



Archived at the Flinders Academic Commons:

<http://dspace.flinders.edu.au/dspace/>

'This is the peer reviewed version of the following article:

Colm McAlinden, Jyoti Khadka, Juliane de Freitas Santos Paranhos, Paulo Schor, Konrad Pesudovs; Psychometric Properties of the NEI-RQL-42 Questionnaire in Keratoconus. *Invest. Ophthalmol. Vis. Sci.* 2012;53(11):7370-7374. doi: 10.1167/iovs.12-9969.

which has been published in final form at

<http://dx.doi.org/10.1167/iovs.12-9969>

This article is made available with the permission of the publisher, Association for Research in Vision and Ophthalmology

Copyright © Association for Research in Vision and Ophthalmology

Psychometric Properties of the NEI-RQL-42 Questionnaire in Keratoconus

Colm McAlinden,¹ Jyoti Khadka,¹ Juliane de Freitas Santos Paranhos,² Paulo Schor,² and Konrad Pesudovs¹

PURPOSE. To assess the psychometric properties of the National Eye Institute Refractive Error Quality of Life (NEI-RQL-42) questionnaire in keratoconus and compare these findings to patients with refractive error correction alone.

METHODS. The Portuguese version of the NEI-RQL-42 Quality of Life questionnaire was completed by 44 patients who had keratoconus before and after implantation of intracorneal ring segments. Rasch analysis was used to assess the use of response categories, success in measuring a single trait per subscale (unidimensionality), ability to discriminate person ability (precision), and targeting of questions to person quality of life (QoL).

RESULTS. Rasch analysis was performed for the questionnaire subscales using stacked preoperative and postoperative data. Three subscales (Symptoms, Dependence on correction, and Suboptimal correction) contained response categories that were not used as intended. Six subscales contained misfitting items indicating multidimensionality. Eleven subscales exhibited inadequate measurement precision. Only the Near vision subscale demonstrated adequate precision with a person separation greater than 2.0. Targeting of items to person QoL was adequate in 11 of the 12 subscales with a mean item location of less than 1 logit.

CONCLUSIONS. Only one NEI-RQL-42 subscale (Near vision) performed adequately in keratoconus. Targeting was better in patients with keratoconus than in patients with refractive error correction alone, but 11 of the 12 subscales remain manifestly inadequate. Better instruments exist for measuring patient-reported outcomes in keratoconus. (*Invest Ophthalmol Vis Sci.* 2012;53:7370–7374) DOI:10.1167/iovs.12-9969

Quality of life (QoL) is a well-established trait measured in clinical and research settings across medicine, including ophthalmology.^{1,2} Indeed, patient-reported outcomes have become an essential component required by funding bodies, ethics committees, third-party papers, and regulatory agencies.³ It is therefore imperative that any patient-reported

outcome, including QoL questionnaires, be well developed, of high quality, and meet standard psychometric properties.

The National Eye Institute Refractive Error Quality of Life (NEI-RQL-42) questionnaire was developed at RAND Corporation under the sponsorship of the National Eye Institute and is a commonly used questionnaire in the ophthalmic community to assess QoL related to refractive error correction.⁴ The questionnaire consists of 42 items (questions) combined to provide 13 subscale scores. Previous studies have evaluated the NEI-RQL-42 questionnaire using more traditional validation techniques.^{5–7} These techniques fail to assess key psychometric properties such as response category ordering, dimensionality, person separation, and targeting of questions to persons. The NEI-RQL-42 also uses Likert scaling with a simplistic summary scoring method that is prone to bias and makes assumptions on item difficulty and unequal response option spacing.⁸ Rasch analysis overcomes these disadvantages and provides a measure that is linear with the latent variable, with measurements conforming to a Guttman scale.^{9,10} Rasch analysis is recognized as the gold standard for questionnaire methodology in ophthalmology. We recently investigated the psychometric properties of the NEI-RQL-42 in spectacle and contact lens wearers undergoing laser refractive surgery using Rasch analysis.¹¹ Rasch analysis is a simple yet effective psychometric model with two main characteristics.⁹ Firstly, it estimates interval-level measurement on a continuous scale from raw questionnaire data (ordinal responses). This estimation reduces measurement noise and enables valid parametric statistical analysis of the output data. Secondly, Rasch analysis delivers unparalleled insight into the psychometric properties of questionnaires, including the appropriateness of the response options, precision of the measurement, the fit of questions to the construct under investigation, and targeting of the questions to the respondents.¹² This has led to its growing use in the development of questionnaires in ophthalmology.^{13–16}

Our study revealed serious psychometric deficiencies with the NEI-RQL-42 QoL questionnaire, including response categories not performing as expected, questions that did not fit with the trait represented by a subscale, and poor precision. Precision problems occurred because the subscales possessed too few items (e.g., two items); some of these items were irrelevant to people with refractive error correction. For example, the questionnaire contains many questions related to activity limitations, yet very few people had such problems. We hypothesized that the NEI-RQL-42 questionnaire may perform better in disease groups where activity limitations may be more relevant.¹¹

To test this hypothesis, we draw upon data from a recent study investigating QoL in people with keratoconus before and after implantation of intracorneal ring segments.¹⁷ This study was a collaborative project between a research team based in Australia and Brazil. The Brazilian group were conducting a larger project on keratoconus surgery outcomes, and the present study was an additional arm of the main project. The

From the ¹National Health and Medical Research Council Centre for Clinical Eye Research, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, Australia; and the ²Federal University of São Paulo, São Paulo, Brazil.

Submitted for publication April 4, 2012; revised August 13 and September 13, 2012; accepted September 14, 2012.

Disclosure: C. McAlinden, None; J. Khadka, None; J. de Freitas Santos Paranhos, None; P. Schor, None; K. Pesudovs, None

Corresponding author: Colm McAlinden, NH&MRC Centre for Clinical Eye Research, Flinders Medical Centre and Flinders University of South Australia, Bedford Park, South Australia, 5042, Australia; colm.mcalinden@gmail.com.

patients in Brazil spoke Portuguese; thus a Portuguese version of the NEI-RQL-42 questionnaire was used. The purpose of this study was to assess the psychometric properties of the NEI-RQL-42 QoL questionnaire in keratoconus, to determine if it functions better in patients with keratoconus than in those with refractive error correction alone.

METHODS

The National Eye Institute Refractive Error Quality of Life Questionnaire (NEI-RQL-42)

The NEI-RQL-42 contains 42 items across 13 subscales (Clarity of vision, Expectations, Near vision, Far vision, Diurnal fluctuation, Activity limitations, Glare, Symptoms, Dependence on correction, Worry, Suboptimal correction, Appearance, Satisfaction with correction). The 42 items of the NEI-RQL-42 are designed with 16 different question/response category formats. Question format is defined as the way in which the questions are asked: "How much difficulty . . . ?" or "Do you suffer . . . ?" The response scales indicate the number of different category response options such as one to six or one to three response options. The question/response category formats are not unique within each of the 13 subscales of the questionnaire, with some subscales incorporating multiple formats. Scoring of the questionnaire involves two steps. Firstly, the original numeric values are recoded following a set of scoring rules across a 0% to 100% range, with higher scores indicating better QoL. Secondly, subscales are scored by averaging together the items within each subscale. The number of items in each subscale varies from one to seven.

Subjects

The translation and validation of the questionnaire have been previously published.¹⁸ In brief, the translation involved a forward and back translation with native ophthalmologists and professional translators. The back-translated version was compared to the original English version, and the questionnaire underwent pilot testing in 20 patients undergoing refractive surgery at the Federal University of São Paulo. The final Portuguese version of the NEI-RQL-42 was administered to 102 patients who underwent refractive surgery (before and after surgery), with 20 patients answering the questionnaire at two different postoperative visits.

Patients with keratoconus with no ocular comorbidities who were undergoing intracorneal ring segments at the refractive sector of the Ophthalmologic Department of Federal University of São Paulo (UNIFESP) were invited to take part in this study. The questionnaire was self-administered by 44 patients. The same 44 patients completed the questionnaire following implantation of intracorneal ring segments. Twenty-nine patients had bilateral implantation, and 15 had monocular implantation. Full details for the surgery have been previously reported.¹⁷ Postoperatively, the questionnaire was completed when the patient had been wearing his or her best correction for at least 40 days, corresponding to between 4.5 and 8 months postoperatively. All patients were 18 years or older, were speakers of Portuguese, and had no severe cognitive impairment. The study was approved by the ethics committee of the Federal University of São Paulo (Rua Botucatu, 572 1º andar conj 14, CEP 04023-062, São Paulo, Brazil). All patients read and signed a consent form, and research was conducted in accordance with the Declaration of Helsinki.

Rasch Analysis

Twelve of the 13 subscales underwent a separate Rasch analysis. The Satisfaction with correction subscale has only one item so cannot undergo Rasch analysis. Items that contained the response options "Never do these activities for other reasons/Don't drive at night for other reasons/Don't drive at dusk for other reasons" were treated as

missing data. The response polarity was matched according to the recoding scores in the NEI-RQL-42 manual. Items 36 to 42 have an "a" part and a "b" part. Part "a" asks for a yes/no response, and part "b" is completed only if the response to part "a" is yes. We considered this a five-response scale.

The Rasch model is based on a probabilistic relationship between item difficulty and person ability, with the difference known as the functional reserve or functional ability. The functional reserve expresses the probability of any person being successful on any item. In the dichotomous model, the probability of success is expressed as a function of the size of the difference between the ability (B) of the person (n) and the difficulty (D) of the item (i). Raw scores are converted into odds of success, the ratio of person percentage success (p) to person percentage failure ($1 - p$). The natural log of this ratio is the difference between the person ability estimate (Bn) and the item difficulty estimates (Di). Therefore, both person ability (Bn) and item difficulty (Di) are expressed on a logit scale with an average logit of zero. Positive logits indicate higher than average probabilities of person success on an item, and negative logits indicate a lower than average probability of success. A polytomous Andrich rating scale model using joint maximum likelihood estimation was applied to each question format using Winsteps (version 3.70.0.2; Winsteps, Chicago, IL).¹⁹ The Wolfe and Chiu stacking procedure for pre- and postoperative evaluation was performed.²⁰ This involves combining pre- and postoperative data in a single analysis. Rasch analysis was used to assess each subscale for response category performance, dimensionality, precision, and targeting.

Category Threshold Order

Response category performance in terms of being used in the order intended was evaluated by observing if the category calibration increased in an orderly fashion in the probability curves. The threshold is the midpoint between response categories and indicates the point where the likelihood of choosing either response category is the same. Items in the questionnaire have between two and six choices, which translates to one to five thresholds, respectively. Each threshold has a location on the logit scale, and each item has a mean location. Hence, one would expect that with decreasing ability, the probability of selecting each category would increase in an ordered fashion from least to most difficult. Disordered thresholds may occur in the event of an underused category or unclear descriptive wording, or if the number of categories exceeds the number of levels the participants can distinguish.²¹ Disordered thresholds can be a source of item misfit. Therefore, in cases of disordered thresholds, response categories were collapsed (adjacent categories combined together) until thresholds were ordered; this was performed prior to any further analysis. Category probability curves were inspected to identify which adjacent categories could be combined in the presence of disordered thresholds. Categories that showed the greatest overlap of curves are usually the most appropriate to combine; however, where an underutilized category could be combined with two adjacent categories, both were performed in turn, with the impact on the fit to the model assessed. The combination that provided the largest improvement in fit was accepted. Categories were considered for combining only when labels made it logical to do so. Ideally, categories should be evenly spaced and should advance step calibrations by at least 1.4 logits.²¹

Dimensionality

Item fit statistics are used in the assessment of unidimensionality, which demonstrates whether the questionnaire or subscale is measuring a single concept. Fit statistics (infit and outfit) focus on two aspects, which can be reported as a mean square (MNSQ) with expected values of 1. The MNSQ residual statistic is normalized to the average expected variance such that a residual of less than 0.70 indicates at least 30% less variance than expected, suggesting a high

TABLE. Overall Performance of the NEI-RQL-42 Quality of Life (QoL) Questionnaire in Keratoconus using Rasch Analysis

Subscales	No. of Items	Thresholds Needing Reordering	No. of Misfitting Items	Misfitting Items	Person Separation	Mean Item Location
Clarity of vision	4	None	1	40	1.44	0.06
Expectations	2	None	None	NA	0.00	-0.43
Near vision	4	None	None	NA	2.81	0.03
Far vision	5	None	3	6, 9, 10	0.87	0.47
Diurnal fluctuation	2	None	None	NA	1.67	-0.44
Activity limitations	4	None	2	33, 34	0.66	0.16
Glare	2	None	None	NA	1.53	-0.05
Symptoms	7	1 (item 25)	3	25, 41, 42	1.91	0.76
Dependence on correction	4	2 (items 15, 16)	1	13	0.00	-0.58
Worry	2	None	None	NA	1.01	-1.50
Suboptimal correction	2	2 (items 31, 32)	None	NA	0.00	0.65
Appearance	3	None	2	27, 29	NA	NA
Satisfaction with correction	1	NA	NA	NA	NA	NA

level of predictability or possible redundancy. Residuals greater than 1.30 indicate at least 30% more variance than expected, suggesting that items may be measuring something different from the overall scale. Therefore, an acceptable infit and outfit is within the range of 0.70 to 1.30.²² Misfitting items were removed from subscales until all items demonstrated unidimensionality, if possible, prior to further analysis.

Precision

Precision refers to whether a measure is able to discriminate along its scale, and person separation statistics indicate the overall precision of the instrument. The ratio of the true variance in the estimated measures to the observed variance can be used to determine how many groups or strata of person ability an instrument can discriminate. Greater precision is indicated with a greater value for the person separation statistics. Instrument precision enables a greater distinction between levels of function. A minimal acceptable cutoff value for the person separation ratio was set at 2.0 for this study.²³

Targeting

Targeting refers to the extent to which the difficulty of the items matches the abilities of the persons in the sample, and is assessed numerically by comparison of the person and item mean values. A perfect targeting instrument would have a difference of zero, whereas a difference of more than 1 logit indicates significant mistargeting.²⁴

RESULTS

The questionnaire was self-administered by 44 patients (mean age 24.1 ± 5.2 years; age range 18–39 years; 25 female, 19 male). A separate Rasch analysis of stacked preoperative and postoperative data was performed for each of the 12 subscales. Category response thresholds were assessed first, before further analysis. Three response category formats had disordered thresholds requiring response categories to be reordered. The three affected subscales were Symptoms, Dependence on correction, and Suboptimal correction (Table). The Figure illustrates the category probability curves for the Symptoms subscale. It can be seen from the Figure that category 3 was underused; at no point was it more likely to be chosen over any other category.

Rasch analysis was repeated with the newly reorganized response formats, and the subscales were assessed for unidimensionality with item fit statistics. Six of the 12 subscales contained misfitting items (Table). Misfitting items were removed prior to further analysis. There were two misfitting items in the Appearance subscale, and once these

were removed, only one item remained. Therefore, further analysis was not possible for this subscale.

Precision was assessed by person separation, with values greater than 2.0 deemed adequate (Table). Only the Near vision subscale demonstrated adequate person separation (2.81). All remaining subscales had a person separation below acceptable standards.

Targeting of person ability to item difficulty was adequate in 11 of the 12 subscales, with a mean item location of less than 1 logit. The Worry subscale had poor targeting, with a mean item location of -1.50 logits (items too difficult).

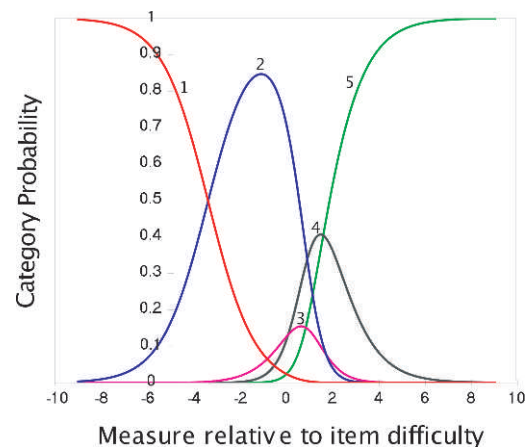


FIGURE. Response category probability curve for the Symptoms subscale. The *x-axis* represents the difference between item and person calibration, and the *y-axis* represents the probability of the category's being chosen. It can be seen that category 3 was underused; at no point was it more likely to be chosen over any other category.

DISCUSSION

Keratoconus is a noninflammatory, progressive, ectatic disease of the cornea. It has well-described clinical signs including stromal thinning, protrusion, irregular astigmatism, and scarring. It can lead to significant visual impairment and decreased QoL.²⁵ There are a range of treatment and management modalities available for keratoconus.²⁶ QoL is an important outcome measure in the assessment of the various treatment options; however, QoL measures have rarely been applied. The NEI-RQL-42 questionnaire was developed to measure refractive error-related QoL; however, the questionnaire's psychometric

properties for use with patients with refractive error correction have been found to be grossly inadequate.¹¹ It was hypothesized that the questionnaire may function better in disease groups where the latent trait of activity limitation may be more important. Hence this study investigated the psychometric properties of the NEI-RQL-42 QoL questionnaire in a keratoconus population.

The results revealed that the questionnaire was deficient in the assessment of QoL in keratoconus, with only one valid subscale of the 12 assessed. This was the four-item Near vision subscale, with an adequate response category, non-misfitting items, adequate person separation, and adequate targeting. The remaining 11 subscales were deficient in one or more ways. Three subscales (Symptoms, Dependence on correction, and Suboptimal correction) had disordered thresholds that required reordering of response categories (e.g., Fig.). This commonly occurs because there are too many response options, as it has been shown that respondents can generally distinguish only four or five response options.²¹ The number of response options for items in the questionnaire varied between 1-2, 1-3, 1-4, 1-5, and 1-6 across the subscales. Subscales with disordered response categories needed to be repaired prior to further analysis; hence all disordered subscales underwent category collapsing to eliminate disordering.

Unidimensionality was assessed with item fit statistics, which indicated that six of the 12 subscales had misfitting items. This indicates that these six subscales were not assessing a single latent trait but rather multiple traits, with the questions within each subscale not measuring what they purport to measure. This is an important issue in measurement that also occurs with clinical measurement. For example, a device that measures intraocular pressure (IOP) and central corneal thickness (CCT) might be very useful if it produces two unidimensional scores (one for IOP and one for CCT). However, if it produces a single multidimensional score, this might have something to do with glaucoma, but it has no clinical utility. The six multidimensional subscales of the NEI-RQL-42 QoL questionnaire have this exact problem. In the previous study investigating the psychometric properties of this questionnaire in refractive error correction, six subscales were also found to have misfitting items, although only three subscales were the same as in the present investigation: Activity limitations, Symptoms, and Dependence on correction.¹¹ The reasons for item misfit are probably related to the design of the questionnaire, particularly the appropriateness of the questions for the subscale and the format of the response options. Misfitting items were removed from the appropriate subscales and Rasch analysis was repeated.

Person separation was found to be inadequate for all except one of the 12 subscales: Near vision. The remaining subscales had person separation values significantly less than the minimum accepted value of 2.0.²³ This indicates that these subscales could not adequately discriminate between the individuals in the sample population. Targeting was adequate for 11 of the subscales. Targeting refers to the matching of the mean person ability estimates and mean item difficulty estimates. This finding suggests that these items are important to the keratoconus disease group. In comparison, targeting in the refractive correction group was poor, as many subjects had no such problems (items were too easy for the patients' ability level).

While the Near vision subscale could be recommended for measuring visual functioning in keratoconus, there are many other visual functioning instruments that could be considered for use.¹⁶ Many researchers who are unfamiliar with the properties of a good-quality questionnaire choose the NEI-RQL-42, as they assume it would have been developed appropriately under the sponsorship of the National Eye Institute and the

subsequent naming of the questionnaire. A number of good visual functioning questionnaires have been revalidated with Rasch analysis for use in a cataract population which have better psychometric properties.²⁷⁻²⁹ There is also the Quality of Vision Questionnaire, which was developed using Rasch analysis for refractive surgery and cataract surgery.^{15,30} It is likely that these instruments will measure visual functioning adequately in keratoconus also. We have previously attempted to repair the NEI-RQL-42 questionnaire in the refractive surgery group without success,¹¹ and it would not be advantageous to attempt repair in this keratoconus data set. Even if it were possible to create some sort of valid measure with adequate psychometric properties, we could not advocate it for keratoconus. The questionnaire was not developed for keratoconus, and it is likely to be content deficient. There are also significant problems with the design of the questionnaire, such as the differences in the wording and the number of response options across the items. Moreover, the Keratoconus Outcomes Research Questionnaire was specifically developed to measure visual functioning and symptoms in keratoconus patients, with good psychometric properties as determined by Rasch analysis, and is a better alternative (Pesudovs K, et al. *IOVS* 2007;48:ARVO E-Abstract 1841).

The main limitations of the present study include possible cross-cultural issues with this analysis, as the questionnaire used was a Portuguese version of the NEI-RQL-42 and the results may be different with the original, English version. The study also contains a small sample size and reports findings similar to those from our previous study in a group of refractive surgery patients.¹¹ However, it has the key benefit of utilizing modern psychometric theory to highlight the shortcomings of this questionnaire.

In conclusion, this study further highlights the deficiencies of the NEI-RQL-42, this time in keratoconus, and demonstrates that the NEI-RQL-42 should not be used in keratoconus studies. Only one subscale was found to be functional in this group, with all other subscales deficient on a number of counts. Better instruments are available, such as the Keratoconus Outcomes Research Questionnaire or the Quality of Vision Questionnaire, depending on which latent trait one wishes to measure.

References

- Higginson IJ, Carr AJ. Measuring quality of life: using quality of life measures in the clinical setting. *BMJ*. 2001;322:1297-1300.
- Lamoureux EL, Maxwell RM, Marella M, Dirani M, Fenwick E, Guymer RH. The longitudinal impact of macular telangiectasia (MacTel) type 2 on vision-related quality of life. *Invest Ophthalmol Vis Sci*. 2011;52:2520-2524.
- Varma R, Richman EA, Ferris FL III, Bressler NM. Use of patient-reported outcomes in medical product development: a report from the 2009 NEI/FDA Clinical Trial Endpoints Symposium. *Invest Ophthalmol Vis Sci*. 2010;51:6095-6103.
- Berry S, Mangione CM, Lindblad AS, McDonnell PJ. Development of the National Eye Institute refractive error correction quality of life questionnaire: focus groups. *Ophthalmology*. 2003;110:2285-2291.
- Nichols JJ, Mitchell GL, Saracino M, Zadnik K. Reliability and validity of refractive error-specific quality-of-life instruments. *Arch Ophthalmol*. 2003;121:1289-1296.
- Hays RD, Mangione CM, Ellwein L, Lindblad AS, Spritzer KL, McDonnell PJ. Psychometric properties of the National Eye Institute-Refractive Error Quality of Life instrument. *Ophthalmology*. 2003;110:2292-2301.
- McDonnell PJ, Mangione C, Lee P, et al. Responsiveness of the National Eye Institute Refractive Error Quality of Life instrument to surgical correction of refractive error. *Ophthalmology*. 2003;110:2302-2309.

8. Townsend JT, Ashby FG. Measurement scales and statistics: the misconception misconceived. *Psychol Bull.* 1984;96:394-401.
9. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danmarks Paedagogiske Institut; 1960.
10. Guttman LA. A basis for scaling quantitative data. *Am Sociol Rev.* 1944;9:139-150.
11. McAlinden C, Skiadaresi E, Moore J, Pesudovs K. Subscale assessment of the NEI-RQL-42 questionnaire with Rasch analysis. *Invest Ophthalmol Vis Sci.* 2011;52:5685-5694.
12. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci.* 2007;84:663-674.
13. Khadka J, Ryan B, Margrain TH, Court H, Woodhouse JM. Development of the 25-item Cardiff Visual Ability Questionnaire for Children (CVAQC). *Br J Ophthalmol.* 2010;94:730-735.
14. Khadka J, Pesudovs K, McAlinden C, Vogel M, Kernt M, Hirneiss C. Reengineering the glaucoma quality of life-15 questionnaire with rasch analysis. *Invest Ophthalmol Vis Sci.* 2011;52:6971-6977.
15. McAlinden C, Pesudovs K, Moore JE. The development of an instrument to measure quality of vision: the Quality of Vision (QoV) questionnaire. *Invest Ophthalmol Vis Sci.* 2010;51:5537-5545.
16. McAlinden C, Gothwal VK, Khadka J, Wright TA, Lamoureux EL, Pesudovs K. A head-to-head comparison of 16 cataract surgery outcome questionnaires. *Ophthalmology.* 2011;118:2374-2381.
17. de Freitas Santos Paranhos J, Avila MP, Paranhos A Jr, Schor P. Evaluation of the impact of intracorneal ring segments implantation on the quality of life of patients with keratoconus using the NEI-RQL (National Eye Institute Refractive Error Quality of life) instrument. *Br J Ophthalmol.* 2010;94:101-105.
18. Nunes LM, Cotizo V, Schor P. Validação de versão em língua portuguesa do questionário NEI-RQL como instrumento de avaliação da qualidade de vida relacionada à visão, em candidatos a cirurgia refrativa. *Rev Bras Oftalmol.* 2004;63:110-118.
19. Andrich D. A rating scale formulation for ordered response categories. *Psychometrika.* 1978;43:561-573.
20. Wolfe EW, Chiu CW. Measuring change across multiple occasions using the Rasch Rating Scale Model. *J Outcome Meas.* 1999;3:360-381.
21. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas.* 2002;3:85-106.
22. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans.* 1994;8:370.
23. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989;70:857-860.
24. Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care.* 2004;42:117-124.
25. Rabinowitz YS. Keratoconus. *Surv Ophthalmol.* 1998;42:297-319.
26. Jhanji V, Sharma N, Vajpayee RB. Management of keratoconus: current scenario. *Br J Ophthalmol.* 2011;95:1044-1050.
27. Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Cataract Symptom Scale: clarifying measurement. *Br J Ophthalmol.* 2009;93:1652-1656.
28. Pesudovs K, Wright TA, Gothwal VK. Visual disability assessment: valid measurement of activity limitation and mobility in cataract patients. *Br J Ophthalmol.* 2010;94:777-781.
29. Gothwal VK, Wright TA, Lamoureux EL, Khadka J, McAlinden C, Pesudovs K. Improvements in visual ability with first-eye, second-eye, and bilateral cataract surgery measured with the visual symptoms and quality of life questionnaire. *J Cataract Refract Surg.* 2011;37:1208-1216.
30. McAlinden C, Skiadaresi E, Pesudovs K, Moore JE. Quality of vision after myopic and hyperopic laser-assisted subepithelial keratectomy. *J Cataract Refract Surg.* 2011;37:1097-1100.