Determination of the Minimal Clinically Important Difference of the University of North Carolina Dry Eye Management Scale

Christopher J. Hwang, MD, MPH, Reeves Ellis, MD, MPH,† Richard M. Davis, MD,* and Sue Tolleson-Rinehart, PhD‡*

Purpose: To establish an initial estimation of the MCID of the University of North Carolina Dry Eye Management Scale (UNC DEMS) and assess its association with patient perceptions of symptom change.

Methods: Thirty-three patients (33.3% men, 67.7% women, mean age 60.5 yrs) with previous DEMS scores were recruited from a UNC ophthalmology clinic in spring 2014. We used anchor-based methods, categorizing important symptom change, to compare the change in the DEMS scores across visits to patient assessments of change; linear regression coefficients estimated the MCID. We correlated clinical assessments, patient perceptions, and DEMS scores.

Results: DEMS score changes correlated with global anchors [−0.4229 (P = 0.014)]. Unadjusted linear regression yielded a beta coefficient of −0.54 (confidence interval, −0.97 to −0.12, R^2 = 0.18, P = 0.014), which estimated the DEMS MCID. Adjusting the regression model for days since the last visit and DEMS score improved the association (beta = −0.56; confidence interval, −0.99 to −0.13; R^2 = 0.43; P = 0.013). Descriptive statistics produced an MCID of 1 point. Patients said that 2 points would represent a significant change. The DEMS modestly correlated with the Schirmer test (−0.4045, P = 0.0266), Oxford Grading Scheme (+0.3713, P = 0.0364), and tear breakup time (−0.3559, P = 0.0456).

Conclusions: The UNC DEMS is a valid, responsive patient-reported outcome measure instrument, which is easy to use in the clinic and capable of showing an MCID of 1 point.

Received for publication January 14, 2017; revision received May 19, 2017; accepted May 19, 2017. Published online ahead of print July 13, 2017.

From the *Department of Ophthalmology, University of North Carolina, Chapel Hill, NC; †Department of Ophthalmology, LSU/Oschner Health Center, Jefferson, LA; and ‡Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC.

Supported by a pilot grant from the NC TraCS Institute, with support from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number UL1TR001111. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The authors have no conflicts of interest to disclose.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.corneajrnl.com).

Reprints: Christopher Hwang, MD, MPH, 5151 Bioinformatics Bldg., CB#7040, Chapel Hill, NC 27599 (e-mail: christopher.hwang@unchealth.unc.edu).

Key Words: dry eye disease, quality of life measures, patient-reported outcome instrument development, responsiveness testing

(*Cornea* 2017;36:1054–1060)

Importance of patient-reported outcome measures (PROs) is reflected in growing policy guidance from the Food and Drug Administration (FDA) and provisions in the Affordable Care Act such as the creation of the Patient-Centered Outcomes Research Institute.[1–4] The necessity of developing valid PROs for clinical trials and health care has led to active instrument development. The field of ophthalmology is no exception.

One common chronic ocular disease responsible for a large proportion of eye clinic visits each year and a substantial burden to patients is dry eye disease (DED). DED affects millions of adults over age 50 and is associated with symptoms such as pain, burning, grittiness, tearing, and light sensitivity; these symptoms harm patients' quality of life.[5–9] Time trade-off methods of estimating the burden of disease have shown that moderate to severe DED is believed to be burdensome as, or more than, dialysis, severe angina, and hip fractures.[10] Many studies have found a relatively poor correlation between clinical assessment of disease severity and patient-reported symptoms.[11] Effective monitoring of disease severity and the appropriateness of treatments could benefit from a measurable patient-reported outcome that can be followed over time, and, in response to this need, many questionnaires that aim to evaluate relevant domains within DED have been developed. The clinical utility of many of these questionnaires is still uncertain.[12]

Over 20 dry eye symptom questionnaires are documented in the current literature, but a recent review identified only 6 capable of assessing quality of life.[13] Two instruments, the Ocular Surface Disease Index (OSDI) and the Instrument of Dry Eye in Everyday Life (IDEEL), have undergone robust validity and reliability studies and have been used in clinical trials.[14–18] The most recent version of the OSDI is a 12-item questionnaire assessing 3 domains: ocular symptoms, vision-related function, and environmental triggers.[14,19] The questionnaire is scored on a scale of 0 to 100 and requires the use of an algorithm and a scale to determine a patient's DED severity. The IDEEL is a 57-item questionnaire developed to assess 3 domains: "dry eye symptom-bother, dry eye impact

on daily life…and dry eye treatment satisfaction".[16(p1)] The length and multistep interpretation of scores may make both instruments challenging for use within the context of a busy eye clinic.

Our colleagues at The University of North Carolina (UNC) developed the UNC Dry Eye Management Scale (DEMS) to meet the need for a valid, reliable, easy-to-administer, and easy-to-interpret instrument that can assess both patient-reported symptoms and their effects on daily life. The UNC DEMS is a single-item instrument that asks patients to rate their symptoms and the effects of those symptoms on daily life on a scale of 1 to 10 over the past 2 weeks. The instrument was created for use in the clinical setting, with emphasis on the importance of ease of use and quick interpretability. The scale, as it is presented to patients, can be found in Supplemental Digital Content 1 (see http://www.med.unc.edu/ophth/files/FINALDEMSwithCopyright11.pdf).

The simple tool includes brief explanations for patients to help them discriminate between mild, moderate, and severe symptoms. Its creators used PROMIS methods for development of patient-reported outcomes (PROs) instruments, and, in recent work, the UNC DEMS has been shown to be both valid and reliable, strongly correlating with scores generated by the current gold standard measure, the OSDI.[2,20]

A PRO tool, however, is only as good as is its ability to detect a meaningful change over time, or its responsiveness. Groups like the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and the FDA have recognized the importance of developing and choosing appropriate PRO tools for research.[1,3,21–24] The FDA recommends determining "a score change in a measure, experienced by an individual patient over a predetermined time period that has been demonstrated in the target population to have a significant treatment benefit," often referred to as the minimal clinically important difference (MCID).[23] In response to this recommendation, we take the next validation step for the UNC DEMS—determining its MCID.

METHODS

Patient Recruitment and Participation

At the time of recruitment, a single ophthalmologist regularly supervised the use of the UNC DEMS to assess patients with dry eye at the UNC Ophthalmology Cornea outpatient clinics. We recruited all 33 patients for this study from this one attending physician during May and June 2014. To be included in the study, patients had to have been 18 years of age or older, have a diagnosis of DED coded *ICD-9* 375.15—tear film insufficiency (at the time of the study, *ICD-10* codes were not yet in use), and have at least one previously documented DEMS score. We did not exclude patients based on etiology of DED. Because the DEMS had only been validated in English at the time, we recruited only English-speaking patients to the study.

After UNC institutional review board approval, we began to identify eligible participants at regularly scheduled clinic appointments and those consenting to participate. We collected patients' age and sex, the date and score of their most recently recorded DEMS score, and, if available, previous clinical assessments of DED including the Schirmer tear production test, tear breakup time (TBUT), and Oxford grading scheme for the dry eye score from the corresponding visit.

Patients underwent their usual clinical assessment and care during their visit for DED, including administration of the UNC DEMS during initial workup by the technician before evaluation by physician and/or researcher. If the technician did not administer the UNC DEMS, then student researchers would administer the DEMS before continuing with data collection. Those administering the UNC DEMS used the following standard administration instructions, used at every administration of the UNC DEMS. Those instructions including reading this sentence to patients: "Using the examples below (pointing to descriptors of eye symptoms with increasing severity) think about your symptoms and how they affect your daily life over the past week." The UNC DEMS was then recorded into the medical record for that visit. The UNC DEMS questionnaire can be found in Supplemental Digital Content 1 (see http://links.lww.com/ICO/A533).

Clinicians, or researchers with clinician supervision and confirmation, also conducted routine assessments for tear production and tear film quality, as well as an evaluation of the ocular surface. These evaluations were made using the Schirmer 1 test, evaluation of TBUT, and ocular surface evaluation through fluorescein staining using the Oxford grading scheme for dry eye.[25,26] To measure TBUT, fluorescein dye was instilled into the eye using BioGlo (HUB Pharmaceuticals, LLC. Rancho Cucamonga, CA) fluorescein strips according to supplied product instructions. Strips were moistened with 1 to 2 drops of sterile irrigating solution to saturate the BioGlo-impregnated portion. Excess moisture was shaken off before instilling the fluorescein into the eye. The tip of the fluorescein strip was touched to the lower palpebral conjunctiva, and the patient was instructed to blink several times before performing the evaluation. This technique avoids instilling excess dye into the eye and avoids artificially raising tear volume before evaluation. TBUT was measured in seconds, counted by the same provider as time from blink to first observed disruption of a uniform fluorescein tear film. Patients' participation in this study did not influence the clinical or therapeutic course of disease management.

Dry Eye Symptom Change Questionnaire

We followed best validation practices by developing a single-item, Likert scale questionnaire to serve as an anchor for within-patient global transition assessment, or what we call the global change assessment (GCA).[23,27–29] The questionnaire, called the Dry Eye Symptom Change Questionnaire (DESCQ), asks patients, "compared to your last visit, how are your dry eye symptoms now?" Patients responded by choosing "much worse, somewhat worse, a little worse, the same, a little better, somewhat better, or much better." The DESCQ also includes questions about patients' perceptions of symptom change and their current therapeutic and/or behavioral methods (medication adherence, avoidance of triggers of dry eye) for managing their disease.

To serve as an additional reference for measuring patients' beliefs on the symptom change as they relate to how patients use the DEMS score, we also asked patients the following questions:

1. "How many points would your score have to change to show that you felt like your symptoms were getting better? That is, how much change would be a meaningful improvement in quality of life for you?"
2. "How many points would your score have to change to show that you felt like your symptoms were getting worse? That is, how much change would mean that your symptoms were making your quality of life worse? "
3. "If you could choose a number on the UNC DEMS that would be your goal score for treatment of your dry eyes —the place you'd like to get to—what would that number be?"

The DESCQ questionnaire in its final administered version can be found in Supplemental Digital Content 2 (see http://links.lww.com/ICO/A534).

Statistical Analysis

To be included in the analysis, participants must have had a previously recorded DEMS score and must have met all other inclusion criteria. Patients were administered the DEMS first, by their technician, and were told their DEMS score, before they interacted with either their ophthalmologist or the student researchers. If the technician had not administered the DEMS, student researchers did so before administering the DESCQ. In all cases, patients knew their DEMS score for that visit, knew it before they answered DESCQ questions, and answered the DESCQ questions second, after having answered the DEMS. To avoid creating recall biases, patients were not informed of their most recent previous DEMS score before administering both DEMS and DESCQ. We used only the most recent score for patients with multiple previous DEMS scores. For those reporting a range of numbers (eg, "1–2 points") for any question, we used the average of the range reported for the statistical analysis (eg, "1–2 points" was considered as "1.5" for data analysis). We initially scoured the data for missing values, which were replaced with the last known value carried forward. If no known previous values were available, the missing entry was omitted from the analysis.

We compared previous DEMS and current DEMS scores to determine difference in severity, from a low of 1 to a high of 10. We calculated differences by subtracting previous from current scores, meaning that a negative value signifies improvement, and vice versa. This scoring choice dictated that we would base the DESCQ on a Likert scale centered at numerical "0" to serve as anchors for patient-reported GCA. The responses "much worse, somewhat worse, a little worse, the same, a little better, somewhat better, and much better" corresponded to an assigned numerical value of "−3, −2, −1, 0, 1, 2, or 3," respectively. This scoring enabled us to assess a relationship between the patient-reported symptom change (GCA) over time and the actual change in the DEMS score.

We recorded clinical evaluations of disease severity using the Schirmer I test, TBUT, and Oxford grading scheme for each eye individually. Scores for the left and right eyes for each test were averaged to be used as a single variable for statistical analysis. Rather than converting test values into arbitrary categories of normal or gradations of severity, we decided to treat each test as a continuous variable for correlation analysis.

The primary outcomes of this study were the change in the DEMS score and the patient-reported GCA. Secondary outcome relationships included the association of scores with age, sex, number of days since the last visit, DEMS score at the last visit, Schirmer test score, TBUT, Oxford score, patient responses to perceptions of smallest incremental improvement/worsening in the DEMS score, and the goal DEMS score.

Because our participants are being treated for their disease, we expected most to report improvement over time or at least no worsening. To account for the likely small sample sizes within the "worsening" GCA categories and to conform to the strategies used to establish MCIDs for OSDI and IDEEL, we "folded" data for statistical comparisons by using the magnitude of change for each incremental category, regardless of the direction.[15,17] Data for "a little worse" were paired with "a little better," "somewhat worse" paired with "somewhat better," and "much worse" paired with "much better" to create 3 categories of incremental change. All other analyses, including estimation of the MCID using linear regression analysis, were done using unfolded data.

Estimation of the DEMS MCID

As did Miller et al,[15] we used linear regression modeling to estimate the MCID but, unlike the OSDI investigators, who used folded data even in their regression, we did not. Linear regression analysis allows us to use the entire spectrum of change in patient-reported GCAs regardless of the direction and magnitude, thereby eliminating concern about the small sample size within subgroups. In this way, our estimation of the MCID represents the predicted change in the DEMS score as a function of the patient's real rating of both magnitude and direction of symptom change.

We used Spearman rank correlations of DEMS scores and GCAs to verify the legitimacy of the anchors. We correlated all other independent variables (age, sex, days since the last visit, previous DEMS score, and the change in the DEMS score since the last visit) with DEMS and the GCA to check for collinearity before allowing them to weight the MCID estimation and to characterize their unadjusted associations. Our first multiple regression entered all independent variables simultaneously, after which we conducted stepwise regression, dropping all variables without a statistically (α = 0.05) or clinically (>10% change) significant relationship with the change in the DEMS score. We fitted the coefficients from the final, reduced model to the data to estimate the MCID.

Correlation of DEMS With Clinical Findings

The likelihood of a nonnormal distribution of data based on DEMS scores and patient-reported GCA directed us to use Spearman rank correlations to assess the relationship between patient-reported disease severity using the DEMS

score and clinical assessments of disease severity using the Schirmer test, TBUT, and Oxford scores. Statistical significance is achieved at $\alpha = 0.05$.

RESULTS

Population Demographics

Most of the 33 participants were women (67.7%), and the average age of all participants was 60.5 years. On average, patients went 122.5 days between visits and had an average previous DEMS score of 5.3 and a current DEMS score of 4.6. Complete data were available for 30 of the 33 patients (90.9%). Those with missing data did not have last values to carry forward and had either missing all 3 clinical assessment tests (1 patient) or deferred the Schirmer test (2 patients). No patients had missing DEMS scores, and all fully completed the DESCQ.

Estimation of the MCID: Linear Regression Analysis

Figure 1 is a plot of the actual DEMS score change against how patients reported that their symptoms had changed since their last visit. An unadjusted linear regression model yielded a statistically significant relationship ($R^2 = 0.18$, $P = 0.014$) with a beta coefficient of -0.54 (confidence interval, -0.97 to -0.12), which represents our estimation of the MCID. The DEMS score change was statistically significantly correlated with the number of days since the last visit ($r = 0.5069$, $P = 0.003$) and the last DEMS score ($r = -0.3741$, $P = 0.032$), as well as with the GCA ($r = -0.4229$, $P = 0.014$). Age and sex were not correlated with the change in the DEMS score. To control for these associations, we performed a linear regression analysis adjusted for both the number of days since the last visit and the previous DEMS score. The adjusted regression yielded a similar beta coefficient value and a greater strength of association (beta = -0.56; confidence interval, -0.99 to -0.13; $R^2 = 0.43$; $P = 0.013$). Based on our linear regression analysis, we estimate that the MCID for the UNC DEMS is approximately one half (½) a point on a 10-point scale, making the UNC DEMS capable of giving both patients and clinicians a fine-tuned indication of improving or worsening disease.

Estimation of the MCID—Descriptive Statistics for GCA Anchors

The majority (28 of 33) of patients reported that they felt their symptoms were the same or better; only 4 patients reported worsening of symptoms since their last visit, and none reported that their symptoms were "much worse." Table 1 shows the average score change using folded data comparing patient-reported GCAs. The literature has described using the mean score change for the smallest GCA anchor as an arbitrary method for determining the MCID.[27–29] The average actual change in the score for those rating their symptom change to be "a little better/worse" was 1.09 (n = 11). By this method, the MCID would be approximately 1 point on the DEMS scale. For the next incremental change rating of "somewhat better/worse," the mean score change was 1.81 (n = 8), which shows an increase in the magnitude of score change of almost 1 additional point paired with greater improvement/worsening of symptoms. However, the largest change in the symptom category "much better/worse" had a mean score change of only 1.00 (n = 6). Although the trends in the score change seem to be consistent

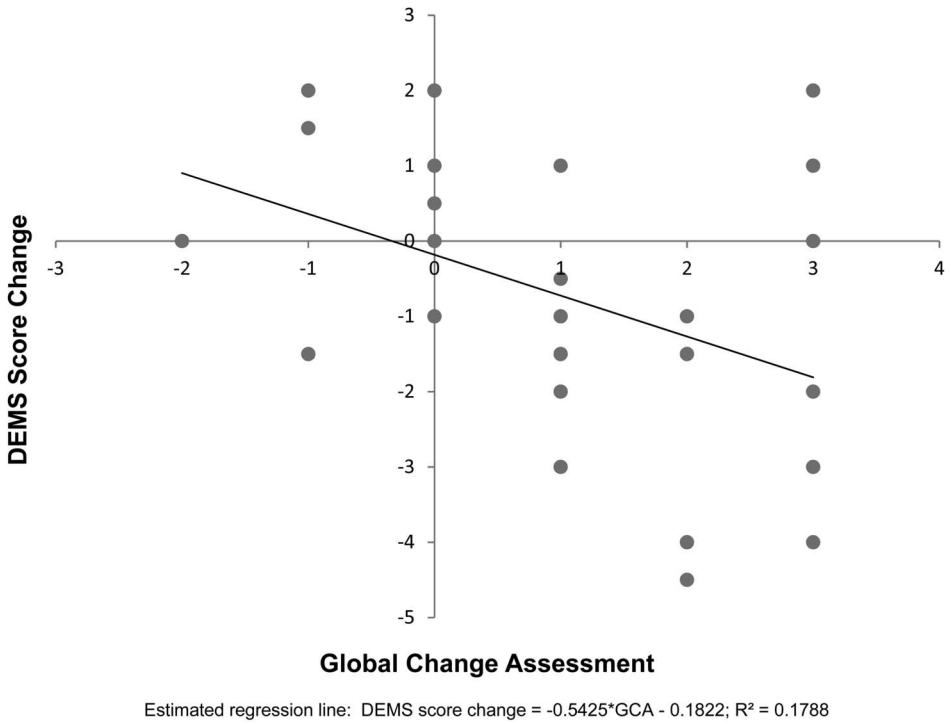

FIGURE 1. Linear regression analysis estimation of the MCID using patients' GCAs and their associated changes in the DEMS score; slope of linear regression estimates the MCID. SOURCE: data collected by first author for the UNC DEMS MCID study.

TABLE 1. Mean DEMS Score Change by Patient-Reported GCA

Response	No.	Mean (SD)
The same	8	+0.44 (1.05)
A little better/worse	11	−1.09 (1.32)
Somewhat better/worse	8	−1.81 (1.58)
Much better/worse	6	−1.00 (2.37)

with an expected increase in the magnitude of change with each incremental rating of the symptom change, the smaller sample size for this category likely limits the certainty of our findings.

Patient Perceptions of Symptom and Score Change—Responses to DESCQ

Patients' own estimations of a meaningful score change in either direction were somewhat different. The modal changes in either direction were between 1 and 2 points, capturing 27/33 participants' views of an improved score, and 22/33 participants' views of worsening symptoms. Two participants felt that only a score change of 4 would mean real improvement, but 7 participants did not think their symptoms would be meaningfully worse without a change of 4 points or more. Table 2 shows patients' average expectations of a meaningful change in either direction. Of note, 23 of the 33 total participants said that they wanted to reach a score of 1, and the average for the entire group was 1.4. This is an ambitious expectation, as only 1 patient in our sample reported a current DEMS score of 1.

Other Findings

The time since patients' last visit is significantly associated with the likelihood of reporting a worse DEMS score. Linear regression results produce a statistically significant trend of estimated worsening of 1.3 points for every 100 days since the last visit ($P = 0.003$) (Fig. 2). For all patients, the average time since the last visit was 122.5 days with a range of 27 to 237 days (SD = 67.0). When we dichotomized participants by the time since the last visit to this mean of ~120 days, dividing patients into those who had gone ≥120 days since the last visit (N = 14) and those who waited less than 120 days between visits (N = 19), patients with less time between visits had an average score change of −1.45 points, whereas those waiting longer had an average score change of +0.17 (Table 3). The paired t test revealed that this was a statistically significant difference of mean values ($P = 0.006$).

TABLE 2. Participant Perceptions of Smallest Meaningful Incremental Change in the DEMS Score and Reported Goal DEMS Score

Category	Mean Reported Score Change (SD, Range)
Smallest improvement	2.0 (0.95, 1–4.5)
Smallest worsening	2.3 (1.27, 1–4.5)
Goal score	1.4 (0.77, 1–3)

Those who had gone ≥120 days since the last visit had higher DEMS scores at that last visit, and had higher DEMS scores and larger changes (data not shown). Clinically, those with better managed symptoms and, thus lower DEMS scores, tend to follow up less frequently, which would be consistent with smaller improvements that are closer to goal scores over longer periods of time. These clinical pictures seem to comport with the establishment of an MCID for the DEMS.

As the literature has often shown that clinical assessments of DED can be poorly correlated with patient-reported symptoms, we also wanted to evaluate the correlation of clinical signs of disease with what patients reported through their DEMS scores.[11] We were able to achieve statistically significant Spearman rank correlations between the DEMS scores and all 3 clinical tests. The DEMS modestly correlated with the Schirmer test (−0.4045, $P = 0.0266$), the Oxford Grading Scheme (+0.3713, $P = 0.0364$), and TBUT (−0.3559, $P = 0.0456$). The direction of the correlation for each test with the DEMS confirms our predictions, and the findings further validate the utility of the DEMS as an indicator of a meaningful health outcome.

DISCUSSION

Many dry eye PRO instruments have been developed to aid physicians and patients as they attempt to manage DED. The International Dry Eye WorkShop (DEWS), a well-recognized group that produces timely publications with summaries of extensive data and literature analysis within the field of dry eye, argues that "clinically meaningful changes in questionnaire scores need to be defined".[30(p105)] We have followed such guidelines by developing and validating the UNC DEMS, a single-item PRO designed for clinical practice, and, now, its MCID.[20]

Despite our confidence in the UNC DEMS's estimated MCID, our study is not without limitations. First, the overall sample size is small, and even smaller for between-group comparisons. Second, using within-patient global transition assessments as anchors for establishing MCIDs requires relying on the patient's ability to recall events in the intervals between visits with some subtlety. Not informing patients of their previous DEMS score before administering both DEMS and DESCQ provided unbiased patient-reported data. However, by not informing them of their last DEMS score, it is unknown to which point in the past they were recalling for comparison. If patients recalled a time in the past when their symptoms were more severe rather than their most recent visit, this may explain why some patients said they were feeling "much better" yet had small or no improvement in their scores.

Third, linear regression estimates of the MCID assume that incremental changes are equivalent across the scale. For example, a 1-point change from 9 to 8 would be equivalent to a change in the score from 3 to 2. The analysis also assumes that the MCID is the same for movements in the score in either direction on the scale; that is, we assume that a 1-point worsening is equivalent to a 1-point improvement, in terms of patient perceptions of symptom change. We recognize the limitations of applying a linear regression analysis model across the entire spectrum of severity, but our small sample sizes limit our ability to determine an MCID for mild,

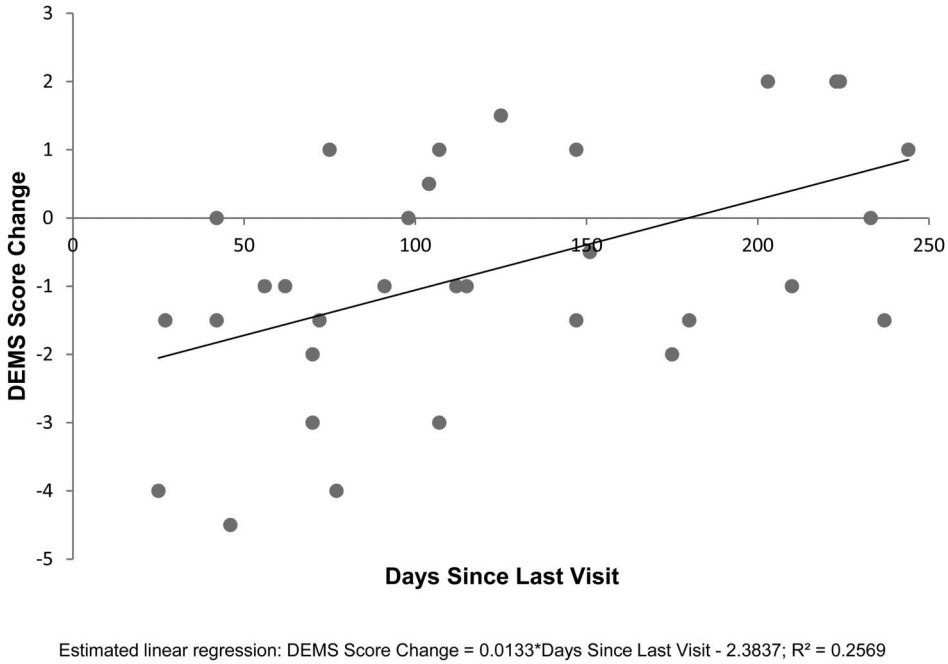

FIGURE 2. Linear regression analysis of the change in the DEMS score versus number of days since the last visit. SOURCE: data collected by first author for the UNC DEMS MCID study.

moderate, and severe dry eye based on the DEMS. The developers of the IDEEL determined a single MCID for the "symptom bother" module but recognized potential differences for those with severe dry eye.[17] Although the developers of the OSDI determined severity-specific MCIDs for mild, moderate, and severe dry eye, they also cited challenges with small sample sizes for patients with worsening symptoms and used folded data to increase sample sizes.[15] Further testing should help unravel these questions of directionality and the effect of disease severity on the MCID.

Additional considerations may contribute to these findings. On the DESCQ, patients were not specifically asked about what the smallest change in the DEMS score would be to reflect the smallest meaningful or noticeable change in symptoms. Although we would expect the answer to this question to be less than the mean of 2 points that we discovered, we also believe that asking in this manner would have been leading. Our questions allowed patients to answer in an unguided fashion and may in fact represent true patient perceptions. However, there is a possibility that some patients had difficulty understanding the meaning of the question and may have been providing responses relative to their actual DEMS score, which would bias our results toward a larger mean change. A few patients did require clarification questions, and there is no way to know whether patients fully understood how to answer the questions. Despite these uncertainties, we believe that the majority of patients understood the questions.

We also did not consider the length of time patients have been coping with their DED, which may also have some relationship with the magnitude of change from visit to visit. Those who have had DED for longer lengths of time may have reached what they have come to accept as reasonable DEMS scores and may be less likely to change from visit to visit; we may have seen these potential floor and ceiling effects among patients with higher DEMS scores but bigger improvements over time. Other limitations include not masking researchers assessing clinical severity of disease to previous/current DEMS scores.

Although there is no standard guideline for the frequency of visits for managing dry eye, clinicians tend to space visits out when a desired level of symptom control has been reached. According to our data, however, waiting for longer periods of time between visits may increase the likelihood of worsening of symptoms. But because the DEMS asks patients to consider their symptoms over the past week, a large variation in symptoms over time may make this association difficult to establish. Another study is required to determine how the length of time between visits actually affects disease management and patient-reported symptom burden.

Despite these limitations, our study has several strengths. Because we recruited patients seen by the same provider, we may have removed one possible source of patient and provider variability, and that may have contributed to our achievement of modest correlations between signs and symptoms. By recruiting patients at their normally scheduled appointments, we also had the advantage of having different follow-up times

TABLE 3. Mean DEMS Score Change by Days Since the Last Visit

Group	No.	Mean DEMS Score at Last Visit (SD, Confidence Interval)	Mean DEMS Score Change (SD, Confidence Interval)
≥120 days	14	4.4 (2.03, 3.26–5.60)	0.17 (1.49, −0.681 to 1.04)
<120 days	19	6.0 (1.68, 5.19–6.81)	−1.45 (1.64, −2.24 to −0.657)
		$P = 0.021$	$P = 0.006$

between visits. This allowed us to see the responsiveness of the DEMS in the natural clinical setting and disease course.

We are also confident that in this small test of the feasibility of establishing, the MCID demonstrates that the UNC DEMS is in fact a responsive PRO instrument that can be used to aid physicians' therapeutic strategies for DED over time. In its initial validation, the UNC DEMS demonstrated good test–retest probability with an intraclass correlation of 0.9, and we believe that this measure supports responsiveness down to 1 point.[20] We aim to validate our findings with further research using larger sample sizes. Modest, statistically significant correlations with clinical tests of disease severity also demonstrate the validity of the DEMS. We find that our identification of the mean MCID and its SD of the 10-point DEMS is similar in range, given the difference in scale, to investigators' determination of significant changes in the 100-point Eye Dryness Score (a change of 10 points) in the Lifitegrast trials (however, to our knowledge, the MCID of the EDS has not been established).[31] We have also developed and validated a Spanish version of the UNC DEMS and hope to share data on its performance in the future. Future research will involve validity and outcomes studies across multiple providers and centers for both English and Spanish versions of the DEMS. Our hope is that the UNC DEMS will help ophthalmologists and patients better manage DED.

ACKNOWLEDGMENTS

The authors thank all the participants in the study who generously supplied the data for our test of the MCID. The authors also thank the North Carolina Translational and Clinical Sciences Institute, UNC's CTSA, for its pilot grant support. Use of the UNC DEMS: The UNC DEMS is available for nonprofit clinical and research use at the following link: https://www.med.unc.edu/ophth/for-patients/clinical-specialties/unc-dry-eye-management-scale.

REFERENCES

1. Bottomley A, Jones D, Claassens L. Patient-reported outcomes: assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. *Eur J Cancer.* 2009;45:347–353.
2. Reeve BB, Burke LB, Chiang Y, et al. Enhancing measurement in health outcomes research supported by agencies within the US Department of health and Human Services. *Qual Life Res.* 2007;16(Suppl 1):175–186.
3. Revicki DA. FDA draft guidance and health-outcomes research. *Lancet.* 2007;369:540–542.
4. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *JAMA*. 2012;307:1583–1584.
5. Nichols KK. Patient-reported symptoms in dry dye disease. *Ocul Surf.* 2006;4:137–145.
6. Miljanovic B, Dana R, Sullivan DA, et al. Impact of dry eye syndrome on vision-related quality of life. *Am J Ophthalmol.* 2007;143:409–415.
7. Mertzanis P, Abetz L, Rajagopalan K, et al. The relative burden of dry eye in patients' lives: comparisons to a U.S. normative sample. *Invest Ophthalmol Vis Sci.* 2005;46:46–50.
8. Schaumberg DA, Sullivan DA, Buring JE, et al. Prevalence of dry eye syndrome among US women. *Am J Ophthalmol.* 2003;136:8–326.
9. Schaumberg DA, Dana R, Buring JE, et al. Prevalence of dry eye disease among us men. *Arch Ophthalmol.* 2013;127:763–768.
10. Schiffman RM, Walt JG, Jacobsen G, et al. Utility assessment among patients with dry eye disease. *Ophthalmology*. 2003;110:1412–1419.
11. 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.
12. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci.* 2013;90:720–744.
13. Guillemin I, Begley C, Chalmers R, et al. Appraisal of patient-reported outcome instruments available for randomized clinical trials in dry eye: revisiting the standards. *Ocul Surf.* 2012;10:84–99.
14. Schiffman RM, Christianson MD, Jacobsen G, et al. Reliability and validity of the ocular surface disease index. *Arch Ophthalmol.* 2000;118:615–621.
15. Miller KL, Walt JG, Mink DR, et al. Minimal clinically important difference for the ocular surface disease index. *Arch Ophthalmol.* 2010;128:94–101.
16. Abetz L, Rajagopalan K, Mertzanis P, et al. Development and validation of the impact of dry eye on everyday life (IDEEL) questionnaire, a patient-reported outcomes (PRO) measure for the assessment of the burden of dry eye on patients. *Health Qual Life Outcomes* 2011;9:111.
17. Fairchild CJ, Chalmers RL, Begley CG. Clinically important difference in dry eye: change in IDEEL-symptom bother. *Optom Vis Sci.* 2008;85:29–34.
18. Grubbs JR, Tolleson-Rinehart S, Huynh K, et al. A review of quality of life measures in dry eye questionnaires. *Cornea*. 2014;33:215–218.
19. Allergan. Ocular surface disease index (OSDI). [The dry eye zone website]. 1995. Available at: dryeyezone.com/encyclopedia/documents/OSDI.pdf. Accessed May 29, 2014.
20. Grubbs J, Huynh K, Tolleson-rinehart S, et al. Instrument development of the UNC dry eye management scale. *Cornea.* 2014;33:1186–1192.
21. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol.* 2006;6:2.
22. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539–549.
23. McLeod LD, Coon CD, Martin SA, et al. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11:163–169.
24. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health*. 2007;10(Suppl 2):S125–S137.
25. Norn MS. Desiccation of the precorneal film. I. Cornea wetting time. *Acta Ophthalmol.* 1969;47:865–880.
26. Bron J, Evans V, Smith J. Grading of corneal and conjunctival staining in the context of other dry eye tests. *Cornea*. 2003;22:640–650.
27. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–415.
28. Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77:371–383.
29. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res.* 2005;14:285–295.
30. Foulks GN, Lemp M. 2007 report of the international dry eye WorkShop (DEWS). *Ocul Surf.* 2007;5:1–204.
31. Holland EJ, Luchs J, Karpecki PM, et al. Lifitegrast for the treatment of Dry Eye Disease: results of a Phase III, randomized, double-masked, placebo-controlled trial (OPUS-3). *Ophthalmology*. 2017;124:53–60.