

Validation of an Objective Measure of Dry Eye Severity

Sezen Karakus¹, Esen K. Akpek¹, Devika Agrawal¹, and Robert W. Massof¹

¹ The Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland, USA

Correspondence: Robert W. Massof, The Wilmer Eye Institute at Johns Hopkins, 600 N Wolfe St, Wilmer B43, Baltimore, MD 21287-2816, USA. e-mail address: bmassof@jhmi.edu

Received: 29 October 2017

Accepted: 22 July 2018

Published: 10 October 2018

Keywords: dry eye; grading; Rasch analysis; severity

Citation: Karakus S, Akpek EK, Agrawal D, Massof RW. Validation of an objective measure of dry eye severity. *Trans Vis Sci Tech.* 2018; 7(5):26, <https://doi.org/10.1167/tvst.7.5.26>

Copyright 2018 The Authors

Purpose: We evaluated the validity of a single dry eye severity measure estimated using Rasch analysis from a battery of clinical tests and patient symptoms.

Methods: This study included 203 dry eye patients and 51 controls. Administered tests included the Ocular Surface Disease Index (OSDI), tear osmolarity, Schirmer's test, noninvasive break-up time, and ocular surface staining. Each of the 12 OSDI questions and each clinical test was defined to be a separate indicator to estimate a single dry eye severity measure from Rasch analysis. Measures of severity were estimated for each subject (person measures) and measures of sensitivity to severity were estimated for each sign and symptom (indicator measures).

Results: The average severity measure for dry eye patients was significantly greater than the average severity measure for controls (-0.39 vs. -1.2 , $P < 0.001$). The distribution of indicator measures was well matched to the distribution of person measures. No indicator carried $>10\%$ of the total information about dry eye severity carried by all indicators together. However, the most informative indicators were corneal and conjunctival staining.

Conclusions: Our study indicated that there is no single "best" dry eye test. Clinical tests and symptoms should be used in combination to estimate a single dry eye severity measure.

Translational Relevance: There is no single "gold standard" testing method for dry eye that correlates with the severity of disease. We propose that Rasch analysis can be used to calculate an objective dry eye severity score from a battery of clinical indicators.

Introduction

Dry eye is a common condition worldwide affecting up to one-third of individuals older than 50 years.¹⁻⁵ Dry eye is characterized by loss of homeostasis of the tear film, which results in tear film instability, increased tear osmolarity, ocular surface inflammation and damage, and neurosensory abnormalities.⁵ These signs are accompanied by symptoms of ocular discomfort and fluctuating vision that impact quality of life.⁵ A variety of clinical tests as well as symptom questionnaires are available to diagnose dry eye and determine disease severity.⁵⁻⁹ However, there is little correlation among or between physician-observed signs and patient-reported symptoms, which leads to confusion in monitoring changes and managing the disease.^{10,11}

Even though dry eye is one of the most prevalent ocular conditions affecting an estimated 25 million

individuals in the United States alone (Market Scope. 2011 Comprehensive Report on the Global Dry Eye Products Market. St. Louis, MO: Market Scope, November 2011), surprisingly few approved and effective treatments of dry eye exist.¹² One reason for this can be that investigators tend to seek a single "gold standard" to measure dry eye disease severity or response to treatment, which has eluded the dry eye field.⁵ The lack of a single "gold standard" sign or symptom that correlates with dry eye state has been noted again in the recently updated Tear Film and Ocular Surface Society (TFOS) Dry Eye Workshop (DEWS) II Diagnostic and Methodology report.⁵ A new diagnostic scheme requiring a combination of symptoms and clinical signs has been proposed instead of using a severity grading table. The original and recently updated scheme follow the recommendations of an expert panel of clinicians and scientists.⁵ The original severity grading scale was modified previously to make it a single explicit continuous

variable for the purpose of exploring its functional relationship to measures or ordinal scaling of dry eye signs and symptoms.¹³ This assumption of a single continuous severity variable for dry eye is implicit in the work of consensus panels,¹⁴ estimation of composite scores or measures from dry eye symptom,¹⁵ and requirement of a “single best test” for evaluating the efficacy of dry eye treatments by regulatory agents.^{13,16}

The hypothesis is that the magnitude of a single dry eye severity variable can be estimated from information carried by clinical signs and patient-reported symptoms using the mapping noise measurement model suggested previously.¹⁷ The mapping noise model reduces to a form equivalent to the Masters partial credit model¹⁸ when the constraints of axiomatic measurement theory (i.e., independent identically distributed random variables) and the assumption of a standardized logistic distribution ($\mu = 0$; $\sigma = \pi/\sqrt{3}$) of the random variables are imposed. Rasch analysis, which uses the Masters partial credit model, assumes that dry eye severity is a latent variable that cannot be observed directly, but has a probabilistic relationship to manifest variables, such as signs and symptoms, that can be observed. In other words, “dry eye severity” is a theoretical construction from observations, and not an observation per se. Rasch models assume that patients differ from each other only in the severity of their dry eye. The magnitude of each observed sign and symptom is reported in its own units (e.g., mm of wetting, elapsed time in seconds, ordinal grade of staining), which the Rasch model assumes can be mapped to a common latent dry eye severity variable and that the unique mapping function for each sign and symptom is the same for every patient. The Rasch model then is used to estimate the most likely value of the patient’s dry eye severity from the pattern of observed values of the patient’s signs and symptoms.¹⁷

We estimated a severity measure using this unique approach and tested its internal and external validity on a sample of physician-diagnosed dry eye patients (including patients with and without Sjögren’s syndrome [SS]) and a sample of nonpatient controls (which could include cases of previously undiagnosed dry eye).

Methods

Patients and Clinical Testing

This prospective clinical observational study was approved by the Johns Hopkins University institu-

tional review board and adhered to the tenets of the Declaration of Helsinki. Dry eye patients with or without SS (previously established based on the American College of Rheumatology [ACR] Criteria¹⁹ endorsed in 2012) older than 18 years were recruited from the Wilmer Eye Institute Ocular Surface Diseases and Dry Eye Clinic, Johns Hopkins University (Baltimore, MD). Patients with a prior physician-made diagnosis of dry eye, who were taking over-the-counter or prescription dry eye treatments, and who had dry eye-related procedures (including but not limited to tear duct occlusion, lipiflow, or intense pulse light treatment) were included in the dry eye group, independent of their ocular surface or tear film parameters. Consequently, the dry eye group subjects had a broad spectrum of severity of findings changing from mild to severe.⁵ Controls were volunteers older than 18 years who did not carry a physician-made diagnosis of dry eye, blepharitis, allergic conjunctivitis, or other ocular surface disease, were not seeking or had not sought eye care for ocular surface disease symptoms before enrollment in the study. Additionally, subjects who had any ocular surgery within 3 months of the study visit were not included. Patients were asked to discontinue use of any eye drops, including artificial tears, 12 hours before examination and patients and controls were asked to stop wearing contact lenses at least 1 week before examination.

The number of subjects in our study was chosen to be comparable to that of an earlier study evaluating the validity of the added noise version of the Rasch measurement model.¹⁷ Subgroups of subjects were chosen specifically to represent the two extremes of the dry eye severity measure distribution (controls and SS-related dry eye patients) and the broad middle of the distribution (non-SS dry eye patients).

After obtaining informed consent in accordance with the Health Insurance Portability and Accountability Act (HIPAA), a detailed review of systems was performed followed by an Ocular Surface Disease Index (OSDI) symptom questionnaire (Allergan, Inc., Irvine, CA) completed by each participant.²⁰ The following tests then were performed with 10-minute intervals in between in the order listed: tear osmolarity (TearLab Corporation, Inc., San Diego, CA), Schirmer’s test without topical anesthesia (Tear Flo; Sigma Pharmaceuticals, Monticello, IA), automated noninvasive break-up time (NIBUT; Tear Stability Analysis System [TSAS] used on the RT-7000 Auto Refractor-Keratometer; [Tomey Corporation, Nagoya, Japan), corneal staining with fluorescein (Fulglo; Akorn, Inc., Lake Forest, IL), and conjunctival

staining with lissamine green (Green-Glo; HUB Pharmaceuticals, LLC., Rancho Cucamonga, CA). The order of tests was adopted from the Sjögren's International Collaborative Clinical Alliance (SICCA) study to obtain most accurate results.²¹ Schirmer's test was recorded at 1 minute and then at 5 minutes for each eye separately. Values of NIBUT greater than 3.0 seconds were considered normal as previously recommended.²² The SICCA grading system²¹ was used to rate corneal and conjunctival staining. The maximum possible corneal staining score for each cornea was 6. Nasal and temporal conjunctiva were graded separately with a maximum score of 3 for each area. The total possible maximum ocular staining score (OSS) was 12 for each eye.

Estimating a Dry Eye Severity Measure

A single dry eye severity measure was estimated for each participant from Rasch analysis using a grouped item version of the Masters partial credit model²³ of clinical test results and patient responses to each of the 12 items in the OSDI questionnaire. Each of the six clinical tests, including tear osmolarity, 1-minute Schirmer's test, 5-minute Schirmer's test, NIBUT, corneal staining, and conjunctival staining, and each of the 12 OSDI questions were defined to be a separate indicator variable, or "item," in the partial credit model with five groupings of the indicators that used the same scoring:²⁴ (1) tear osmolarity, (2) 1- and 5-minute Schirmer's tests, (3) NIBUT, (4) corneal and conjunctival staining, and (5) 12 OSDI items. The order of the tests must be the same within the study cohort to obtain a valid severity measure; however, a different order also can be used, if desired, as long as one test does not affect the following test adversely. The mean between the two eyes was calculated for each participant to determine a single variable per person for each clinical test, since the OSDI items do not differentiate between eyes. For the purpose of analysis, continuous clinical test variables (tear osmolarity, 1- and 5-minute Schirmer's tests, and NIBUT) were binned into quintiles and assigned rank scores. Rank scores also were assigned to the ordinal OSDI response categories for each question. Raw ordinal clinician rating scores of corneal and conjunctival staining were accepted at face value for Rasch analysis. The polarities of the indicator scores were adjusted so that in each case a greater score corresponded to greater dry eye severity. Measures of dry eye severity were estimated for each participant (person measures) and measures of sensitivity to dry eye severity were estimated for each clinical test and

OSDI question (indicator measures). The probability of observing each ordinal category as a function of dry eye severity also was estimated for each indicator.

Evaluating Internal Validity of the Measurement Model

The mapping error measurement model assumes there is a single systematic source of true variance (i.e., dry eye severity) and that any other sources of variance in the observed indicators, whether within or between persons, are random. If the estimated dry eye severity measurement model is internally valid (i.e., observations fit the model assumptions), then the ratio of the variance in the mean squared residuals (squared difference between observed and expected indicator rank scores) to the expected variance of the noise should be distributed as χ^2/df across persons and across indicators.²⁵ This ratio is equivalent to squared residuals summed across persons, to evaluate indicator fit, or across indicators, to evaluate person fit, multiplied by Fisher information, summed across persons or indicators, respectively (i.e., information-weighted mean square fit statistic or "infit").²⁶ The cube-root of χ^2 is approximately normally distributed.²⁷ Therefore, to simplify evaluation, we transformed the infit mean squares to z-scores (i.e., standard normal distribution), which have an expected value of 0 and a standard deviation of 1 (with 95% of the values falling within ± 2 standard deviations [SD] of the expected value).

If the estimated measure is a unidimensional variable with a single source of true variance, then the unexplained variance should be random. To test this hypothesis, we evaluated differential person functioning (DPF) and we performed a principal components analysis on the indicator score residuals (difference between the observed indicator value and the value expected by the model) and evaluated the scree plot to determine if there was evidence of nonrandom structure in the variance.

Evaluating External Validity of the Estimated Measures

If the estimated measures are, indeed, measures of dry eye severity, then the measures should discriminate cases from controls with cases having worse dry eye. This hypothesis was tested by comparing the distributions of case person measures to control person measures and performing a *t*-test on these two distributions. Also, from clinical experience, we expect that patients with SS will have worse dry eye

Table 1. Characteristics of Subjects According to Dry Eye Status

	Dry Eye, <i>n</i> = 203	Control, <i>n</i> = 51	<i>P</i> Value
Demographics			
Age, years	57.8 (12.9)	45.5 (16.1)	<0.001
Female, <i>n</i> (%)	175 (86.2%)	34 (66.7%)	0.02
Dry eye measures			
OSDI score, 0–100	37.5 (24.5)	15.7 (16.5)	<0.001
Osmolarity, mOsm/L	303.4 (12.0)	304.9 (11.8)	0.42
Osmolarity inter-eye difference, mOsm/L	9.4 (11.2)	10.0 (14.0)	0.77
Schirmer's test at 1 minute, mm	6.5 (6.6)	9.2 (8.7)	0.04
Schirmer's test at 5 minutes, mm	14.3 (9.8)	18.2 (10.5)	0.02
NIBUT, second	3.4 (3.1)	5.2 (3.3)	0.001
Ocular staining score, 0–12	4.8 (3.5)	1.4 (1.9)	<0.001
Corneal staining score, 0–6	2.4 (1.8)	0.7 (1.0)	<0.001
Conjunctival staining score, 0–6	2.4 (2.2)	0.8 (1.3)	<0.001

Results are represented as mean (SD) for continuous variables and number (percentage) for binary variables.

than do non-SS patients.²⁸ This hypothesis also was tested by comparing distributions of estimated person measures between SS and non-SS cases and performing a *t*-test. In addition, a linear regression analysis after controlling for age and sex was performed to determine the associations between the dry eye status (SS or non-SS) and the magnitude of the severity measure.

Results

Among the 254 participants, 203 were dry eye patients (55 had a previously established diagnosis of

SS based on 2012 ACR Criteria¹⁹) and 51 with no previous diagnosis of dry eye or any ocular surface disease were included as controls. Participant characteristics are summarized in [Tables 1](#) and [2](#). The dry eye group included a significantly higher proportion of females ($P = 0.02$). Except for tear osmolarity, dry eye measures were significantly worse in patients with dry eye compared to controls. Similarly, except for the OSDI score and tear osmolarity, dry eye measures were significantly worse for SS than for non-SS dry eye patients. As itemized in [Table 3](#), Spearman correlations between OSDI questions (symptoms) and clinical signs were extremely low

Table 2. Characteristics of Subjects With SS-Related Dry Eye Versus Non-SS Dry Eye

	SS-Dry Eye, <i>n</i> = 55	Non-SS Dry Eye, <i>n</i> = 148	<i>P</i> Value
Demographics			
Age, years	54.3 (12.6)	59.1 (12.8)	0.04
Female, <i>n</i> (%)	51 (92.7%)	124 (83.8%)	0.06
Dry eye measures			
OSDI score, 0–100	38.8 (24.5)	37.0 (24.5)	0.65
Osmolarity, mOsm/L	304.2 (9.5)	303.1 (12.8)	0.49
Osmolarity inter-eye difference, mOsm/L	11.4 (12.4)	8.6 (10.7)	0.15
Schirmer's test at 1 minute, mm	4.6 (6.2)	7.2 (6.7)	0.01
Schirmer's test at 5 minutes, mm	11.6 (9.5)	15.3 (9.8)	0.02
NIBUT, second	2.6 (3.0)	3.7 (3.1)	0.03
Ocular staining score, 0–12	7.2 (3.5)	3.9 (3.1)	<0.001
Corneal staining score, 0–6	3.3 (1.6)	2.1 (1.7)	<0.001
Conjunctival staining score, 0–6	4.0 (2.1)	1.8 (1.9)	<0.001

Results are represented as mean (SD) for continuous variables and number (percentage) for binary variables.

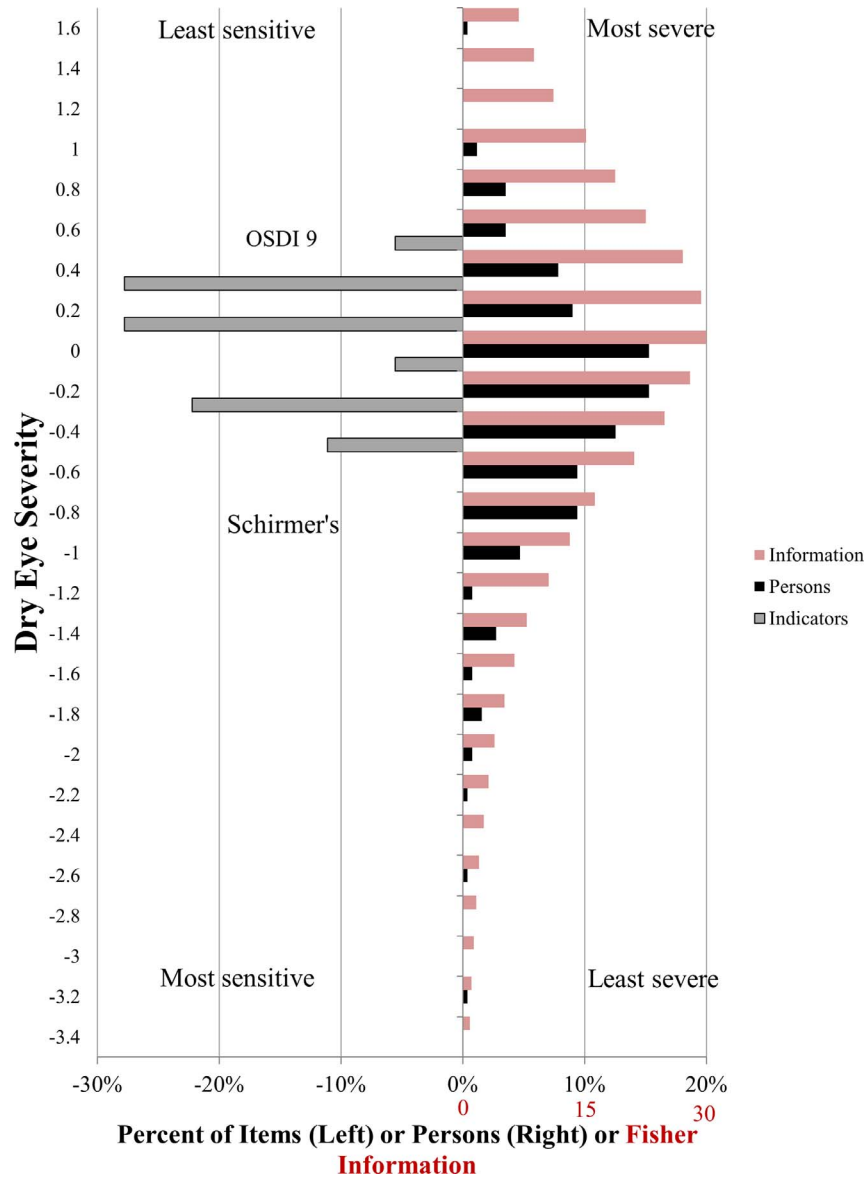


Figure 1. The distribution of estimated indicator sensitivity measures (*gray bars*) and the distribution of estimated dry eye severity measures for persons (*black bars*). *Orange bars* illustrate the Fisher information (in logit^{-2} units) carried by all 18 indicators combined as a function of the measure.

(mean = 0.12; range = 0.25–0.02). Correlations among OSDI questions were moderate to high (mean = 0.54; range = 0.36–0.79). The two Schirmer’s test measures (at 1 and 5 minutes) were highly correlated (0.90), as were corneal and conjunctival staining scores (0.69).

The distribution of indicator sensitivity measures was well matched to the combined distribution of dry eye severity measures for dry eye patients and controls (Fig. 1). Schirmer’s tests at 1 and 5 minutes were the most sensitive of these 18 indicators, followed closely by tear osmolarity. OSDI item 9 (watching TV) was

the least sensitive indicator, preceded by corneal staining. Overall measurement reliability was 0.82 for persons (i.e., on average, 18% of variance between persons in the observed dry eye severity distribution can be attributed to estimation error) and 0.97 for indicators (i.e., on average, 3% of variance between indicators in the observed indicator sensitivity distribution can be attributed to estimation error). The maximum dry eye severity information carried by each indicator ranged from 1.1 for NIBUT to 3.4 for corneal and conjunctival staining (Table 4, Fig. 2). No single indicator carried >10% of the total dry eye

Table 3. Matrix of Spearman's Inter-Item Correlations Between Dry Eye Indicators

	Osmolarity	Schirmer's 1-min	Schirmer's 5-min	NIBUT	Corneal Staining	Conjunctival Staining	OSDI 1	OSDI 2
Osmolarity	1.00							
Schirmer's 1-min	-0.03	1.00						
Schirmer's 5-min	0.03	0.90	1.00					
NIBUT	0.11	0.38	0.41	1.00				
Corneal staining	-0.03	0.33	0.40	0.36	1.00			
Conjunctival staining	-0.05	0.40	0.44	0.38	0.69	1.00		
OSDI 1	0.13	0.10	0.14	0.12	0.20	0.23	1.00	
OSDI 2	0.18	0.04	0.09	0.03	0.10	0.10	0.46	1.00
OSDI 3	0.16	0.04	0.10	0.08	0.18	0.13	0.57	0.59
OSDI 4	0.11	0.10	0.13	0.09	0.20	0.19	0.55	0.37
OSDI 5	0.08	0.12	0.10	0.07	0.23	0.21	0.53	0.37
OSDI 6	0.16	0.13	0.17	0.15	0.22	0.15	0.55	0.44
OSDI 7	0.02	0.10	0.10	0.08	0.13	0.11	0.58	0.39
OSDI 8	0.11	0.07	0.10	0.13	0.12	0.13	0.62	0.42
OSDI 9	0.18	0.07	0.08	0.13	0.15	0.13	0.50	0.45
OSDI 10	0.24	0.02	0.10	0.09	0.19	0.10	0.53	0.56
OSDI 11	0.25	0.12	0.16	0.10	0.12	0.15	0.54	0.61
OSDI 12	0.20	0.09	0.12	0.02	0.12	0.08	0.46	0.54

Spearman's rank correlation coefficient (ρ) was used to analyze the correlations between variables.

Table 4. Item Measures, Maximum Item Information, and Infit Mean Square Fit Statistics for the 18 Indicators

Indicator	Item Measure (SE)	Maximum Item Information	Infit Mean Square
Osmolarity	-0.35 (0.06)	1.6	8.91
Schirmer's test at 1 minute	-0.59 (0.05)	1.65	3.88
Schirmer's test at 5 minutes	-0.41 (0.05)	1.65	2.63
NIBUT	-0.32 (0.07)	1.1	2.70
Corneal staining	0.02 (0.04)	3.4	2.00
Conjunctival staining	0.01 (0.04)	3.4	5.47
OSDI 1 (sensitive to light)	-0.2 (0.05)	1.9	-1.43
OSDI 2 (feel gritty)	0.08 (0.06)	1.9	-2.04
OSDI 3 (painful or sore eyes)	0.23 (0.06)	1.9	-3.91
OSDI 4 (blurred vision)	0.06 (0.06)	1.9	-4.21
OSDI 5 (poor vision)	0.32 (0.06)	1.9	-1.88
OSDI 6 (difficulty with reading)	0.35 (0.06)	1.9	-2.61
OSDI 7 (difficulty with driving at night)	0.23 (0.06)	1.9	0.42
OSDI 8 (difficulty working on computer)	0.39 (0.06)	1.9	-3.61
OSDI 9 (difficulty with watching TV)	0.57 (0.07)	1.9	-2.35
OSDI 10 (windy condition)	-0.27 (0.06)	1.9	-1.67
OSDI 11 (areas with low humidity)	-0.18 (0.06)	1.9	-2.19
OSDI 12 (areas that are air-conditioned)	0.07 (0.06)	1.9	-1.98

SE, Standard error.

Table 3. Extended

	OSDI 3	OSDI 4	OSDI 5	OSDI 6	OSDI 7	OSDI 8	OSDI 9	OSDI 10	OSDI 11	OSDI 12
Osmolarity										
Schirmer's 1-min										
Schirmer's 5-min										
NIBUT										
Corneal staining										
Conjunctival staining										
OSDI 1										
OSDI 2										
OSDI 3	1.00									
OSDI 4	0.48	1.00								
OSDI 5	0.46	0.72	1.00							
OSDI 6	0.54	0.61	0.58	1.00						
OSDI 7	0.49	0.55	0.59	0.59	1.00					
OSDI 8	0.62	0.61	0.54	0.79	0.65	1.00				
OSDI 9	0.55	0.53	0.53	0.66	0.62	0.77	1.00			
OSDI 10	0.61	0.49	0.45	0.55	0.50	0.54	0.47	1.00		
OSDI 11	0.60	0.48	0.43	0.46	0.53	0.51	0.47	0.73	1.00	
OSDI 12	0.54	0.42	0.36	0.46	0.49	0.51	0.46	0.73	0.76	1.00

severity information carried by all 18 indicators together (ALL ITEMS).

Of the participants, 5% had infit mean squares that exceeded expectations by >2.5 SD and 2% had infit mean squares that fell short of expectations by >2.5 SD (Fig. 3). These outliers represented significant departures from the expected distribution ($P = 0.011$ for the 2-tail Kolmogorov-Smirnov [K-S] test). When the distribution was truncated to remove these outliers, then the distribution was not significantly different from the expected normal distribution ($P = 0.16$ for the 2-tail K-S test). The indicator infit mean square z -scores had a bimodal distribution with large positive values for all clinical test indicators (average infit mean square in Table 4 is 1.375) and large negative values for all but one of the OSDI items (average infit mean square in Table 4 is 0.818; Fig. 4). The indicator measure is negatively correlated with the infit mean square z -score (Pearson $r = 0.60$), which suggests nonuniform DPF.²⁹ These results indicated that there must be two sources of noise, with the average variance 1.7 times greater for clinical signs than it is for symptoms. Also, there was a trend for clinical signs to be more sensitive than symptoms to dry eye severity.

As illustrated by the scatter plots in Figure 5, the OSDI-based person measures and the clinical sign-based person measures are linear with the person

measures estimated from all 18 indicators. Consistent with the hypothesis of two different sources of noise with 1.7 times more noise variance for clinical signs, unexplained random variance about the regression line is 2 times greater for clinical sign-based measures ($R^2 = 0.34$) than it is for OSDI-based measures ($R^2 = 0.68$). The difference in slopes of the regression lines (1.75 for OSDI and 0.87 for clinical signs) also was consistent with differences in noise variance associated with the two sets of indicators, as opposed to uniform DPF, which would suggest the two sets of indicators were sampling two different latent variables in the sample of patients (see the Supplementary Material for the theoretical interpretation of the slopes of the regression lines).

Principal components analysis (PCA) of response residuals showed that the estimated measures explained only 36% of the observed variance (Fig. 6A). The first two components of the residuals together accounted for 40% of the remaining variance (26% of the total variance). The higher order components accounted for $<5\%$ of the total variance and can be considered part of the scree (unstructured background noise). Figure 6B illustrates that the variance of response residuals for clinical sign indicators (black points) was in the direction of the first component and the variance of response residuals for OSDI items (gray points) was in the direction of the second

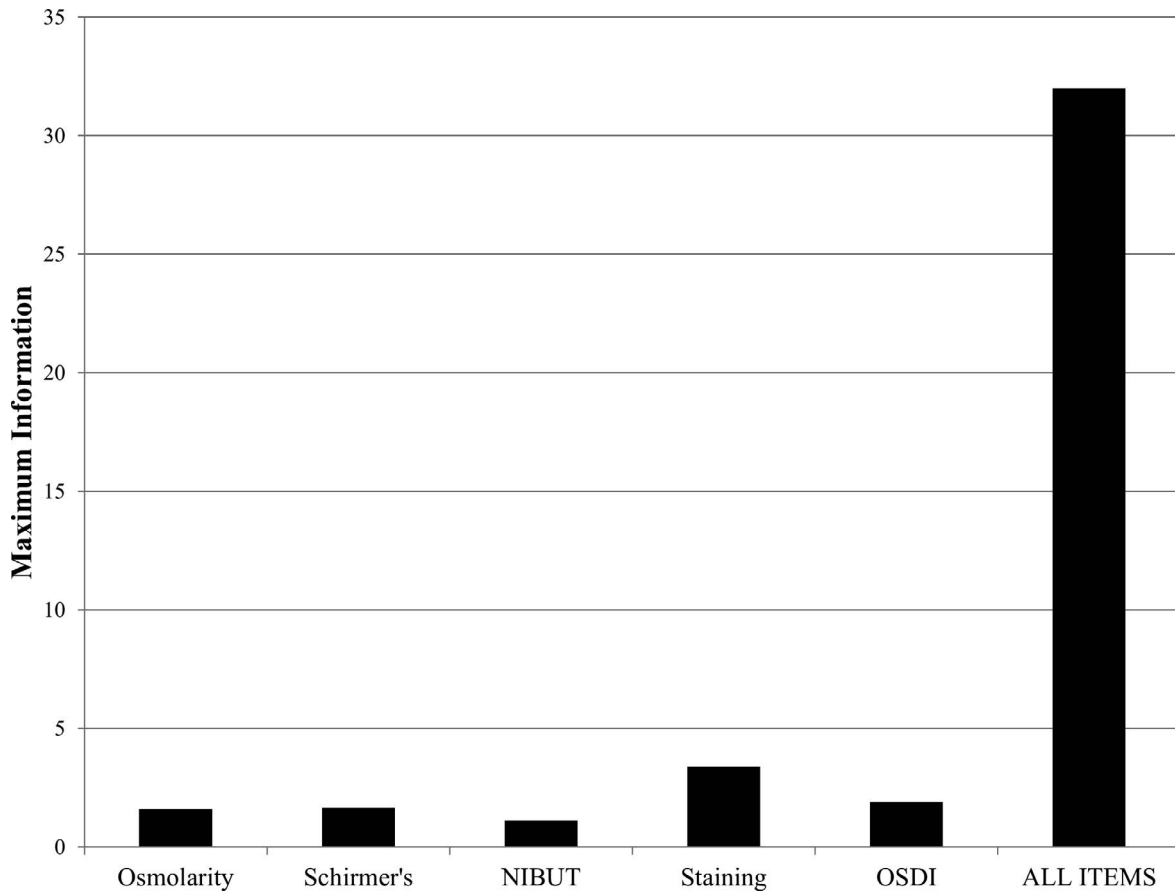


Figure 2. Maximum dry eye severity information carried by each indicator and all indicators.

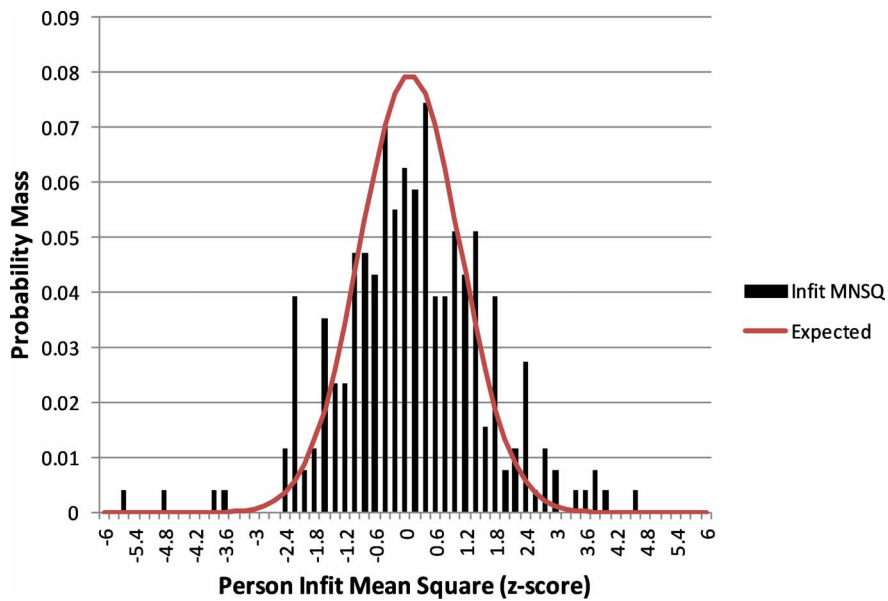


Figure 3. The probability mass distribution (black bars) of infit mean square z-scores for the persons compared to the probability mass distribution expected by the measurement model (red curve).

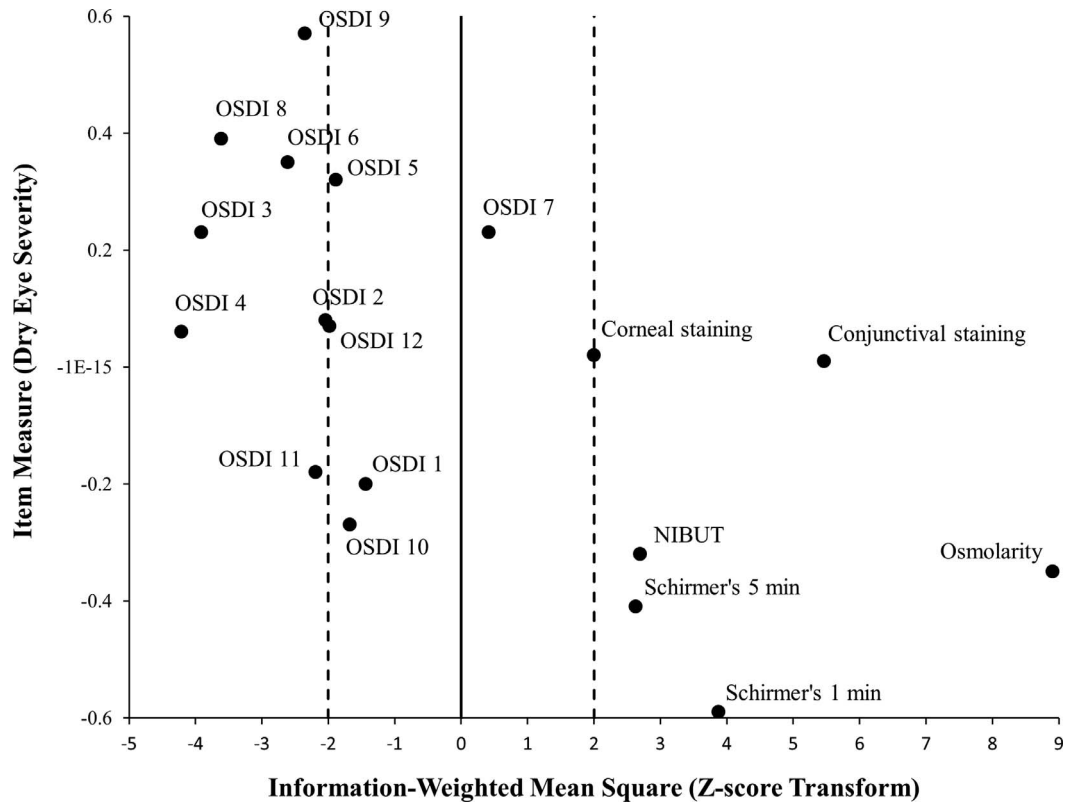


Figure 4. Scatter plot of infit mean square z-scores for each indicator (*horizontal axis*) versus the corresponding estimated indicator measure (*vertical axis*). The *solid vertical line* is the expected infit mean square and the *dashed vertical lines* define the boundaries for ± 2 SD from the expected value.

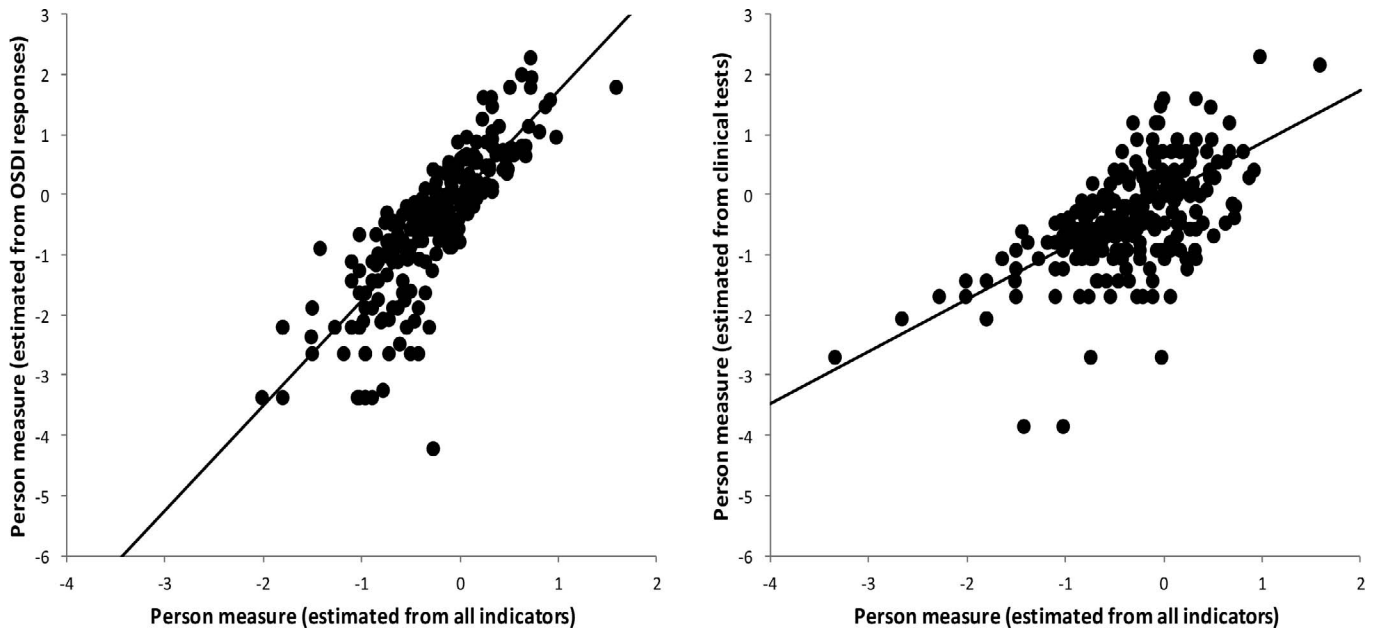


Figure 5. The scatter plots of the OSDI-based person measures and the clinical sign-based person measures.

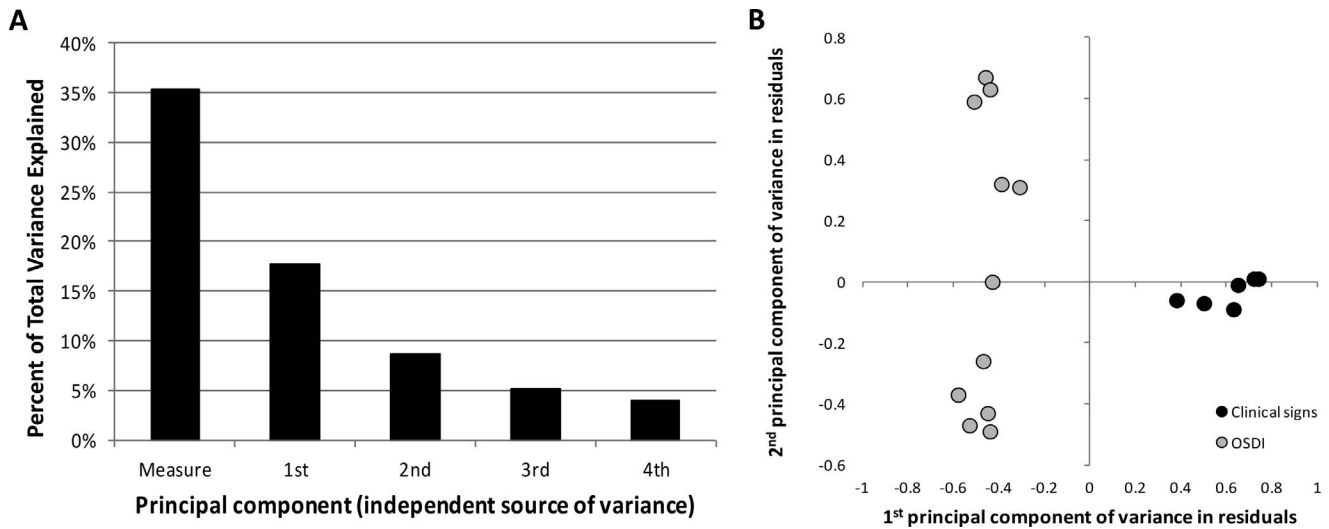


Figure 6. (A) PCA of response residuals. (B) The variance of response residuals for clinical sign indicators (black points) and OSDI items (gray points).

component. Consistent with the interpretation of the bimodal distribution of infit mean square residuals for indicators (Fig. 4), these results confirm that there are two independent sources of noise variance contributing to the distribution of response residuals.

As expected, the average severity measure for dry eye patients (-0.39 logit, $SD = 0.59$) was significantly greater than that for controls (-1.2 logit, $SD = 0.83$; 2-tailed t -test; $P < 0.001$; Fig. 7A). Again as expected, the average severity measure of patients with SS-related dry eye (-0.06 logit, $SD = 0.50$) was significantly greater than that of patients with non-SS dry eye (-0.30 logit, $SD = 0.51$; 2-tailed t -test; $P = 0.003$; Fig. 7B). After adjusting for age and sex in

linear regression models, having dry eye was significantly associated with greater person measure compared to not having dry eye (0.70 logit, 95% confidence interval [CI] = 0.51–0.89, $P < 0.001$) and having SS-related dry eye also was significantly associated with greater person measure compared to having non-SS related dry eye (0.23 logit, 95% CI = 0.70–0.39, $P = 0.004$).

Discussion

We demonstrated that dry eye signs and symptoms can work together to determine a single dry eye

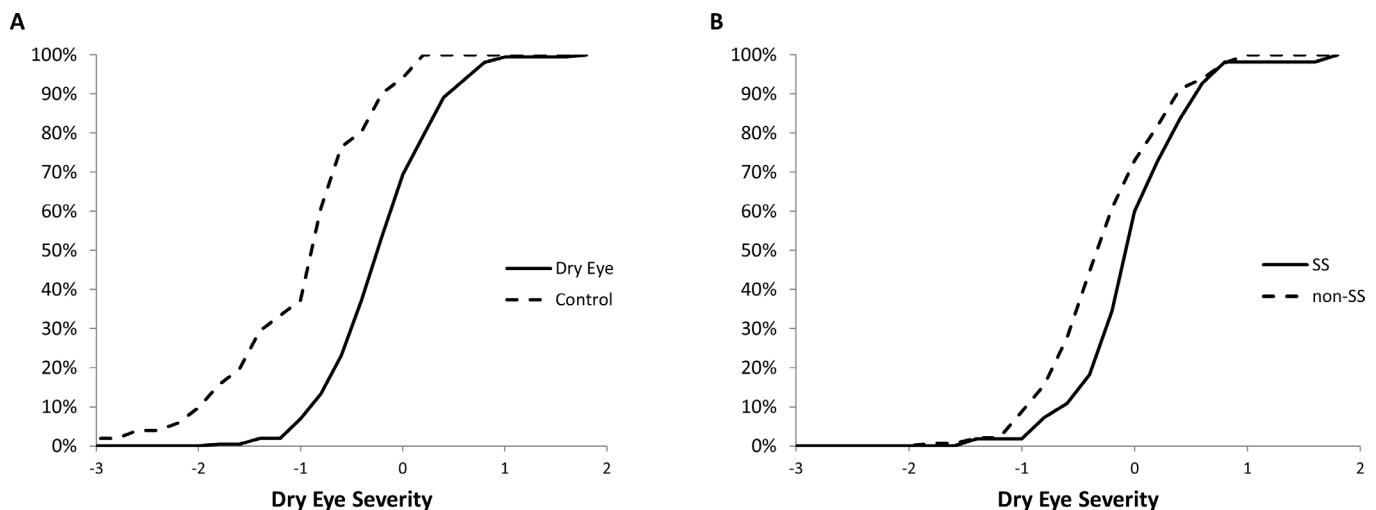


Figure 7. (A) The cumulative person measure distribution for dry eye patients (solid line) versus controls (dashed line). (B) The cumulative person measure distribution for patients with SS (solid line) to that of non-SS patients (dashed line).

severity variable that is measurable on an interval scale (Fig. 1). Previously, a formal theoretical framework was proposed for estimating and validating a latent dry eye severity measure from clinical signs and patient symptoms.¹⁷ The theory assumes that each clinical observation and symptom carries information about dry eye severity, but it is degraded and/or masked by “noise,” which gives rise to random and systematic perturbations in the observations. A similar study aiming to create a single objective dry eye severity index evaluated the use of similar dry eye tests using the original DEWS severity scale; however, usefulness of the severity index was not studied in that report.¹³ Distinctively, we treated each OSDI question as a separate item in our analysis instead of using the total OSDI score and included slightly different set of clinical tests. Additionally, we used Rasch analysis to estimate the measure, which is a unique approach that assumes the magnitude of each observed symptom and sign can be mapped to a latent dry eye severity variable. Other research studies also used the Rasch analysis to investigate the functioning of the dry eye questionnaires.^{15,30} However, the main purpose of these studies was not to determine a variable to use as a severity measure for dry eye. To our knowledge, this is the first study to combine clinical observations with the symptom questions to estimate a single severity measure using a Rasch model.

Even though the lack of correlation between dry eye signs and symptoms is well known,^{10,11} the most commonly used grading systems, such as the original DEWS severity table, require presence of severe signs and symptoms together to diagnose severe dry eye. Recently, the ODISSEY European Consensus Group also proposed an algorithm for evaluating the severity of dry eye.³¹ However, both previous algorithms are based on consensus methods without any prospective studies. In addition to the overall discordance between signs and symptoms, conflicting signs also are an issue in dry eye severity evaluation.⁵ For example, a low Schirmer’s test can be seen without any significant ocular surface staining or low tear film break-up time in the same patient. The limited use of severity tables due to lack of strong association between the features of dry eye has been noted in the TFOS DEWS II Diagnostic and Methodology report.⁵ For this reason, the TFOS DEWS II scientific committee offered a new diagnostic scheme that suggests positive symptomatology should be accompanied by significant worsening in one of the clinical signs (NIBUT,

osmolarity, or ocular surface staining) for the diagnosis of dry eye.⁵ However, we know that patients with severe debilitating symptoms with no significant clinical findings also exist. Although neuropathic pain rather than dry eye is suggested to be considered in this situation,⁵ absence of significant clinical signs may be momentary and should not exclude the diagnosis of dry eye. We previously demonstrated that evaluating tear film and ocular surface parameters at rest may miss clinical findings that can be seen in the same patient after a 30-minute reading, and baseline symptoms correlate better with signs measured after reading activity.³² This further proves that the concurrent presence of symptoms and signs should not be a requirement for dry eye diagnosis and traditional scoring algorithms remain insufficient.¹³ The severity measure that we estimated from Rasch analysis uses each information that each item provides. Additionally, our method does not ignore the less severe sign or symptom if another one indicates more severity, but rather combines all information available. We showed that no single test carries information for >10% of all items, and that corneal and conjunctival staining were the most informative, while NIBUT was the least informative (Fig. 2). However, the information value of any indicator alone was small – the 18 indicators working together provided 10 times the information of the single most informative indicator.

We noted two sources of variability for the dry eye severity measure using the 18 indicators: there was more variance than expected for clinical signs and less variance than expected for patient symptoms, which potentially challenges the internal validity of the item measures because the error variance was not homogeneous across items (Fig. 4). The two sources of error variance do not appear to represent two different dry eye variables as no evidence of DPF is illustrated by the regressions in Figure 5 and their interpretation described in the [Supplementary Material](#). The error variance is high (the measures account for only 36% of the observed variance – Fig. 6).

There appear to be two independent sources of random error variance (Fig. 6). Potential candidates are that OSDI scores result from patient judgments and clinical signs involve a variety of physical measures, which can be affected by a number of physiologic and methodologic parameters, and clinician judgments. Also, OSDI scores are person level, clinical signs are for each eye separately. Thus, differences between eyes in dry eye severity can

contribute to increased error variance among the signs. As expected if the measures are externally valid (i.e., we are measuring what we claim to measure), dry eye severity measures are significantly greater for dry eye cases than for controls (Fig. 7A) and significantly greater for dry eye cases diagnosed as SS than for cases that do not have SS (Fig. 7B).

An objective single severity measure would be simple and very useful in clinical trials, particularly to determine eligibility, and evaluate the effectiveness of the therapy. Traditional algorithms may cause misclassification of dry eye patients under some circumstances. Some dry eye patients may be missed due to conflicting signs and symptoms. However, these patients exist and they ought to be included in the studies as well. In addition, when evaluating the effectiveness of any therapy, it is difficult to document the progress especially in case of conflicting signs. A therapy can be effective overall; however, this effect cannot be shown adequately if only one of the clinical signs is expected to be improved. Alternatively, the therapy might reduce all of the clinical signs as determined by the physician, but if patient symptoms do not improve, therapy can be deemed ineffective. The difficulties in clinical trials to evaluate the effectiveness of therapies for dry eye has been discussed extensively.^{6,12,33–35} To obtain approval for a dry eye therapy in the United States, it is expected that signs and symptoms are improved by the suggested therapy in clinical trials. The multifactorial nature of dry eye, variability of signs and symptoms, and lack of widely accepted guidelines for the approval of a therapy for dry eye are the main drawbacks resulting in failure of many clinical trials leading to only a few approved treatment options.^{6,12,33,34}

This report demonstrated that currently available clinical tests for dry eye are insufficient due to high variability. Although the tests were done at 10-minute intervals in our study, the order of tests may have had an effect on variability, such as performing NIBUT after Schirmer's test. More informative tests with less variations are warranted. As new clinical tests are proposed, this method also can be used as a tool to assess their use. In fact, including more tests, such as tests evaluating lipid layer of tear film, meibomian glands, or neurosensory abnormalities, may yield more information on severity of dry eye. This approach currently is limited to be used in clinical trials or research studies as it requires complex calculations to estimate a severity measure for each person. The next level of the study will be including a

large sample from multiple centers to create a large enough database to explore the dimensionality of the intrinsic variance. With doing so, we will be able to calibrate item measures for different clinical tests, and an application or a web-based calculator then could be used by clinicians to transform test results to the single dry eye severity measure, which in turn could be used to assess the severity and/or treatment response or progress of dry eye by clinicians in daily practice.

In conclusion, our study indicated that standard dry eye clinical signs and symptoms can work together to define and measure a latent dry disease severity variable. There is no single “best” dry eye severity measure; rather, the most information about dry eye severity is carried by the battery of clinical indicators. We believe that the dry eye severity measure that we estimated from Rasch analysis would be very helpful to overcome most of the challenges of clinical trials.

Acknowledgments

Supported in part by research grants provided by the King Khaled Eye Specialist Hospital (KKESH) and Jerome L. Greene Sjögren's Center, Johns Hopkins University. The osmolarity cards were donated by Tearlab (TearLab Corporation, Inc., San Diego, CA).

Presented at the annual meeting of the Association for Research in Vision and Ophthalmology (ARVO), Denver, CO, May 3–7, 2015.

Disclosure: **S. Karakus**, None; **E.K. Akpek**, None; **D. Agrawal**, None; **R.W. Massof**, None

References

1. Schaumberg DA, Sullivan DA, Buring JE, Dana MR. Prevalence of dry eye syndrome among US women. *Am J Ophthalmol.* 2003;136:318–326.
2. Chia EM, Mitchell P, Rochtchina E, Lee AJ, Maroun R, Wang JJ. Prevalence and associations of dry eye syndrome in an older population: the Blue Mountains Eye Study. *Clin Experiment Ophthalmol.* 2003;31:229–232.
3. Schaumberg DA, Dana R, Buring JE, Sullivan DA. Prevalence of dry eye disease among US men: estimates from the Physicians' Health Studies. *Arch Ophthalmol.* 2009;127:763–768.

4. Paulsen AJ, Cruickshanks KJ, Fischer ME, et al. Dry eye in the beaver dam offspring study: prevalence, risk factors, and health-related quality of life. *Am J Ophthalmol.* 2014;157:799–806.
5. Wolffsohn JS, Arita R, Chalmers R, et al. TFOS DEWS II Diagnostic Methodology report. *Ocul Surf.* 2017;15:539–574.
6. Savini G, Prabhawasat P, Kojima T, Grueterich M, Espana E, Goto E. The challenge of dry eye diagnosis. *Clin Ophthalmol.* 2008;2:31–55.
7. Grubbs JR Jr, Tolleson-Rinehart S, Huynh K, Davis RM. A review of quality of life measures in dry eye questionnaires. *Cornea.* 2014;33:215–218.
8. Schiffman RM, Christianson MD, Jacobsen G, et al. Reliability and validity of the Ocular Surface Disease Index. *Arch Ophthalmol.* 2000;118:615–621.
9. Ozcura F, Aydin S, Helvaci MR. Ocular Surface Disease Index for the diagnosis of dry eye syndrome. *Ocul Immunol Inflamm.* 2007;15:389–393.
10. Nichols KK, Nichols JJ, Mitchell GL. The lack of association between signs and symptoms in patients with dry eye disease. *Cornea.* 2004;23:762–770.
11. Bartlett JD, Keith MS, Sudharshan L, Snedecor SJ. Associations between signs and symptoms of dry eye disease: a systematic review. *Clin Ophthalmol.* 2015;9:1719–1730.
12. Sullivan DA, Hammitt KM, Schaumberg DA, et al. Report of the TFOS/ARVO Symposium on global treatments for dry eye disease: an unmet need. *Ocul Surf.* 2012;10:108–116.
13. Sullivan BD, Whitmer D, Nichols KK, et al. An objective approach to dry eye disease severity. *Invest Ophthalmol Vis Sci.* 2010;51:6125–6130.
14. Behrens A, Doyle JJ, Stern L, et al. Dysfunctional tear syndrome: a Delphi approach to treatment recommendations. *Cornea.* 2006;25:900–907.
15. Dougherty BE, Nichols JJ, Nichols KK. Rasch analysis of the Ocular Surface Disease Index (OSDI). *Invest Ophthalmol Vis Sci.* 2011;52:8630–8635.
16. Suzuki M, Massingale ML, Ye F, et al. Tear osmolarity as a biomarker for dry eye disease severity. *Invest Ophthalmol Vis Sci.* 2010;51:4557–4561.
17. Massof RW, McDonnell PJ. Latent dry eye disease state variable. *Invest Ophthalmol Vis Sci.* 2012;53:1905–1916.
18. Masters GN. A Rasch model for partial credit scoring. *Psychometrika.* 1982;47:149–174.
19. Shiboski SC, Shiboski CH, Criswell L, et al; Sjögren's International Collaborative Clinical Alliance (SICCA) Research Groups. American College of Rheumatology classification criteria for Sjögren's syndrome: a data-driven, expert consensus approach in the Sjögren's International Collaborative Clinical Alliance cohort. *Arth Care Res (Hoboken).* 2012;64:475–487.
20. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the Ocular Surface Disease Index. *Arch Ophthalmol.* 2000;118:615–621.
21. Witcher JP, Shiboski CH, Shiboski SC, et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjögren's Syndrome International Registry. *Am J Ophthalmol.* 2010;149:405–415.
22. Gumus K, Crockett CH, Rao K, et al. Noninvasive assessment of tear stability with the tear stability analysis system in tear dysfunction patients. *Invest Ophthalmol Vis Sci.* 2011;52:456–461.
23. Linacre M. Winsteps Rasch Tutorial 3, 2012. Available at: <https://www.winsteps.com/a/winsteps-tutorial-3.pdf>. Accessed on July 6, 2018.
24. Partial credit Rasch model: Available at: <http://www.winsteps.com/winman/partialcreditmodel.htm>. Accessed on September 27, 2017.
25. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. *J Appl Meas.* 2003;4:153–163.
26. Massof RW. Understanding Rasch and item response theory models: applications to the estimation and validation of interval latent trait measures from responses to rating scale questionnaires. *Ophthalmic Epidemiol.* 2011;18:1–19.
27. Wilson EB, Hilferty MM. The distribution of chi-square. *Proc Natl Acad Sci U S A.* 1931;17:684–688.
28. Akpek EK, Klimava A, Thorne JE, Martin D, Lekhanont K, Ostrovsky A. Evaluation of patients with dry eye for presence of underlying Sjögren syndrome. *Cornea.* 2009;28:493–497.
29. Johanson G, Alsmadi A. Differential person functioning. *Educ PsychMeasure.* 2002;62:435–443.
30. Gothwal VK, Pesudovs K, Wright TA, McMonnies CW. McMonnies questionnaire: enhancing screening for dry eye syndromes with Rasch analysis. *Invest Ophthalmol Vis Sci.* 2010;51:1401–1407.
31. Baudouin C, Aragona P, Van Setten G, et al; ODISSEY European Consensus Group members. Diagnosing the severity of dry eye: a clear and practical algorithm. *Br J Ophthalmol.* 2014;98:1168–1176.

32. Karakus S, Agrawal D, Hindman HB, et al. Effects of prolonged reading on dry eye. *Ophthalmology*. 2018;125:1500–1505.
33. Foulks GN. Challenges and pitfalls in clinical trials of treatments for dry eye. *Ocul Surf*. 2003;1:20–30.
34. Novack GD, Asbell P, Barabino S, et al. TFOS DEWS II Clinical Trial Design Report. *Ocul Surf*. 2017;15:629–649.
35. Alves M, Fonseca EC, Alves MF, et al. Dry eye disease treatment: a systematic review of published trials and a critical appraisal of therapeutic strategies. *Ocul Surf*. 2013;11:181–192.