median 2.43 Logit; third quartile 5.73 Logit). Skewness of the score was -0.20 (standard error 0.24) while Kurtosis was -0.90 (standard error 0.48). According to Kolmogorov-Smirnoff test, results were non-normal ($P < 0.001$).

Upon comparing Rasch score for both calls, Spearman Rho Score was $R = 0.921$ ($P < 0.001$) [Fig. 1]. Average measures ICC was 0.952 (95% confidence interval 0.929–0.968).

Functional compromise (KEPAQ-F)

All of the patients answered the KEPAQ-F in two occasions seven days apart, and none of them referred to have any problem understanding or answering the questions. Regarding the first call, the mean score for the KEPAQ-F according to Rasch scoring was 1.20 ± 2.05 Logit (first quartile -0.03 Logit; median 0.91 Logit; third quartile 2.24 Logit). Skewness of the score was 0.05 (standard error 0.24) while Kurtosis was 0.13 (standard error 0.47). According to Kolmogorov-Smirnoff test, results were non-normal ($P = 0.005$). Regarding the second call, the mean score for the KEPAQ-E according to Rasch scoring was 1.22 ± 2.28 Logit (first quartile -0.01 Logit; median 0.92 Logit; third quartile 2.47 Logit). Skewness of the score was -0.04 (standard error 0.24) while Kurtosis was 0.28 (standard error 0.47). According to Kolmogorov-Smirnoff test, results were normal ($P = 0.052$).

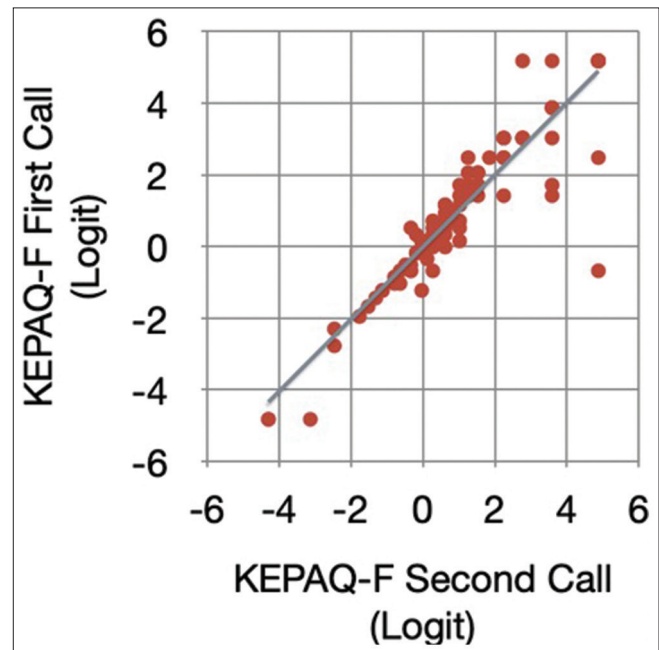


Figure 2: Spearman rank-order correlation between the score of the first and the second call for the functional compromise subscale of the KEPAQ (KEPAQ-F)

Upon comparing Rasch score for both calls, Spearman Rho Score was $R = 0.921$ ($P < 0.001$) [Fig. 2]. Average measures ICC was 0.952 (95% confidence interval 0.929–0.968).

Discussion

Scale construction and validation is a long and demanding process, in which the final objective is to build a set of questions that adequately measure a latent trait of interest (called a *construct*) and develop a way of measuring a final score that can be both logical and sensitive for statistical analysis. Aside from the actual development process (such as Rasch analysis to eliminate misfitting or redundant questions and principal component analysis to determine unidimensionality)(5, 6), post-development studies are of utmost importance in order to determine that scale results are reliable, which means they actually measure what they are designed to measure in the first place. Test-retest reliability reflects the variation in measurements taken by an instrument on the same subject under the same conditions.^[9] It helps determine that results obtained by the scale at any point in time actually correspond to a clinical state of the patient, instead of being caused by mere chance. If test-retest reliability is poor (meaning measuring the same subject under the same conditions generates far too different scores) then it would mean that the instrument is poorly designed and is not measuring an actual construct but is being subjected to random noise.

This study was designed precisely to demonstrate whether the KEPAQ was capable of producing comparable results when a patient was measured twice while keeping their clinical conditions stable. In order to achieve this, each patient received a total of two calls, seven days apart, and answered a complete version of the KEPAQ in every call. Seven days was arbitrarily designated as a period long enough for the patient not to be able to exactly remember their original answers, whereas at

Table 4: Transforming KEPAQ-F raw score to Person Measure, which is the value that should be used for epidemiological and clinical applications according to Rasch analysis theory

Raw score	Person measure equivalent	Standard error
0	-5.43	1.84
1	-4.19	1.03
2	-3.45	0.75
3	-2.98	0.63
4	-2.63	0.56
5	-2.33	0.52
6	-2.07	0.50
7	-1.84	0.48
8	-1.61	0.47
9	-1.40	0.46
10	-1.18	0.46
11	-0.97	0.47
12	-0.75	0.47
13	-0.52	0.48
14	-0.28	0.50
15	-0.02	0.51
16	0.24	0.53
17	0.54	0.56
18	0.86	0.58
19	1.21	0.60
20	1.59	0.63
21	2.00	0.66
22	2.47	0.70
23	3.00	0.76
24	3.63	0.83
25	4.41	0.94
26	5.49	1.18
27	6.97	1.93

the same time being short enough to ensure their visual and clinical state remained unchanged. If the KEPAQ was found to be well constructed, results between both calls should be very similar as demonstrated by statistical analysis.

A number of different epidemiological and statistical approaches have been proposed to evaluate test-retest reliability in clinical scenarios. Some studies have used a paired *Student-t* or a Bland-Altman plot to evaluate reliability. Nevertheless, these tests were developed to analyze only agreement, and not correlation, and hence they are nonideal measures of reliability.^[9] Therefore, Spearman/Pearson correlation and, specially ICC, have been proposed as better measures for reliability. ICC is especially useful as it reflects both degree of correlation and agreement between measurements.^[9]

Our study starts with obtaining an exact score for the KEPAQ in every one of the two calls through Rasch analysis. Then, the correlation between the two calls was initially evaluated through Spearman rank-order correlation. The reason for selecting Spearman instead of Pearson was the non-normal distribution of the KEPAQ score, as designated

by Kolmogorov-Smirnoff test. For Spearman rank-order correlation, an *R* value over 0.90 demonstrates an excellent correlation, and the results obtained in our study are well over this value for both subscales. This demonstrates that a greater KEPAQ score in the first call was very predictive of a greater score in the second call, a very suggestive characteristic of adequate test-retest reliability. Nevertheless, this test alone does not evaluate agreement between the two scores. To solve this, we have also used ICC based on a mean rating ($k = 2$), absolute-agreement, 2-way mixed-effects model, according to Koo and Li's^[9] recommendations. This ensures an adequate evaluation of both correlation and agreement between the score of the two calls. Results from our study demonstrate that mean ICC value is well over 0.90, suggesting excellent reliability. This is further confirmed by looking at the 95% confidence interval, in which the lower limit is also over 0.90 so even in the most pessimistic scenario, test-retest reliability is still excellent.

These results support the notion that using the KEPAQ to evaluate the quality of life in keratoconus patients provides well-structured reliable results, which adequately measures both the *Emotional* (KEPAQ-E) and *Functional* (KEPAQ-F) constructs. This should provide the clinician with enough confidence to warrant the use of KEPAQ for both clinical and research endeavors.

Conclusion

The KEPAQ is a robust well-developed scale designed to measure both emotional and functional handicap due to keratoconus. It has been previously shown to be well-fitting,^[5] unidimensional,^[6] and to correlate with clinical variables.^[7] Results from this study also demonstrate it to have excellent test-retest reliability, showing that the scale measures what it is supposed to measure, and is not subject to random noise which may impair its use. These results suggest that the KEPAQ is an excellent scale and can be confidently used for both clinical and research use.

Acknowledgements

The authors are in debt with the patients who participated in this study. This article is dedicated to the memory of Captain Kepa Balparda Sr.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Khadka J, Schoneveld PG, Pesudovs K. Development of a keratoconus-specific questionnaire using Rasch Analysis. *Optom Vis Sci* 2017;94:395-403.
2. Kandel H, Pesudovs K, Ferdi A, Mills R, Chen JY, Watson A, *et al*. Psychometric properties of the Keratoconus outcomes research questionnaire: A save sight Keratoconus registry study. *Cornea* 2020;39:303-10.
3. Balparda K, Galarreta-Mira D, Vanegas-Ramírez CM. Translation and validation of the "Questionnaire for research on keratoconus results" (KORQ) in the Colombian population. *Archivos de la Sociedad Espanola de Oftalmologia* 2020;95:579-85.
4. Kandel H, Pesudovs K, Watson SL. Measurement of quality of life in keratoconus. *Cornea* 2020;39:386-93.

5. Balparda K, Herrera-Chalarca T, Silva-Quintero LA, Torres-Soto SA, Vanegas-Ramirez CM. Development and validation of the "Keratoconus End-Points Assessment Questionnaire" (KEPAQ), a disease-specific instrument for evaluating subjective emotional distress and visual function through rasch analysis. *Clin Ophthalmol* 2020;14:1287-96.
6. Balparda K, Herrera-Chalarca T, Torres-Soto SA, Silva-Quintero LA. Both sub-scales of the "Keratoconus End-Points Assessment Questionnaire" (KEPAQ) are unidimensional and reliable. *Graefes Arch Clin Exp Ophthalmol* 2020;258:2233-9.
7. Balparda K, Herrera-Chalarca T, Silva-Quintero LA, Torres-Soto SA, Segura-Munoz L, Vanegas-Ramirez CM. Both subjective emotional distress and visual handicap correlate with Belin ABCD classification in the worse eye as measured with the "Keratoconus End-Points Assessment Questionnaire" (KEPAQ). *Clin Ophthalmol* 2020;14:1839-45.
8. Paiva CE, Barroso EM, Carneseca EC, de Padua Souza C, Dos Santos FT, Mendoza Lopez RV, *et al.* A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: A systematic review. *BMC Med Res Methodol* 2014;14:8.
9. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-63.