McMonnies Questionnaire: Enhancing Screening for Dry Eye Syndromes with Rasch Analysis

Vijaya K. Gothwal, Konrad Pesudovs, Thomas A. Wright, and Charles W. McMonnies

PURPOSE. To determine, by using Rasch analysis, whether the McMonnies questionnaire possesses the properties of a measure and whether screening for dry eye syndromes (DESs) can be enhanced by using different scoring approaches.

METHODS. The questionnaire was self-administered by 45 female Sjögren syndrome patients (>45 years) recruited from a specialized rheumatology clinic and 140 age-matched control subjects. Data were scaled by using Rasch analysis and were assessed for response category behavior and ability to reliably discriminate between severity of the participant’s dry eye symptoms (i.e., person separation reliability; minimum acceptable value, 0.80). Standard summary statistics of screening performance were calculated for raw and Rasch-scaled scores from receiver-operating characteristic analysis including area under the curve (AUC). Best predictors (i.e., questions) from a discriminant analysis were used to calculate a discriminant function for both Rasch-scaled and raw scores.

RESULTS. Response categories were not used as intended, necessitating a collapse of categories. Person separation reliability was inadequate (0.75). A Rasch-scaled discriminant cutoff score of −2.29 logits from seven items provided an AUC of 0.99 with 95% sensitivity. However, discriminant raw score from modification in the scoring of a question (e.g., use of medications), used as one rather than multiple questions, provided an AUC (0.97) that was not significantly different (z = 1.11, P = 0.27), with 98% sensitivity, and required only two questions.

CONCLUSIONS. In this population, the McMonnies questionnaire does not function as a measure. However, various scoring methods can be used to efficiently screen for DES. (Invest Ophthalmol Vis Sci. 2010;51:1401–1407) DOI:10.1167/iovs.09-4180

The McMonnies questionnaire is among the earliest and most widely used screening instruments for dry eye syndromes (DESs) with sensitivity reportedly varying between 97% and 98% and specificity between 87% and 97% (Golding TR, et al. IOVS 1995;36:ARVO Abstract 823).1–3 Possible reasons for this variation in estimates of sensitivity include the differences in experimental populations and criteria used for disease classification. Another possible reason is the use of various scoring methods for the questionnaire since its development. For example, McMonnies et al.1 used the weighted-scale algorithm based on clinical experience and obtained 92% sensitivity and 93% specificity in a group of 50 women with Sjögren’s syndrome (group with severe dry eye), compared with corresponding values of 98% and 97%, using raw scores derived from discriminant analysis in a subsequent study involving women with or without keratoconjunctivitis sicca.2 However, the results of this study were perhaps affected by selection bias, because efficacy was assessed on the data from the same sample of patients from whom the cutoff values for diagnosis were derived and not from an independent sample of new patients. Recently, Nichols et al.3 modified the scoring system further and assessed its screening ability in various degrees of DES severity. They obtained a sensitivity and specificity of 82% and 36%, respectively, with a cutoff score of 14.5. However, they did not assess the performance in DES versus non-DES populations, making it difficult to compare the results with those of other studies.

Two major concerns surround the use of the McMonnies questionnaire for DES. First, there is no standardized scoring protocol. Second, there is uncertainty about whether the questionnaire can be used to grade disease severity. Although the primary intended purpose of the questionnaire was to assess risk for DES (i.e., screening), it has been used as a measure of symptom severity.1–5 Although instruments designed for screening merely must be able to separate the diseased from the nondiseased populations, instruments that are designed to measure the positioning of individuals along the continuum of the construct under measurement. Therefore, the ability of the McMonnies questionnaire to perform such a dual function should be tested.

Rasch models produce scales that have interval-level properties and offer several potential advantages.6 As they pertain to this study, one includes its usefulness in testing whether items from a questionnaire measure a unidimensional (i.e., all items measure a single trait) construct, which is necessary to justify the summation of scores.7,8 Second, Rasch models can determine the precision with which persons can be scaled so as to establish whether the questionnaire is capable of measurement or simply screening. Rasch analysis was used to investigate the performance of four dry eye questionnaires, including the McMonnies questionnaire, and it was shown to be unidimensional, albeit by the simple use of fit statistics.9 Recent studies, however, suggest the use of principal components analysis (PCA) to supplement the results of fit statistics to
We performed the analyses in three phases as follows: phase I, to determine whether the McMonnies questionnaire is a measure; phase II, to determine whether the questionnaire is a screening instrument by using the three proposed scoring systems (McMonnies, 198614 and 1987,2 and Nichols, 20043); and phase III, to determine whether the screening properties could be enhanced by using Rasch scaling, DA, or confirm unidimensionality in Rasch analysis10–13; this method has not been explored as yet for the McMonnies questionnaire.

Given the popularity of the McMonnies questionnaire in screening for DES, it is critical that its screening properties using the three proposed scoring systems (namely, McMonnies and Ho,14 McMonnies and Ho,2 and Nichols et al.3) are investigated in detail, to propose the most appropriate system to maximize its screening properties. To be able to answer this broad research question, we subdivided it into three smaller parts as follows: (1) Is the McMonnies questionnaire a measure? We used Rasch analysis to determine the answer. (2) Is the McMonnies questionnaire a screening instrument? We used receiver operating characteristics (ROC) curve analysis for this purpose. (3) Can the screening ability of the questionnaire be improved further? We performed discriminant analysis (DA) to determine whether the screening abilities could be enhanced by using Rasch scaling.

METHODS

McMonnies Questionnaire

Table 1 provides the item (question) content of the McMonnies questionnaire published in 1987.2 As mentioned, there are two additional scoring systems: one published by McMonnies and Ho in 198614 and the other by Nichols et al. in 20043.

The 1987 version of the McMonnies questionnaire was used in the present study. It is presented on a single page and includes 12 questions that focus on clinical risk factors for DES. The questions employ polytomous response options that vary in number and type. For example, question 1 has three response categories consisting of yes (2), no (0), and uncertain (1), whereas question 9 has four response categories, consisting of never (0), sometimes (1), often (2), and constantly (3).

For the final two questions (11 and 12), the three response categories are yes (2), no (0), and uncertain (1). However, the original version of the questionnaire included sometimes as a response to these questions,15 which appears appropriate, so we used this response option for a score of 1. All other response options were retained. For the purposes of the present study, each of the components of the two questions (2 and 7) were treated as individual items for analysis. The response option for each of these questions was dichotomous (yes, 1; and no, 0). Thus, we analyzed 25 questions in this version.

<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1†</td>
<td>Have you ever had drops prescribed or other treatment for dry eye?</td>
</tr>
<tr>
<td>2‡</td>
<td>Do you ever experience any of the following eye symptoms?</td>
</tr>
<tr>
<td>3§</td>
<td>How often do your eyes have these symptoms?</td>
</tr>
<tr>
<td>4∥</td>
<td>Do you regard your eyes as being unusually sensitive to cigarette smoke, smog, air conditioning, central heating?</td>
</tr>
<tr>
<td>5∥</td>
<td>Do your eyes easily become very red and irritated when swimming in chlorinated fresh water?</td>
</tr>
<tr>
<td>6∥</td>
<td>Are your eyes dry and irritated the day after drinking alcohol?</td>
</tr>
<tr>
<td>7‡</td>
<td>Do you take antihistamine tablets, antihistamine eye drops, diuretics, sleeping tablets, tranquilizers, oral contraceptives, or medications for peptic ulcers or duodenal ulcer, digestive problems, high blood pressure, or others?</td>
</tr>
<tr>
<td>8∥</td>
<td>Do you suffer from arthritis?</td>
</tr>
<tr>
<td>9∥</td>
<td>Do you experience dryness of the nose, mouth, throat, chest, or vagina?</td>
</tr>
<tr>
<td>10∥</td>
<td>Do you suffer from thyroid abnormality?</td>
</tr>
<tr>
<td>11∥</td>
<td>Are you known to sleep with your eyes partly open?</td>
</tr>
<tr>
<td>12∥</td>
<td>Do you have eye irritation as you wake from sleep?</td>
</tr>
</tbody>
</table>

* Response options for the present study were modified for four questions: questions 2 and 7 for purposes of ROC analysis; and uncertain was replaced with sometimes for questions 11 and 12.
† Response options for questions 1, 8, and 10: yes/no/uncertain.
‡ Response options for questions 7 and 2: yes/no (each component of these questions was treated as separate and so a total of 25 questions were analyzed).
§ Response options for questions 4, 5, 6, 11 and 12: yes/no/sometimes.
∥ Response options for questions 3 and 9: never/sometimes/often/constantly.

Participants

Forty-three female patients aged 45 years and over, with a confirmed diagnosis of Sjögren’s syndrome, were recruited through a rheumatology practice in the University of New South Wales, Sydney, Australia.1 The inclusion criterion was an established diagnosis of keratoconjunctivitis sicca, either alone or in conjunction with rheumatoid arthritis.

One hundred forty patients (with age and sex distribution similar to that of the included cases of DES) from a private optometric practice in the same area who presented for correction of refractive error without dry eye symptoms constituted the control group.

The McMonnies questionnaire was self-administered by each participant before any ocular examination. Those who could not understand written English were excluded from the study. All participants signed an Institutional Review Board informed consent form before evaluation. This study was approved by the Ethics Committee for Human Research, University of New South Wales, Sydney, Australia, and was conducted in accordance with the tenets of the Declaration of Helsinki.

Clinical Assessment

After completion of the questionnaire, all patients underwent routine clinical assessments for evaluation and confirmation of the diagnosis of DES. Specifically for patients attending the rheumatology clinic, these diagnostic tests included slit lamp (SL) examination of the eye, Schirmer test I and tear lysozyme assay. For the control group, the examination of asymptomatic patients (i.e., those who did not report a history of symptoms or denied them on specific questioning) included SL examination for corneal and conjunctival staining with fluorescein, hypofluorescence, reduced tear prism height, excess tear debris, short tear breakup times, and meibomian gland dysfunction (abnormal specular reflex interference pattern, reduced gland patency). This examination served the purpose of identifying any cases of dry eye disease that occurred without significant symptoms.

Statistical Analyses

We performed the analyses in three phases as follows: phase I, to determine whether the McMonnies questionnaire is a measure; phase II, to determine whether the questionnaire is a screening instrument by using the three proposed scoring systems (McMonnies, 198614 and 1987,2 and Nichols, 20043); and phase III, to determine whether the screening properties could be enhanced by using Rasch scaling, DA, or...
rescoring the questionnaire. The related statistical methods are described in the following sections.

**Rasch Analysis**

Data were analyzed by using the Andrich rating scale with Winsteps software (ver. 3.68). A four-rating model was used because there were four different groups of questions with various response categories. The Rasch model transforms raw ordinal measures into a linear continuous measure expressed as log-odds units, or logits. Before assessing the fundamental properties, we investigated the behavior of the response categories (specifically, the ordering of the thresholds). Thresholds are levels at which the response to either of two adjacent categories is equally likely. Disordered thresholds suggest that the participants have not used the response categories as intended and therefore collapsing the categories may be useful for improving performance. Once the thresholds were ordered, the fundamental psychometric properties were assessed as follows: (1) person separation reliability or the extent to which the items reliably distinguish strata of dry eye symptom severity in the patients. A person separation statistic of 2.0 is comparable to a reliability of 0.80, which is considered the minimum acceptable and enables the distinction of three strata. (2) fit or the extent to which the items in the questionnaire measure a single construct (i.e., unidimensionality) was assessed by fit statistics (infit mean-square, or infit MnSq) with values outside the range of 0.7 to 1.3 used to diagnose misfitting items. If these measurement properties of the Rasch model were met, then further assessments of validity such as PCA and differential item functioning were considered. A complete description of the Rasch model and these techniques is available elsewhere in the literature.

**ROC Analysis**

We performed the ROC analysis on the raw and Rasch-scaled scores to enable a comparison of the screening properties using these two approaches. ROC curves plot sensitivity on the y-axis versus (1-specificity) on the x-axis and provide the area under this curve (AUC). Values of the AUC range between 0 and 1, and the closer the value to 1, the better the overall diagnostic ability of the test. That is, for a random test with a discriminatory ability no better than chance (i.e., the test segregate participants randomly into control and dry eye groups), the AUC is 0.5; a value of 1.0 represents perfect discriminatory ability. However, the interpretation of AUC depends on the context of a test. Furthermore, real clinical cutoff values are influenced by the prevalence of the disease in the population. Summary measures of diagnostic utility were also generated (sensitivity, specificity, and negative predictive value, NPV) along with 95% confidence intervals (CI) for the AUC. Using z statistics, we compared the AUC for the Rasch-scaled discriminant score and the approach involving rescoring the questionnaire (these are described below).

**Rescoring to Optimize the Screening Properties of the Questionnaire**

Question 7 is related to the use of nine medications and is scored as yes (1) or no (0) for each of these, thereby, treating the score for each medication as an individual component. For example, if a participant took only one medication (e.g., antihistamine tablets) compared with another participant who took two medications (e.g., both an antihistamine and a diuretic), then whereas the former participant was assigned a score of 1, the latter was assigned a score of 2, a higher score for more medications, indicative of a higher risk of dry eye. This additivity of scores (with addition of each medication) implies that question 7 contributes the properties of a measure. Although such a scoring strategy suits the properties of a measure, it could compromise the screening properties of the questionnaire as a whole. Assuming that the use of any of the nine medications suggests risk of dry eye, the screening properties of the questionnaire may be enhanced if question 7 was scored dichotomously. Therefore, we rescorded this question to determine whether the screening properties of the questionnaire could be optimized.

**Discriminant Analysis**

A linear stepwise model of DA was used to determine which questions were the best predictors of DES, thereby reduce the number of questions required for screening. All the questions served as independent variables and the group membership (i.e., DES versus control) as the dependent variable. The DA used the Wilk’s , which calculates the usefulness of a given variable (i.e., question) and the F statistic, which shows which of the questions’ contributions are significant (with F = 3.84 to enter and F = 2.71 to remove, set by default in SPSS). We used both the raw and Rasch-scaled scores to compute classification (discriminant) functions designed to identify and combine the best items to differentiate control from DES patients.

In descriptive analyses, all P values were two-sided, with statistical significance set at P < 0.05 (SPSS for Windows ver. 15.0; SPSS, Chicago, IL).

**RESULTS**

**Phase I: Is the McMonnies Questionnaire a Measure?**

**Rasch Analysis: Assessment of Rating Scale Structure.** Category thresholds were disordered for two question groups, indicating that the response categories did not function as intended. In one group, the option uncertain (for questions 1, 8, and 10) was used by only 49 (10%) participants, suggesting infrequent use. However, combining uncertain with the other options, yes or no, appeared illogical, and therefore this category was dropped, and the new categories formed were yes (1) and no (0) (i.e., 0-missing-1). Likewise, the option sometimes was used by only 19% of the participants as a response to the other question group (4, 5, 6, 11, and 12). However, sometimes was considered to be affirmative and therefore was recorded as yes. Thus, the new categories formed were yes (1) and no (0) (i.e., 0-1-1) for these questions. After the reorganization of the response categories, they were all placed in order, and we proceeded to further analyses.

**Overall Performance.** The person separation reliability was poor (0.76 and 0.75 before and after collapsing categories, respectively) indicating that the McMonnies questionnaire was unable to distinguish between sufficient strata or groups (i.e., at least three) of a participant’s dry eye symptoms and therefore failed to function as a measure. Furthermore, two items misfit. One was related to symptoms secondary to swimming and the other to consumption of alcohol. After deletion of these items iteratively, the remaining 23 items fit the Rasch model, yet the person separation reliability failed to improve. Thus, the questionnaire was able to distinguish between only two strata (i.e., disease present versus absent), which confirms the original intention of its development; screening. Because of its failure to meet the minimum requirements of the Rasch model (i.e., as a measure), further assessment of psychometric properties was not considered.

**Phase II: Is the Questionnaire a Screening Instrument?**

Table 2 provides the results of the ROC analysis: sensitivity, specificity, negative predictive value (NPV), and area under the curve (AUC; with 95% CI) for the total raw scores from the three proposed scoring systems. The AUC was high (>0.8) for each system.
Accordingly, it does not meet the requirements of the Rasch McMonnies questionnaire does not function as a measure and, based on the findings in this population, we concluded that the McMonnies questionnaire be improved further.

**Table 2. Screening Properties of the McMonnies Questionnaire Using Different Scoring Systems**

<table>
<thead>
<tr>
<th>Scoring Basis</th>
<th>Cut-Off Score</th>
<th>Type Of Score</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>NPV</th>
<th>AUC (95% CI)</th>
<th>Questions Used (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMonnies and Ho¹⁴</td>
<td>19</td>
<td>Total raw</td>
<td>87</td>
<td>89</td>
<td>95</td>
<td>0.94 (0.89–0.97)</td>
<td>12</td>
</tr>
<tr>
<td>McMonnies and Ho²</td>
<td>8</td>
<td>Total raw</td>
<td>93</td>
<td>85</td>
<td>97</td>
<td>0.95 (0.91–0.98)</td>
<td>25</td>
</tr>
<tr>
<td>Nichols et al.³</td>
<td>15</td>
<td>Total raw</td>
<td>87</td>
<td>81</td>
<td>94</td>
<td>0.88 (0.81–0.93)</td>
<td>11</td>
</tr>
<tr>
<td><strong>Phase III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescored questionnaire*</td>
<td>9</td>
<td>Total raw</td>
<td>100</td>
<td>83</td>
<td>100</td>
<td>0.97 (0.94–0.99)</td>
<td>16</td>
</tr>
<tr>
<td>Rasch analysis</td>
<td>–14.3</td>
<td>Total Rasch</td>
<td>100</td>
<td>86</td>
<td>100</td>
<td>0.96 (0.92–0.99)</td>
<td>25</td>
</tr>
<tr>
<td>Discriminant analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMonnies and Ho¹⁴</td>
<td>3.18</td>
<td>Discriminant raw</td>
<td>93</td>
<td>89</td>
<td>97</td>
<td>0.96 (0.91–0.98)</td>
<td>5</td>
</tr>
<tr>
<td>McMonnies and Ho²</td>
<td>2.1</td>
<td>Discriminant raw</td>
<td>93</td>
<td>90</td>
<td>97</td>
<td>0.96 (0.92–0.98)</td>
<td>8</td>
</tr>
<tr>
<td>Nichols et al.³</td>
<td>0.22</td>
<td>Discriminant raw</td>
<td>87</td>
<td>99</td>
<td>97</td>
<td>0.97 (0.92–0.99)</td>
<td>7</td>
</tr>
<tr>
<td>McMonnies and Ho²</td>
<td>–2.29</td>
<td>Discriminant Rasch</td>
<td>95</td>
<td>94</td>
<td>99</td>
<td>0.99 (0.96–0.99)</td>
<td>7</td>
</tr>
<tr>
<td>Rescored questionnaire*</td>
<td>8.5</td>
<td>Discriminant raw</td>
<td>98</td>
<td>85</td>
<td>99</td>
<td>0.97 (0.94–0.99)</td>
<td>2</td>
</tr>
</tbody>
</table>

All discriminant functions use the standard canonical coefficients. * McMonnies 1987² scoring criterion, with a single modification: Question 7, on medication, was treated as a single item.

**Phase III: Can the Screening Ability of the Questionnaire Be Improved Further?**

**Rescoring the Questionnaire.** The sensitivity and NPV were 100% and AUC was 0.97 at a cutoff score of 9.

**Rasch Analysis.** The total Rasch-scaled score provided an AUC of 0.96 at a cutoff score of 8.5 logits using all 25 questions.

**Discriminant Analysis.** All the questions were included in a forward linear stepwise DA and questions that offered little or no contribution to the diagnosis of DES were removed from the analysis. DA was performed for three scoring systems using the raw scores, for rescoring the questionnaire, and for the Rasch-scaled scores (Table 2).

The AUC improved for all the scoring systems except for the approach involving rescoring the questionnaire. Of all the scoring systems, the discriminant Rasch-scaled score had the largest AUC (i.e., 0.97; 95% CI 0.97–1.00), but it was not significantly different from that of the rescored questionnaires (0.97 vs. 0.99; z = 1.11, P = 0.27). Figure 1 compares these two ROC curves for each participant in the control and the DES groups.

In all cases, DA reduced the number of questions for each of the scoring systems with minimum being two questions that was required for the DA of the rescored questionnaire.

**Discussion**

Based on the findings in this population, we concluded that the McMonnies questionnaire does not function as a measure and, accordingly, it does not meet the requirements of the Rasch model, because of its inability to distinguish between more than two strata (i.e., present or absent) of participants’ symptoms (person separation reliability, <0.80). However, it is possible that in a different population, with a greater spread of disease severity, the McMonnies questionnaire can function as a measure. This possibility should be tested in future studies.

**Figure 1.** Comparison of AUCs for identification of dry eye syndromes by the McMonnies questionnaire (Rasch-scaled discriminant score versus discriminant raw score from the rescored questionnaire). Data in parentheses are standard errors.
Results from the present study imply that the McMonnies questionnaire cannot be used to grade disease severity. Therefore, higher questionnaire scores signify a higher likelihood of dry eye, rather than more severe dry eye. Also, the McMonnies questionnaire is not an appropriate endpoint for clinical trials. Given these findings, the conclusions from studies that have used the questionnaire to grade disease severity may be questionable. Furthermore, the inadequate person separation explains the poor diagnostic accuracy reported by Nichols et al. on a sample that included different disease severity.

Our finding regarding the inability of the McMonnies questionnaire to grade disease severity is based on the use of a single model (i.e., the Rasch model, from the item response theory [IRT]). However, other IRT-based measurement models—for example, the nonparametric Mokken model—also exist. Mokken scaling has been used in disability studies and nursing research. The Mokken model offers a method of selecting items in situations in which it is useful to have items representing different difficulty levels in a scale. The McMonnies questionnaire appears to be an obvious example, where each item, and not just the total score, contains information about the presence of dry eye, were participants to endorse it. Therefore, if the McMonnies questionnaire can be demonstrated to have hierarchical properties using Mokken scaling, then it would indicate that the items are ordered relative to one another and also along the latent trait, the presence of dry eye. Future research could be directed toward the use of Mokken scaling for the McMonnies questionnaire.

In the second phase, we assessed whether the questionnaire, as originally intended, possesses the qualities of a good screening instrument. Except for the total raw score from the scoring system of Nichols et al., all others demonstrated high sensitivity (>90%). Nevertheless, the AUC was >0.8, regardless of the scoring system used, indicating that the questionnaire possesses the required screening abilities. In comparison, however, the raw score from Nichols et al. had the largest AUC (0.88) in the present study; this relatively lower AUC is probably related to their scoring system. Nichols et al. modified the scoring system (for three questions), albeit with inconsistencies. For example, an affirmative response is always assigned a higher score for all questions except question 4. In addition a score of 0 was assigned to both categories uncertain and no for question 11, but these categories were assigned different scores for another question (12). Such incongruity may partly explain the smallest AUC (0.88) with the use of this scoring system in the present study.

In the third phase, using three approaches—rescoring the questionnaire, Rasch scaling, and DA—we then investigated, whether the screening abilities could be further enhanced. Rescoring the questionnaire (i.e., question 7, raised the sensitivity and NPV to 100%, from 93% for the original scoring) and the AUC to 0.97. Similarly, Rasch-scaling of raw scores improved the sensitivity and NPV to 100% as well as the AUC. This increase in the sensitivity suggests that there was noise in the computation of the raw scores. The DA further enhanced the AUC (except for the approach involving rescoring the questionnaire) and decreased the number of questions required for all the scoring systems, indicative of redundancy. Notably, the discriminant Rasch-scaled score provided the largest AUC of 0.99 (although not significantly) at a cutoff of −2.29 logits, albeit with seven questions. By comparison, the approach involving rescoring the questionnaire required only two questions. These questions were use of medications and eye irritation on waking up from sleep for screening of DES patients. Shorter questionnaires help reduce respondent burden. Furthermore, our results revealed that shortening the questionnaire did not disadvantage screening for DES in our sample. This finding was evidenced by the lack of reduction in the AUC during DA. On the other hand, the AUC increased (albeit insignificantly) for the Rasch-scaled score. The screening properties of the questionnaire in the present study are comparable to or are slightly superior to results previously reported by McMonnies et al. and McMonnies and Ho. Although the AUC obtained from the Rasch-scaled score appears greater than that in a recent study (AUC = 0.88) comparing the performance of four dry eye questionnaires using Rasch analysis in a Canadian non-contact-lens-wearing population, the differences in the populations do not permit valid comparisons. Nevertheless, the Rasch-scaled discriminant cutoff score of −2.29 logits in the present study provided high sensitivity and NPV (>90%) and used seven items. Although it would be tempting to recommend this set of seven items for future applications, the nature of the sample (extreme cases) included in the present study would preclude any such recommendation (Table 3). We speculate that the extreme sample dichotomy (dry eye versus no dry eye) led to much larger differences in between-group responses than variability in within-group responses. In these circumstances, very few items are needed, and this is exactly what we observed. Nevertheless, this small set of seven items may not be applicable to another population with a broad mix of disease severity where the group separation is less. We believe that rather than fit the questionnaire to the sample as we did in the present study, it would be more appropriate to fit it to the population it represents (i.e., a range of severity of dry eye), and any modifications in a questionnaire to optimize performance should be confirmed in a different sample.

Nevertheless, we believe that in the seven-item format, the McMonnies questionnaire can be administered in a relatively short time and can be scored immediately by nonmedical personnel. Using ROC analysis, the particular cutoff score chosen could be based on the relative costs of false negatives versus false positives. Although the consequences of a false negative could be significant for some potentially sight-threatening conditions, the consequences are perhaps not so detrimental for dry eye because of the low risk of irreversible ocular damage and the option of the patient’s presenting again to the practitioner in case of increase in symptoms. The clinician or researcher could choose the cutoff scores, depending on the purpose, to optimize the sensitivity or specificity of the scale. However, high sensitivity as was the case with the seven-item format is potentially useful for researchers who would like to ascertain high-risk samples for further study and also in specialty clinics—for example, a rheumatology clinic—

### Table 3. Item Content of the Revised Seven-Item McMonnies Questionnaire

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>2*</td>
<td>Do you ever experience dryness in your nose?</td>
</tr>
<tr>
<td>2</td>
<td>Do you ever experience grittiness in your eyes?</td>
</tr>
<tr>
<td>2</td>
<td>Do you ever experience burning in your eyes?</td>
</tr>
<tr>
<td>4</td>
<td>Do you take diuretics (fluid tablets)?</td>
</tr>
<tr>
<td>5</td>
<td>Do you take medication for digestive problems?</td>
</tr>
<tr>
<td>6</td>
<td>Do you suffer from arthritis?</td>
</tr>
<tr>
<td>7*</td>
<td>Do you ever experience dryness of the nose, mouth, throat, chest, or vagina?</td>
</tr>
</tbody>
</table>

* Response options: yes/no.
† Response options: never/sometimes/often/constantly.
where clinicians could screen for DES in a short time and then refer the patient to an eye care professional.

The added advantage of using the Rasch-scaled score is that it computes an estimate from available data rather than requiring imputation of scores (such as average score) found in classic methods. Thus, missing data are accounted for during Rasch analysis. This is of significant importance in the McMonnies questionnaire as two questions (related to secondary symptoms after swimming and drinking alcohol) did not apply to a large number (36%) of participants. The high missing response rate concurs with the findings of the present study. The present study shares many of the drawbacks of any screening study (Table 4)—most important, selection bias. That is, the included sample influenced the degree of generalizability of the results. Patients in whom the diagnosis is uncertain are often the cases that the clinician in a primary eye care clinic would encounter, and these patients were excluded from the study. Therefore, the sample included essentially represented a dichotomy (dry eye versus no dry eye). Rather than a dichotomy, dry eye represents a continuum of disease with severity ranging from mild to severe. Thus, ideally, one would include a better representation of the dry eye disease spectrum, but this was not possible in the present study. Nevertheless, this study provides valuable information about the McMonnies questionnaire. However, the results from the present study need replication in a less selective population (i.e., with a disease spectrum that is more representative of populations, such as milder forms of dry eye disease) such as that encountered by clinicians in primary eye care.

In conclusion, the McMonnies questionnaire is not a measure, but is an effective screening instrument. Rasch scaling is beneficial in screening instruments, as it helps reduce noise and thereby improve discrimination, although its benefits are not as great as when used to enhance the measurement properties. The revised format of the seven questions of the McMonnies questionnaire will render it an appropriate, inexpensive, low-burden, and easy-to-use tool for identifying patients at risk of having DES in primary eye care clinics.

Acknowledgments

The authors thank Denis Wakefield, University of New South Wales, Australia, for cooperation in the acquisition of responses from patients with keratoconjunctivitis sicca syndrome.

References


Table 4. Review of Studies That Used the McMonnies Questionnaire

<table>
<thead>
<tr>
<th>Study</th>
<th>Cutoff Score</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>McMonnies and Ho²</td>
<td>14.5</td>
<td>98</td>
<td>97</td>
<td>NA</td>
</tr>
<tr>
<td>Golding and Brennan*</td>
<td>14.0</td>
<td>87</td>
<td>87</td>
<td>NA</td>
</tr>
<tr>
<td>McMonnies and Ho¹</td>
<td>14.5</td>
<td>92</td>
<td>93</td>
<td>NA</td>
</tr>
<tr>
<td>Nichols et al.³</td>
<td>14.5</td>
<td>82</td>
<td>36</td>
<td>0.65</td>
</tr>
<tr>
<td>Simpson et al.²</td>
<td>NA</td>
<td>80 (approximately)†</td>
<td>90 (approximately)†</td>
<td>0.88</td>
</tr>
</tbody>
</table>

† Discriminant functions provided in Table 2.