TFOS DEWS II Diagnostic Methodology report

James S. Wolffsohn, FCOptom, PhD [a, 1, *], Reiko Arita, MD, PhD [b], Robin Chalmers, OD [c],
Ali Djalilian, MD [d], Murat Dogru, MD, PhD [e], Kathy Dumbleton, MCOptom, PhD [f],
Preeya K. Gupta, MD [g], Paul Karpecki, OD [h], Sihem Lazreg, MD [i],
Heiko Pult, MSc (Optom), PhD [a, j, k], Benjamin D. Sullivan, PhD [l],
Alan Tomlinson, FCOptom, DSc [m], Louis Tong, FRCS, PhD [n], Edoardo Villani, MD [o],
Kyung Chul Yoon, MD, PhD [p], Lyndon Jones, FCOptom, PhD [q],
Jennifer P. Craig, MCOptom, PhD [r]

[a] Ophthalmic Research Group, Aston University, Birmingham, UK
[b] Department of Ophthalmology, Itoh Clinic, Saitama, Japan
[c] Clinical Trial Consultant, Atlanta, GA, USA
[d] Illinois Eye and Ear Infirmary, UIC Department of Ophthalmology & Visual Sciences, Chicago, IL, USA
[e] Department of Opthalmology, Keio University School of Medicine, Shinjukuku, Tokyo, Japan
[f] School of Optometry, University of California, Berkeley, CA, USA
[g] Cornea & Refractive Surgery, Duke Eye Center, Durham, NC, USA
[h] Kentucky Eye Institute & University of Pikeville, KY, USA
[i] Cabinet Ophtalmologie, Alger Centre, Algiers, Algeria
[j] 'Dr Heiko Pult — Optometry and Vision Research', Weinheim, Germany
[k] School of Optometry and Vision Sciences, Cardiff University, Cardiff, UK
[l] Tearlab, San Diego, CA, USA
[m] Department of Vision Sciences, Glasgow Caledonian University, Glasgow, UK
[n] Corneal and External Eye Disease, Singapore National Eye Center, Singapore
[o] Department of Clinical Sciences and Community Health, University of Milan & Eye Clinic San Giuseppe Hospital, IRCCS Multimedica, Milan, Italy
[p] Department of Ophthalmology, Chonnam National University Hospital, Gwangju, South Korea
[q] Centre for Contact Lens Research, University of Waterloo, Waterloo, Ontario, Canada
[r] New Zealand National Eye Centre, Department of Ophthalmology, The University of Auckland, Auckland, New Zealand

ARTICLE INFO

ABSTRACT

The role of the Tear Film and Ocular Surface Society (TFOS) Dry Eye Workshop (DEWS) II Diagnostic Methodology Subcommittee was 1) to identify tests used to diagnose and monitor dry eye disease (DED), 2) to identify those most appropriate to fulfil the definition of DED and its sub-classifications, 3) to propose the most appropriate order and technique to conduct these tests in a clinical setting, and 4) to provide a differential diagnosis for DED and distinguish conditions where DED is a comorbidity. Prior to diagnosis, it is important to exclude conditions that can mimic DED with the aid of triaging questions. Symptom screening with the DEQ-5 or OSDI confirms that a patient might have DED and triggers the conduct of diagnostic tests of (ideally non-invasive) breakup time, osmolarity and ocular surface staining with fluorescein and lissamine green (observing the cornea, conjunctiva and eyelid margin). Meibomian gland dysfunction, lipid thickness/dynamics and tear volume assessment and their severity allow sub-classification of DED (as predominantly evaporative or aqueous deficient) which informs the management of DED. Videos of these diagnostic and sub-classification techniques are available on the TFOS website. It is envisaged that the identification of the key tests to diagnose and monitor DED and its sub-classifications will inform future epidemiological studies and management clinical trials, improving comparability, and enabling identification of the sub-classification of DED in which different management strategies are most efficacious.

© 2017 Elsevier Inc. All rights reserved.

* Corresponding author.
E-mail address: j.s.w.wolffsohn@aston.ac.uk (J.S. Wolffsohn).
[1] Subcommittee Chair

1. Introduction

The Diagnostic Methodology Subcommittee set out to first identify tests used to diagnose and monitor dry eye disease (DED) from a comprehensive review of the academic literature, with a particular emphasis on changes since the original Tear Film and Ocular Surface Society (TFOS) Dry Eye Workshop (DEWS) [1]. Studies of test efficacy and/or performance are influenced by the fact that subjects have often been selected based on the same tests that are under scrutiny. Similarly, the performance of any "new" test may be compromised when the test is assessed in a population of DED patients who have been diagnosed using non-standardized criteria.

Secondly the committee identified those tests that are most appropriate to fulfil the definition of DED and its sub-classifications, and the most appropriate order and technique to conduct these tests in a clinical setting. The committee also identified areas in which new tests are emerging, which may influence the future of DED diagnosis and monitoring. While the original TFOS DEWS recommended categories of tests that were considered appropriate to include in DED screening, diagnosis and monitoring, as well as a series of templates to standardize these tests [2], the variety of tests in some categories precluded easy comparison of epidemiological studies or clinical trials of potential DED management techniques. In addition, the previous definition of DED from the original TFOS DEWS "Dry eye is a multifactorial disease of the tears and ocular surface that results in symptoms of discomfort, visual disturbance, and tear film instability, with potential damage to the ocular surface. It is accompanied by increased osmolarity of the tear film and inflammation of the ocular surface" identified key elements presumed to be required for the diagnosis of dry eye (symptoms of discomfort, visual disturbance, tear film instability, increased osmolarity and inflammation of the ocular surface) which might all be expected to be present (perhaps sub-clinically) [1]. However, all these aspects are rarely inclusion criteria of studies. Also the definition implied that dry eye can occur without ocular surface damage, yet staining is often listed as an inclusion criterion.

The other main aim of the Diagnostic Methodology Subcommittee was to provide a differential diagnosis rationale chart for primary DED. 'Mystery patient' studies have identified that DED is poorly recognized by non-ophthalmic health professionals, who are often consulted on self-management [3]. Hence it was deemed important to provide guidance as to the best questions to ask in order to differentiate primary DED from conditions that can mimic some characteristics of DED or cases when the dry eye is secondary to an underlying condition. Managing the underlying condition may alleviate the dry eye or change its severity and therefore its appropriate management. The chart also identifies when specialist tests and eye observation equipment are needed and, from this, determines when a referral to an appropriately equipped eye care practitioner is necessary.

2. Goals of the Diagnostic Methodology Subcommittee

The goals of the Diagnostic Methodology Subcommittee were to determine the most efficacious battery of tests for diagnosing and monitoring DED as per the revised definition, and to propose the most appropriate order and technique to conduct these tests in a clinical setting. Key diagnostic tests were to be differentiated from tests that inform subset etiologies. Recommended differential diagnostic procedures for excluding other forms of disease that may mimic some of the signs and symptoms of dry eye were also to be articulated. To be widely adopted, a diagnosis must be based on tests available in clinical practice.

3. Definition of dry eye disease (DED)

The definition of dry eye has been amended by the TFOS DEWS II to "Dry eye is a multifactorial disease of the ocular surface characterized by a loss of homeostasis of the tear film, and accompanied by ocular symptoms, in which tear film instability and hyperosmolarity, ocular surface inflammation and damage, and neurosensory abnormalities play etiological roles." [4] Hence any indication that specific signs must be present for a patient to be diagnosed with dry eye has been removed and an emphasis has been placed on the homeostasis of the tear film. Loss of homeostasis implies the body has lost the ability to maintain equilibrium, resulting in a hyperosmolar, unstable tear film with associated sequelae, e.g., increased osmolarity, inflammation, neuropathy and reduced function (compromised lubrication, hydration). Hence diagnosis requires knowledge of what is considered normal, even though this may vary with patient demographics such as sex, age and ethnicity. There are many aspects of the tear film that could be considered abnormal, such as its stability, volume, osmolarity, pH and constituents, many of which are interrelated.

4. Classification of sub-categories of dry eye disease (DED)

The TFOS DEWS II Definition and Classification report identifies that sub-categories of DED can be considered from those where the signs are predominantly evaporative (such as from a deficient lipid layer in meibomian gland dysfunction (MGD)) to those where the signs indicate aqueous deficiency (a reduced tear volume) more strongly, and the spectrum in between [4]. The severity of signs together with the evaporative-to-aqueous bias also form part of the sub-classification 'diagnosis' to aid the management of the patient's DED.

5. Diagnostic considerations

5.1. Diagnosis and monitoring

Forming an accurate clinical diagnosis is the mathematical equivalent to the problem of classification, where a multidimensional input vector of observed clinical parameters is mapped onto a discrete set of output classes, using joint probabilities and history to inform a pattern recognition algorithm. Optimal segregation of the variable space is determined by a combination of risk factors and training data. In one dimension, this concept is represented by the familiar overlapping histograms shown in Fig. 1a. True positives (TP) and false positives (FP) are represented by the portion of the affected and unaffected distributions to the right hand side of the cut-off. True negatives (TN) and false negatives (FN) are represented by the portion of the affected and unaffected distributions to the left hand side of the cut-off. In this example, the cut-off is set to achieve a high sensitivity, as defined by the ratio of true positives to the total number of affected subjects in the study. Accordingly, sensitivity = TP/(TP+FN).

In Fig. 1b, the cut-off is set to achieve a higher specificity, as defined by the ratio of true negatives to the total number of unaffected subjects in the study. Accordingly, specificity = TN/(TN+FP).

In any one dimension, sensitivity and specificity are inversely related, meaning that a more sensitive cut-off will cause a higher rate of false positives, and a more specific cut-off will cause a higher rate of false negatives.

5.2. Risk factor considerations — selecting an appropriate cut-off

The level of risk of an incorrect diagnosis generally governs the optimal cut-off for an individual sign or symptom. While there are a

variety of valid, statistical risk models that can be used to choose an optimal cut-off, for example, maximizing the ratio of true positives to false positives, receiver operator characteristic apex, etc, [5] clinical risk should supersede purely statistical methods when relying on a small number of signs or symptoms. For example, if a cataract surgeon understands the impact of an unhealthy ocular surface on biometry and visual outcomes [6–9], a more sensitive cut-off is preferable, as there is little to no safety hazard in treating a DED false positive with lubricants or other first line therapy. Conversely, the systemic costs of over-diagnosis must be considered in general practice, suggesting that a cut-off that produces equivalent risk of false positives and false negatives is more generally applicable for an individual marker. Equivalent risk results in a cut-off at the intersection between the affected and unaffected distributions if the measures of signs or symptoms are normally distributed.

5.3. Aspects of test validation

No single "gold standard" sign or symptom that correlates perfectly with the DED state has been established. If one existed, the distributions of this theoretical marker would be very similar to Fig. 1a and b, with a very small overlap in the affected and unaffected curves. Instead, there is a significant overlap between normal and DED distributions of currently available metrics, as all signs and symptoms fluctuate over time and vary significantly within different levels of disease severity [10,11]. Actual histograms are far more similar to Fig. 1c than to the idealized tests of Fig. 1a & Fig. 1b [12,13].

5.3.1. Sampling & spectrum bias

The lack of a gold standard makes it very difficult to establish true referent histograms when evaluating new diagnostic tests. The traditional approach to DED classification requires DED subjects to satisfy all criteria within a series of *sensitive* thresholds (such as Ocular Surface Disease Index (OSDI) ≥ 13, Schirmer < 10 mm/5 min, TBUT < 10 s, positive staining) and normal controls to satisfy all criteria within another, non-overlapping set (such as: OSDI < 7, Schirmer ≥ 10 mm/5 min, TBUT ≥ 10 s, negative staining) [14]. While this approach can produce strikingly high sensitivities and specificities of the diagnostic methods under evaluation, as has been done for the recent introductions of both matrix metalloproteinase-9 (MMP-9) (85% sensitivity) [14], and tear osmolarity (87% sensitivity) [15], this approach excludes a large number of DED patients, as signs and symptoms are uncorrelated across the broad population and do not move in synchrony [16–22]. For instance, it is very common to encounter a patient with a high level of symptoms and yet a lack of evidence of staining. Similarly, patients can be asymptomatic but exhibit obstructed meibomian glands, short breakup time and high osmolarity [22]. Excluding these uncategorized individuals prevents randomization across the broad population and describes sampling bias. Gaps in the inclusion criteria lead to spectrum bias, where normal patients are compared to more severe patients, to the exclusion of the mild to moderate subjects that are difficult to categorize [14]. Both sampling and spectrum bias will improve the sensitivity and specificity of a particular study, but will also increase the mean of the affected sample, shift the intersection of the two histograms to produce an unreasonably high cut-off, and result in unexpectedly poor sensitivity in the broad population. This is particularly relevant to regulatory trials, where labelled performance may not be replicated in the field, if tested on populations that are different to those included in the trial. For example, in the 510(k) summary of a new MMP-9 test, one site reported 97% sensitivity in diagnosing mild DED subjects at ≥ 40 ng/mL, while the other three sites reported 66%, 67% and 76% sensitivity (https://www.accessdata.fda.gov/cdrh_docs/pdf13/K132066.pdf). In milder populations that
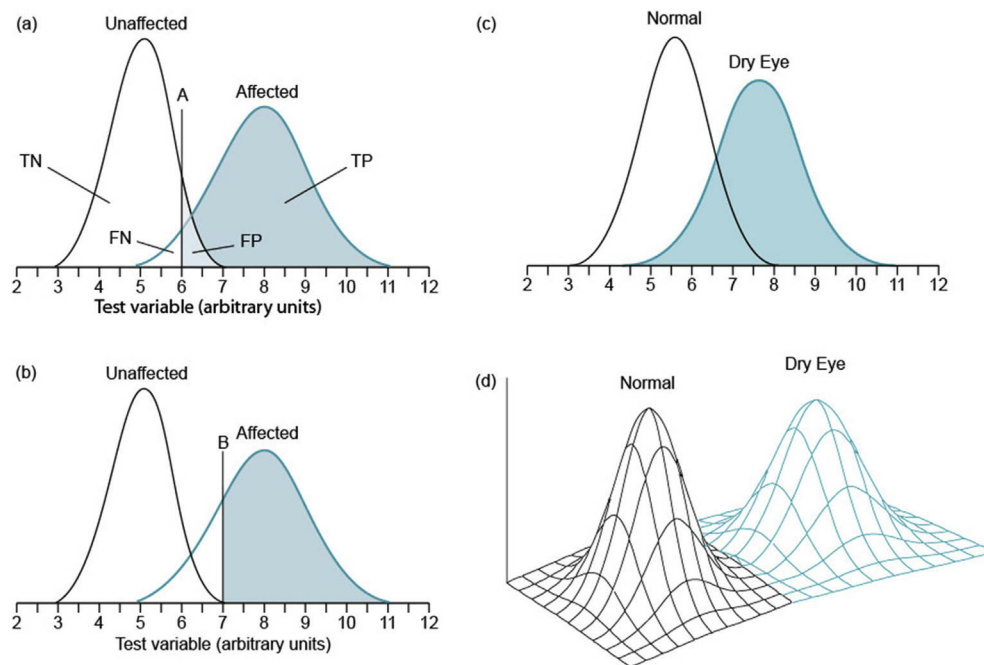
Fig. 1. Segregation of data (such as tear osmolarity or stain grade) for diagnosis concepts: a) representation by overlapping histograms. True positives (TP) and false positives (FP) are represented by the portion of the affected and unaffected distributions to the right hand side of the cut-off (A). True negatives (TN) and false negatives (FN) are represented by the portion of the affected and unaffected distributions to the left hand side of the cut-off (A). In this example, the cut-off is set to achieve a high sensitivity, as defined by the ratio of TP to the total number of affected subjects in the study (TP+FN); b) cut-off (B) set to achieve a higher specificity, as defined by the ratio of TN to the total number of unaffected subjects in the study (TN+FP); c) in reality there is significant overlap between normal and DED distributions of currently available metrics, as all signs and symptoms fluctuate over time and vary significantly within different levels of disease severity; d) additional observations (represented by multiple dimensions to the diagnostic vector), increases sensitivity and specificity simultaneously, eventually allowing clear segregation of the affected and unaffected populations at higher orders, even if there is significant overlap in lower dimensions.

did not apply the regulatory trial inclusion criteria, the 40 ng/mL cut-off demonstrated an 11% sensitivity [23]. Similarly, using an equivalent risk threshold of $\geq$312 mOsm/L applied to a broad population segregated by uncorrelated clinical signs, tear osmolarity reported a sensitivity of 73% [13], and 67% in a milder population outside the trial setting [23].

As a counterpoint in evaluating new diagnostic metrics, it is not clear whether spectrum bias is undesirable when there is no reliable gold standard to definitively diagnose DED. Without a competent benchmark for delineating affected and unaffected populations, histograms of subject populations will significantly broaden and overlap due to misclassification [11], leading to artificially low sensitivity and specificity of the new diagnostic metric under test. Because symptoms and classical DED signs are so variable over clinically relevant timescales [10,24], inclusion criteria that rely upon these metrics will result in a heterogeneous bias, impossible for even an ideal diagnostic metric to achieve good trial performance. When compared to uncorrelated inclusion criteria across the broad population, novel test sensitivities in the 40−70% range are statistically pre-determined for a single metric, regardless of how informative a test is for monitoring therapeutic efficacy or explaining mechanism of action. A good example of how subject misclassification can affect the evaluation of diagnostic metrics can be seen in Huang et al., 2012, which assessed interleukin (IL)-8 and IL-1 receptor agonist (IL-1Ra) as DED biomarkers [25]. In that study, the prospective criteria fully partitioned patients from controls in many DED measures (controls OSDI < 13, corneal staining = 0; DE1 OSDI $\geq$ 13, corneal staining < 4; DE2 OSDI $\geq$ 13, corneal staining = 4−7; and DE3 OSDI $\geq$ 13 and corneal staining >7), but resulted in sizeable overlap of IL-8 (inflammatory) and IL-1Ra (anti-inflammatory) levels between the tears of normal subjects and mild to moderate DED subjects. Conversely, a post-hoc partitioning of the patient space excluded the patients that are difficult to categorize with symptoms between OSDI 13−19 (OSDI < 13, corneal staining = 0, TBUT > 7; DE1 OSDI $\geq$ 20, corneal staining < 4 and TBUT $\leq$ 7; DE2 OSDI $\geq$ 20 corneal staining = 4−7, TBUT $\leq$ 7; and DE3 OSDI $\geq$ 20, corneal staining > 7, TBUT $\leq$ 7) that resulted in clear, significant differences between the subset of controls and mild subjects [25]. The true performance of a diagnostic metric is therefore somewhere in between the superlative performance in trials with spectrum bias and the compromised performance in trials across a broad population, using symptoms and traditional signs as inclusion criteria.

5.3.2. Selection bias

Selection bias occurs when efficacy of metrics that were used in the selection and differentiation of subjects are directly compared to a novel test that was *not* used as part of the inclusion criteria [26]. As clinical signs and symptoms are generally uncorrelated in DED, novel tests evaluated in this manner will necessarily fail. Many biomarkers (such as MMP-9, tear osmolarity, IL-1Ra, IL-8, interferon gamma-induced protein (IP)-10, S100 calcium binding protein A9) provide novel insight into disease pathogenesis [25,27−29], but because this information is unavailable from clinical observation, comparing performance of novel diagnostic metrics against the traditional signs such as staining, TBUT and symptoms will result in an apparently poor performance. This creates a paradox where, if a novel test is correlated to older metrics, it will have strong performance in a clinical trial − but there would be no need to measure the new information. Selection bias can also occur when a novel test is compared in subjects defined as having a history of DED, as these are usually based on established diagnostic tests, which the novel test is then compared to. Furthermore, trials that evaluate new markers must also prospectively align the time-courses of therapy or washout subjects, as different markers respond at different rates.

Failure to account for therapeutic timing is also a type of selection bias that artificially rewards lagging indicators if leading indicators have already responded [30].

5.3.3. Clinically important difference

In order for a new diagnostic metric to be most useful for monitoring, the marker should a) play a direct role in the pathogenesis of the disease, b) significantly improve upon treatment with an effective therapy, with the best markers traversing a large dynamic range and c) be specific to DED. Given the inherent temporal variation in all DED signs and symptoms, knowing whether a therapy has in fact altered the distribution of a sign or symptom at a single visit is non-trivial, especially with a single additional observation on follow up.

The term Minimal Clinically Important Difference was first described by Jaeschke and colleagues in 1989 as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" [31]. Some changes are found to be statistically significant, but of a magnitude too small to be noticed by a patient or to influence clinical management decisions. While clinically important differences are subjective, possibly changing with circumstances and time, they inform sample size calculations. In the field of DED research, only the Impact of Dry Eye on Everyday Living (IDEEL) − Symptom Bother scale [32] and OSDI [24] questionnaires have been assessed to determine a clinically important difference (12 and 7.0−9.9 points, respectively). The differences required in clinical studies for signs and symptoms of DED, and the resulting sample sizes needed for clinical studies using these metrics to determine a significance difference, are described in Table 1.

Another statistical approach, in the form of the reference change value (RCV), provides a confidence interval that new observations are not simply within the statistical error of the original distribution. To calculate RCV, the percentage analytical variation of a method as measured on traceable control materials (denoted CVA; or for clinical observations CVA would be intra-observer variation) and the percentage intra-individual (within-subject) variation (denoted CVI) within a population are needed.

$$\text{RCV} = 2^{1/2}Z(\text{CV}_A{}^2 + \text{CV}_i{}^2)^{1/2}$$

The Z-score for a one-sided, 95% probability that the change in sign or symptom is "very likely real" is 1.65 [48,49]. In applying an RCV, the CVI is typically derived from the literature [48]. Since most DED metrics do not have published data with thousands of subjects from which to draw reliable CVI data (unlike clinical chemistry studies for example), it is recommended that one should subtract the CVA from the CVI before calculating the RCV if the two CV measures are derived from the same study, so as to avoid double-counting. If the change in a sign or symptom between visits exceeds the calculated RCV, there is a high probability that the therapy is working. Very few papers have endeavoured to estimate the RCV of different tests for DED. Fortes et al., estimated the RCV for tear osmolarity to be 13 mOsm/L [49], although they did not correct the CVI for the CVA in the same study. The Fortes estimate would require a patient with a 338 mOsm/L reading upon initial visit to measure 325 mOsm/L or below for a clinician to believe a therapy was very likely having an effect. A corrected CVI would result in an RCV of about 10 mOsm/L. The authors are not aware of any peer-reviewed studies that reported an RCV for common clinical DED tests, but applying published longitudinal data [10], the CVI can be derived from the average and standard deviation of the subjects (n = 52), applying a zero CVA for convenience. RCVs for TBUT = 6.3 s (average over 3 months = 4.8 $\pm$ 2.7 s), 9.6/15 for

Table 1
Clinical differences to detect and resulting sample size calculation based on 2-sample *t*-test comparison with 80% power and p < 0.05 significance level http://www.statisticalsolutions.net/pssTtest_calc.php. Note that in more complex experiments, such as those requiring repeated measures analysis of variance (ANOVA), it is better to consider the number of degrees of freedom (based on both the number of treatments/visits and the number of replicates), with at least 15 recommended (such as 5 subjects being followed up 4 times during treatment [33]. As dry eye metrics often deviate from a normal distribution, it is recommended that the subject numbers are increased by 10% to compensate [34].

Test	SD of repeated measures	Healthy population mean	Clinical difference to detect	Minimum sample size per group
OSDI	6.7 on 100 point scale [35]	9.6 ± 12.2 [35] 7.8 ± 3.1 [24] 3.7 ± 6.9 [36]	MCID 4.5 to 7.3 mild/moderate & 7.3 to 13.4 severe disease [24]	14-35 mild-moderate; 4-14 severe
DEQ-5	Unknown	2.7 ± 3.2 [37]	6 (based on variation between severity classifications) [37]	Not possible
NIBUT	7.2 [38] 2.0 [36]	11.2 ± 6.8 [38] 10.4 ± 4.2 [36]	5 s	33 Tearscope 3 Keratograph 5m
FBUT	2.9 average of 2 repeats [39]	7.6 ± 10.4 [38] 9.1 ± 3.5 [36]	5 s	6
Lipid quality (Tearscope)	Unknown			Not possible
TMH	0.15 (slit lamp) [39] 0.05 (Keratograph 5m) [36]	0.29 ± 0.13 mm (slit lamp) [39] 0.29 ± 0.04 mm (Keratograph 4) [40] 0.27 ± 0.12 (Keratograph 5m) [36] 0.19 ± 0.02 mm (with OCT) [41] 0.34 ± 0.15 mm(with OCT) [42]	0.1 mm	36 slit lamp 4 Keratograph 5m
Bulbar redness (Efron)	0.4 [43] 0.4–0.7 [44]	No reported means- clinically normal taken as grade 0-1	0.5 grading	6–16
Staining grading (Efron)	Only weighted k/ICC [39]	No reported means- clinically normal taken as grade 0-1	Not reported	Not possible
LWE	No repeatability studies	Grade 1 (2–4 mm horizontal staining, 25–50% sagittal staining) [45]	Not reported	Not possible
LIPCOF	No repeatability studies	Grade 1	Not reported	Not possible
Schirmer Test (without anesthetic)	3.9 [39] 11.3 [46]	16.8 [46] 15.5 ± 8.7 [36]	5 mm	5–41
Phenol Red Thread test	7.5 [39] 10.0 [46]	29.0 [46]	5 mm	18–32
Osmolarity (Tearlab)	4.8 [47]	301mOsm/L [15] 299mOsmol/L [47]	5mOsm/L	15

Footnotes: OSDI = Ocular Surface Disease Index; DEQ-5 = Dry Eye Questionnaire – 5 item; NIBUT = non-invasive breakup time; FBUT = fluorescein breakup time; TMH = tear meniscus height; LIPCOF = Lid Parallel Conjunctival Folds; LWE = Lid Wiper Epitheliopathy.

Oxford corneal fluorescein staining (2.9 ± 4.1), 14.2/27 for Foulks/Bron meibomian gland grading (11.0 ± 6.1), and 55.3/100 for OSDI (34.7 ± 23.7) [10]. Like other statistical tests, the RCV should be used only as a guide and not an absolute value – the clinician still needs to take into account all available information when making a determination about therapeutic efficacy. Comparing the statistically derived RCVs to the published minimal clinically important difference suggests that the actual value is somewhere between these two approaches, less than the RCV and greater than the minimal clinically important difference.

5.3.4. Parallel testing

In order to increase sensitivity and specificity simultaneously, it is necessary to expand a diagnostic input vector to include multiple dimensions. As shown in Fig. 1d, extra observations eventually allow a clear segregation of the affected and unaffected populations at higher orders, even if there is significant overlap in lower dimensions. For example, if one wanted to classify trees based on leaf colour alone, it would be a very noisy, imprecise way to separate elm trees from oak trees. If you add in leaf shape, tree height, sap characteristics, bark texture and so forth, the task becomes more straightforward. Sensitivity is optimized in parallel testing by diagnosing disease if any one of a series of highly specific signs is measured to be abnormal [50]. Also known as a logical "OR" operator in computing, parallel sensitivity is calculated by subtracting the product of the two individual sensitivities from their sum (Sensitivity A + Sensitivity B − (Sensitivity A x Sensitivity B)), while parallel specificity is simply the product of the individual test specificities (Specificity A x Specificity B). Each additional metric will increase sensitivity, while multiplicatively reducing specificity.

Low specificity tests quickly degrade the combined specificity. Three parallel tests at 50% sensitivity and 97% specificity achieve 87.5% sensitivity and 91.3% specificity, which is far better than any one individual test. However, if the three tests had only 90% specificity, the parallel specificity would degrade to 72.9%. Therefore, when adding markers in parallel, more specific diagnostic metrics allow for greater confidence – which is somewhat paradoxical, as most clinicians judge new diagnostic metrics based on their sensitivity, not their specificity. As an example, parallel testing of multiple tear proteins has been shown to be very effective in diagnosing DED, despite each protein marker being individually quite insensitive (≈40–60% sensitive); when used in parallel as part of a panel, the combined measurements produce greater than 90% combined sensitivity and specificity [27,50,51].

5.4. Sequence of testing

As even non-invasive tests of DED require alternation of blinking or bright illumination, the sequence of testing can affect the results. It is recommended that the tests are performed from the least to the most invasive [52].

6. Recommendations of appropriate tests for diagnosis and assessment of dry eye

This section reviews the development and enhancement of diagnostic metrics of DED, particularly since the previous TFOS DEWS report. The order in which the tests are reviewed is not a reflection of their importance, nor their invasiveness. Due to the issues highlighted in Section 5 with regard to comparing the

sensitivity and specificity of tests, recommendations are based on the level of evidence combined with the invasiveness of the test and its ability to be conducted in a standard clinical setting, ideally without highly specialist instrumentation. The recommended diagnostic 'homeostatic marker' tests are the minimum data set to be collected from all patients identified by the screening questionnaire (as many patients do not elicit symptoms unless specifically asked) and in all DED clinical trials. However, additional DED metrics should be applied to identify the subtype of DED and the specific aspects (such as inflammatory markers or environmental triggers) relevant to a clinical trial.

6.1. Symptoms

As in the previous TFOS DEWS definition of DED [53], the current TFOS DEWS II definition for DED describes the presence of ocular surface symptoms and other signs of DED [4]. Although the relationship between symptoms and signs of DED is not linear and varies across individuals and types of DED [54], the ability to accurately quantify ocular surface symptoms is an important screening tool that can assist in establishing the medical necessity for additional DED evaluation. It is also critical for monitoring the progression of the condition and response to treatments. In this regard, symptom measurements are very similar to clinical signs of DED. It is therefore recommended that a validated symptom questionnaire be administered at the beginning of the patient interaction.

6.1.1. Current questionnaires

In the clinical setting, symptoms or other subjective reports are typically captured through the patient case history [55,56]. Symptoms reported during non-scripted verbal interviews are very difficult to standardize and quantify. To enhance standardization in clinical research, symptoms are typically gathered through the use of questionnaire instruments that are most often self-administered by the patient or research subject without input from the clinician or researcher. In DED, these instruments either measure ocular surface discomfort or vision symptoms associated with DED, the impact of DED on everyday function, or health-related quality of life. Table 2 gives a summary of the most frequently used DED questionnaires, their original and recent citations, and the forms of validation supported by the literature cited.

For questionnaires that are additionally intended as outcome measures for registration studies at the US Food and Drug Administration (FDA), an FDA guidance document describes a path for the development of a Patient Reported Outcome (PRO) [84]. For most DED research and clinical care, the majority of symptom tools focus primarily on the measurement of symptoms associated with DED, and these instruments, while valid, do not follow the full psychometric development plan for PROs. However, even for symptom questionnaires that are not supporting FDA claims, it is critical that they be validated for their discriminative ability. It should also be noted that validation is generally using patient groups with Sjogren Syndrome versus non-Sjogren Syndrome ADDE and/or healthy controls, with little focus on EDE. A recent thorough review by Guillemin and co-workers in 2012 covers the topic of questionnaire validation, and strengths and weaknesses of many DED questionnaires [85].

It is helpful if instruments are also shown to be reproducible and responsive to change in the DED condition. For clinicians, it can be helpful to have published diagnostic score criteria to screen patients who may need further testing. Table 2 covers these aspects of the DED instruments currently in use. Several of these questionnaires are undergoing translation for use in other populations [65,67]. New DED questionnaires are undoubtedly in development, and they can all be assessed for the features cited here.

6.1.2. Diagnostic test recommendation and technique

In general, the OSDI is the most widely used questionnaire for DED clinical trials. The OSDI measures frequency of symptoms, environmental triggers and vision related quality of life. Many other questionnaires have established concurrent validity against the OSDI in recent publications. The consensus view of the committee was to use the OSDI due to its strong establishment in the field or the DEQ-5 due to its short length and discriminative ability [37]. The continuous nature of visual analogue scales is attractive for clinical trials compared to discrete Likert-based question rating, so questionnaires such as the severity scale of the Symptoms Analysis in Dry Eye (SANDE) should be considered for repeated comfort assessment.

6.2. Visual disturbance

6.2.1. Current tests

6.2.1.1. Symptoms. A number of patient-reported outcome questionnaires have been developed which have items or subscales that assess patients' visual experiences of DED. These include:

6.2.1.1.1. Ocular Surface Disease Index (OSDI). The OSDI includes 6 questions related to visual disturbance (blurred vision, or poor vision) or visual function (problems reading, driving at night, working on a computer, or watching TV). A study showed that the DED group of 87 patients had worse OSDI composite and subscale scores for vision-related function, compared to a group of 71 patients without DED [86].

6.2.1.1.2. Dry Eye Questionnaire (DEQ-5). The DEQ has 4 questions related to visual disturbance, including the frequency of visual changes, how noticeable the visual disturbance is in the morning and at night, as well as how much the visual fluctuation bothers the respondent. Visual symptoms generally increase in intensity over the day, suggesting that open-eye conditions might affect symptom progression [58]. One study using the DEQ found that 10% of patients with non-Sjögren syndrome DED and 30% of patients with Sjögren syndrome complained of impaired vision while others reported that between 42% and 80% of patients with primary Sjögren syndrome experienced "disturbances in daily vision" [18,87,88].

6.2.1.1.3. Impact of Dry Eye on Everyday Living (IDEEL). The IDEEL questionnaire has 2 items related to visual disturbance including the extent to which a person is bothered by "blurry vision" or "sensitivity to light, glare, and/or wind". Statistically significant differences in responses to the IDEEL questionnaire scores across varying levels of DED severity have been observed [89].

6.2.1.1.4. National Eye Institute's Visual Function Questionnaire (NEI VFQ-25). The National Eye Institute's Visual Function Questionnaire (NEI VFQ-25) is a generic visual function questionnaire with seven visual domains including general vision, distance vision, peripheral vision, driving, near vision, color vision, and ocular pain. DED patients have poorer NEI VFQ-25 scores for the subscales of general health, general vision, ocular pain, short distance vision activities, long distance vision activities, vision-related social function, vision-related mental health, vision-related role difficulties, vision-related dependency, and driving [86,90].

6.2.1.1.5. Dry Eye-Related Quality-of-Life Score (DEQS). The Dry Eye-Related Quality-of-Life Score (DEQS) questionnaire developed in Japan has shown strong correlations with 4 subscales (Ocular Pain, Near Vision, Distance Vision, and Mental Health) of the NEI VFQ-25 [62].

6.2.1.1.6. Computer-vision symptom scale (CVSS17). The Computer-Vision Symptom Scale (CVSS17) is a Rasch-based linear-scale that contains 17 items exploring 15 different symptoms of computer-related visual and ocular symptoms. The CVSS17 includes a broad range of symptoms such as photophobia (items A33

Table 2
Features of a range of Dry Eye Questionnaires & supporting literature. Clinical utility of these questionnaires is summarised on Table 6 of the Epidemiology subcommittee report of TFOS DEWS II [57].

Name	Primary & Recent References	Dry Eye Screening Criteria	Type of Validation	Other Comments
Dry Eye Questionnaire (DEQ)	**Primary:** Begley et al. (2002) [58]	None	Discriminant focus ADDE	Indiana University Frequency & Intensity
5-Item Dry Eye Questionnaire (DEQ-5)	**Primary:** Chalmers et al. (2010) [37] **Recent:** Camp et al. (2015) [59] Galor et al. (2015) [60] Fernandez et al. (2013) [61]	≥6 KCS ≥12 suspect SS	Discriminant focus ADDE Subgroup Glaucoma Across post traumatic stress disorder, Depression	Indiana University Frequency & Intensity
Dry Eye-Related Quality-of-Life Score (DEQS)	**Primary**: Sakane et al. (2013) [62]	None	Content Face Psychometric Reproducibility	Frequency & Degree
Impact of Dry Eye on Everyday Life (IDEEL)	**Primary:** Abetz et al. (2011) [63] **Recent:** Fairchild et al. (2008) [32]	Mild 40–50 Moderate 51–63 Severe >64	Content Psychometric Discriminant focus ADDE Responsiveness CID = 8 Symptom Bother	Alcon Research, Ltd., MAPI Values Symptom bother only
McMonnies' Questionnaire (MQ)	**Primary:** McMonnies & Ho (1987) [64] **Recent:** Tang et al. (2016) [65]	>14.5 Dry Eye	Chinese Translation & Validation	Frequency only
Ocular Comfort Index (OCI and OCI-C)	**Primary:** Johnson & Murphy (2007) [66] **Recent:** Chao et al. (2014) [67] Golebiowski et al. (2016) [68]		Rasch scaled items Item reduction Responsiveness CID = 3 Chinese Translation & Validation MGD Female Cross-section	Frequency & Intensity
Ocular Surface Disease Index (OSDI)	**Primary:** Schiffman et al. (2000) [35] **Recent:** Amparo et al. (2015) [69] Asiedu et al. (2016) [70] Baudouin et al. (2014) [71] Finis et al., 2014) [72] Galor et al. (2015) [60] Miller et al. (2010) [73] Ogawa et al. (2013) [74]	Mild 13–22 Moderate 23-32 Severe ≥ 33	Concurrent with SANDE Concurrent with SPEED Severe ≥ 33 Concurrent with SPEED Concurrent with DEQ5 CID = 7.0–9.9 GVHD Subgroup	Allergan, Inc. Better for Research than SANDE Better for ADDE than SPEED Frequency & Intensity
Symptom Assessment in Dry Eye (SANDE)	**Primary:** Schaumberg et al. (2007) [75] **Recent:** Amparo et al. (2015) [69] Saboo et al. (2015) [76]		Concurrent with OSDI Concurrent with OSDI, NEI-VFQ	Frequency & Intensity Visual Analogue Scale Better for Clinical than OSDI
Standard Patient Evaluation of Eye Dryness (SPEED)	**Primary:** Blackie et al. (2009) [77] **Recent:** Asiedu et al. (2016) [70] Finis et al. (2014) [72]		Concurrent with OSDI Concurrent with OSDI	Frequency & Intensity Better for MGD Dry Eye
Developed for Use with Contact Lens Wearers				
Contact Lens Dry Eye Questionnaire (CLDEQ)	**Primary:** Begley et al. (2001) [78] Nichols et al. (2002) [79]	Yes Screening		Frequency & Intensity
8-Item Contact Lens Dry Eye Questionnaire (CLDEQ-8)	**Primary:** Chalmers et al. (2012) [80] **Recent:** Chalmers et al. (2016) [81]	≥12 = CLD	Discriminant Concurrent with Overall Opinion of CLs CID = 3 Responsiveness Concurrent with Overall Opinion of CLs, Eye Dryness & Eye Sensitivity	Frequency & Intensity Soft Contact Lenses
Contact Lens Impact on Quality of Life (CLIQ)	**Primary:** Pesudovs et al. (2006) [82] **Recent:** Erdurmus et al. (2009) [83]	Yes QoL Keratoconus only	Rasch scaling Across CL types	Frequency of bundled symptoms More of a contact lens related QoL questionnaire than a direct measure of symptoms

Footnotes: Abbreviations in alphabetical order: ADDE = Aqueous Deficient Dry Eye, CLD = Contact Lens Discomfort, MGD = Meibomian Gland Dysfunction, QoL = Quality of Life. CID = clinically important difference, GVHD = Graft Versus Host Disease, NEI-VFQ = National Eye Institute - Visual Function Questionnaire.

and C23) and "blinking a lot" (item A20), and has been reported to be valuable in the evaluation of computer-related visual and ocular symptoms [91].

6.2.1.2. Functional tests. Conventional distance and near visual acuity testing, employing Early Treatment Diabetic Retinopathy Study (ETDRS) and Lighthouse near vision charts, showed significant deterioration in symptomatic and asymptomatic OSD subjects, which improved temporarily with instillation of artificial tear drops [92,93]. Similar static tests that require reporting the orientation of sine wave gratings of varying contrast have also been utilised pre- and post-artificial tear instillation [94,95]. Dynamic methodologies to assess visual function in DED patients include detection of randomly located targets of differing contrast during a driving simulation [96]. Ridder et al. employed computer-generated sine-wave gratings that were presented briefly (16 msec duration), and demonstrated that DED patients exhibit a decrease in contrast sensitivity with tear film breakup [97].

Functional visual acuity (FVA) was first defined by Goto and colleagues, as functional vision for daily activities [98]. It corresponded to the visual acuity measured with the patient's habitual prescription, during 10–20 s of sustained eye opening without blinking, aided by anesthesia. To better standardize the test, a commercialized system was developed by Ishida and colleagues, with Landolt optotypes presented in one of four orientations; an increasing optotyope size is presented when a previous presentation is incorrectly identified or when there is no response within the set display time (selectable from 1 to 5 s); a decreasing size is presented when the answer is correct (SSC-350; Nidek, Gamagori, Japan) [99]. The visual maintenance ratio is the average FVA divided by the baseline visual acuity. FVA is reduced in DED patients, Sjögren syndrome and Stevens Johnson syndrome, more than in controls, due to irregularity of the ocular surface and induced higher order aberrations (HOAs), and it improves with treatment [98–101]. The application of FVA measurements in ocular surface disease (OSD) has identified a significant decline of FVA relating to decreased tear clearance in the elderly and associated with the short BUT type of DED in office workers, atopic keratoconjunctivitis, conjunctivochalasis and in elderly drivers [102–106].

6.2.1.3. Aberrations. Initial work examined the optical and visual impact of tear breakup during periods of non-blinking by quantifying vessel contrast in the fundus images and by monitoring the psychophysical contrast sensitivity and the spatial distribution of tear thickness changes by retroillumination [107,108]. Advances in wavefront aberrometers has enabled assessment of real-time changes in the ocular optics by evaluating refractive anomalies at multiple sites over time. Laser in situ keratomileusis (LASIK)-related dry eyes had greater optical aberrations due to increased tear film irregularity, compared to healthy controls [109]. Serial measurements of higher order and double pass (objective scatter) aberrations after blinking in patients with DED is associated with increased HOAs resulting, in part, from superficial punctate keratitis (SPK) overlying the optical zone [110–112].

6.2.1.4. Light scatter. Scheimpflug imaging has been used to show that the ocular forward light scattering and corneal backward light scattering from the anterior cornea are greater in dry eyes than in normal eyes and that increased corneal backward light scattering in dry eyes, at least partially, again resulted from central SPK overlying the optical zone [113].

6.2.2. Diagnostic test recommendation
Visual disturbance is currently assessed subjectively through ocular symptomology questionnaires. Until well-established

objective clinical measures of visual disturbance become widely available, there is no specific additional vision test that can be recommended by TFOS DEWS II for the diagnosis of DED. This does not preclude vision tools, that are currently under development, being used to enhance understanding of individual cases of dry eye.

6.3. Tear film stability

The Definition and Classification subcommittee of TFOS DEWS II have included "tear film instability" in their revised definition of DED [4]. Impaired tear film stability has been one of the fundamental diagnostic criteria for diagnosing abnormality of the tear film and many ways of evaluating tear film stability have been described [114].

6.3.1. Current tests
6.3.1.1. Tear film breakup time. In clinical practice, the most frequently employed test of tear film stability is the measurement of the tear film breakup time (TBUT); this is the interval of time that elapses between a complete blink and the appearance of the first break in the tear film [115,116].

6.3.1.2. Fluorescein breakup time. Sodium fluorescein may be instilled to enhance visibility of the tear film, when the test is referred to as the fluorescein breakup time (FBUT); however, fluorescein reduces the stability of the tear film and therefore the measurement may not be an accurate reflection of its status [117,118]. The fluorescein can be instilled in varying volumes and concentrations using either a micropipette, or more commonly impregnated strips [66]. Since controlling the volume instilled with strips may be difficult, the use of narrow (1 mm) strips and dry sterile applicators have been proposed [118–121]. A standardized methodology is also important and instructions are generally given to blink naturally three times and then to cease blinking until instructed [66]. The reference value for DED diagnosis when fluorescein is used ranges from a cut-off time of less than 10 s [122], to less than 5 s when smaller, more controlled volumes of fluorescein are used [123,124]. The sensitivity and specificity of the test have been reported to be 72.2% and 61.6%, respectively, in individuals with Sjögren Syndrome [88]; however, mild and moderate DED patients have a broad range of FBUT values and the diagnostic value is less certain for these DED sufferers [13,125]. A significant downfall of the measurement of FBUT is its dependence on subjective assessment by the observer and attempts have consequently been made to automate the measurement [126,127]. Despite the drawbacks of using fluorescein to assess tear film stability, FBUT still remains one of the most commonly used diagnostic tests for DED in clinical practice [128–132].

6.3.1.3. Non-invasive tear breakup time. Since tear film stability can be affected by fluorescein, temperature, humidity and air circulation, non-invasive breakup time (NIBUT) measurements have become more popular in both clinical practice and research. Many of these techniques involve the observation of the specular reflection of an illuminated grid pattern from the tear film [133], and these typically result in longer measured values of time to breakup than stability assessment techniques involving fluorescein instillation [118,134,135]. NIBUT can also be measured through observations of placido disk images that are reflected from the anterior ocular surface with many of the currently marketed corneal topography systems [136], and specific software has been developed to assess localized changes in corneal power, as an indication of surface irregularities and breakup of the tear film, with some instruments [137–140]. Automated assessment of tear film stability is also possible with specific software on

instruments such as the Keratograph (Oculus, Wetzlar, Germany), which detects and maps locations of tear breakup over time [141,142], The NIBUT recorded with automated systems was initially reported to be shorter than other subjective measurements of NIBUT, and even conventional FBUT measurements [140–142], however, a recent study described the reverse finding [143]. A standardized methodology is also important when conducting NIBUT measurements with similar instructions to blink naturally three times and then to cease blinking until instructed to blink again [66].

A different approach has been used by other groups of researchers in which high-speed videokeratoscopy is used to estimate the variance of the number of rings detected radially from the centre of the videokeratoscopic image [144–146]. The changes in this variance indicate the instability in image quality, which is directly related to the quality of the tear film, and this has been used as an estimate of the NIBUT. This technique has been further refined by Downie using the E300 corneal topographer (Medmont International Pty Ltd., Victoria, Australia) to measure Tear Film Surface Quality Break-up Time [147]. The algorithm used identifies and eliminates images with excessive movement and is able to recognize shadows arising from eyelashes.

Interferometry is also used to assess the stability of the tear film in a non-invasive manner [148]. Using this technique, the time between the blink and the first appearance of a discontinuity in the lipid layer can be measured, and instruments have been developed specifically for this purpose [38,149–152]. More recently an instrument employing interferometry has been developed, for clinical use, to measure the thickness of the lipid layer (TearScience® LipiView®, TearScience, Morrisville, NC) [77,153]; however, this cannot be used to measure the tear breakup time since only the tear film over the lower half of the cornea is analyzed and the area of initial break could occur anywhere across the cornea, and is noted frequently at the upper lid margin [154]. Instruments that do not allow the assessment of the entire area of the cornea exposed during eye opening may fail to detect areas of tear film abnormality.

The sensitivity and specificity of the NIBUT varies according to the specific technique used, with values of 82–84% sensitivity and 76–94% specificity being reported [134,142,147]. A cut-off value of less than or equal to 10 s has been reported to be indicative of DED when viewing the reflection of an illuminated grid pattern [134]; The absolute values for breakup time have been reported to be longer for non-invasive techniques, with a mean difference of 3.7 s being reported [38]; however, when breakup times are shorter, the differences between the two techniques have been reported to be of lesser magnitude [155].

6.3.1.4. Thermography. Evaporation of the tear film results in a cooling of the ocular surface [156], therefore measuring the absolute temperature and the spatial and temporal changes in temperature during the inter-blink period, can be used as an index of tear film stability. Infrared thermography is able to measure the temperature of the ocular surface in a non-invasive manner and provide an objective, quantitative output [157]. Purslow and Wolffsohn demonstrated the ocular surface temperature measured using infrared thermography is principally related to the tear film [158]. The evidence in the literature indicates that the cooling rate of the ocular surface is faster in individuals with DED than in normal eyes, which is assumed to be as a result of a greater rate of tear film evaporation [156,159–161].

Advances in instrumentation have allowed measurement of the ocular surface temperature with increasing accuracy, resolution, and speed [160–163]. Recently, thermography has been used to differentiate between DED of differing aetiologies, with the lowest

temperatures and greatest cooling rates being reported for presumed ADDE, and lower rates in dry eyes of presumed evaporative aetiology [164].

Studies have also been conducted in which ocular surface temperature and FBUT have been measured concurrently [165,166]. Su et al., demonstrated that areas of ocular surface cooling and breakup were co-localized [166], and Li et al., reported a direct relationship between FBUT and ocular surface cooling, implying that localized increases in evaporation are contributing to tear film thinning and breakup [165]. Using a customized ocular surface thermography device, a method has been demonstrated in which the exact area showing temperature reduction can be determined by analysing a series of images over a period of 9 s [167]. From this analysis, a "thermal breakup area" and "thermal breakup time" can be reported. Furthermore, the subjective sensation of discomfort has been reported to occur earlier in the interblink period in patients with DED than in controls (during forced eye opening), and that the subjective symptoms were correlated to low corneal temperatures and enhanced tear evaporation [168]. Sensitivity and specificity values of around 80% have been reported [160,161].

6.3.1.5. Osmolarity variability. An in-depth review of the evidence relating to osmolarity testing in the diagnosis of DED is provided in Section 6.5.1.1; however, it is also important to consider how spatial and temporal variations in tear osmolarity might affect tear film stability. There is greater inter-eye variability of osmolarity in DED than in normals [12,15,169], and the inter-eye differences increase with disease severity [13]. Moreover, this inter-eye variability has been shown to substantially reduce over time with successful treatment of DED [125].

While repeated measurements over a period of time have been shown to be low and stable in normal subjects, DED subjects showed relatively elevated and unstable readings [170]. This finding is termed heteroscedasticity, or increasing variation with increasing value [171]. Keech et al., further reported that the variability of tear osmolarity of normal subjects was indistinguishable from the analytical variability of measurements of a control solution of known osmolarity, suggesting that normal individuals retain an effective tear film with little variation from blink-to-blink and day-to-day [170]. In contrast, the tears of individuals with DED demonstrated increasing variation and the authors speculated that this was due to "a combination of chaotic or incomplete mixing between blinks and spatially variable tear film breakup, leading to a stochastically increased evaporation rate."

In a small study conducted by Liu et al., a link was reported between hyperosmolarity and tear instability, suggesting that transient increases in tear osmolarity may be observed under conditions of tear instability [172]. More recently, Peng et al., purported that increases in evaporation, that resulted during prolonged interblink periods or as a result of environmental factors (such as increased humidity and wind speed), drive tear film breakup, and predicted "massive" increases in osmolarity at the centre of areas of rupture of the tear film [173].

Indeed variability of osmolarity has been recommended to be a feature that clinicians should specifically be looking for when trying to identify patients with DED [13]. Sullivan advocates that between-eye differences beyond the threshold of 8 mOsm/L should be considered an indication of the loss of tear film homeostasis that occurs with DED [171].

6.3.1.6. Tear evaporation rate. An intact lipid layer may be necessary to prevent tear film evaporation [174]. The tear film evaporation rate is used as an indicator of tear film stability [175]. Evaporation of the tear film has been measured using a number of different techniques including a vapour pressure gradient [176,177],

and the velocity of relative humidity increase (resistance hygrometry) within a goggle cup placed over the eye [178–181]. Using these techniques, higher evaporation rates between blinks have been reported to be associated with poor tear film stability [148], and DED symptoms [179,182,183]. An absent, or non-confluent lipid layer has been determined to be associated with a four-fold increase in evaporation rate [148], and a two-fold increase in evaporation rate has been reported in patients with keratoconjunctivits sicca [179]. The rate of evaporation of the tear film has also been shown to be higher in the presence of a contact lens, and the effect remains for a period of 24 h after ceasing contact lens wear [184,185]. Since the evaporation rate is dependent on ambient temperature [186], humidity [175,180,187], and time of day [181,188], and can be affected by evaporation from the skin surrounding the eye, use of tear evaporation rate as a diagnostic and monitoring tool is challenging due to variable measurements.

In an attempt to address these issues, further techniques to measure tear evaporation rate have been proposed [189–191]. Using an infrared thermography camera [192], tear evaporation rate can be measured non-invasively while excluding the influences of the surrounding skin and sealed chambers [189,191]. Rohit et al. have recently described the modification and re-calibration of a dermatology instrument by attaching a swim goggle cup [190]. Using this instrument, the authors reported being able to obtain absolute rather than relative evaporation rates both with, and without, contact lens wear. Despite these developments, a "normal" tear evaporation rate has yet to be established questioning the diagnostic relevance of this measurement at the current time; in addition individual differences in evaporation rate contribute to the challenge.

6.3.2. Diagnostic test recommendation and technique

It should be emphasised that tear film stability test results are highly variable [125]. When performing tests to assess tear film stability, clinicians need to be meticulous about the procedures and factors that may influence the measurements. Thermography and tear evaporation rate evaluation are not well-established clinical techniques. Measurement of the tear breakup time with a non-invasive technique (NIBUT) is considered preferable to the FBUT [193] and the two techniques are well correlated [118,194]. Since there are several different methods for conducting the measurement, standardization is needed for consistency. The measurement should be made before any other invasive tests are conducted (such as eyelid manipulation or staining of the ocular surface). The patient should be instructed to blink naturally three times and then to cease blinking until instructions are given to blink again, and then to blink freely between measurements [66]. Where possible, an automated measurement system is recommended [193], since subjective measurements taken with a videokeratoscope and the Tearscope/Tearscope Plus (Keeler, Windsor, UK) have been shown to vary between measurement sessions and observers [38,195]. A NIBUT cut-off value of less than, or equal to, 10 s has been reported to be indicative of DED in Caucasians, when viewing the reflection of an illuminated grid pattern [134], but the cut-off value with automated measurement systems is generally shorter [141]. The difference might be attributable to the slower response rate of the observer in subjective techniques as well as the objective software detecting interference in the image capture process and interpreting these as breaks in the tear film.

6.4. Tear volume

Although not mentioned directly within the definition of DED, the tear film volume is important for ocular surface health and its loss of homeostasis (aqueous deficiency) may be at the same time a key pathogenic mechanism and a diagnostic sign in DED patients, independent of evaporative dry eye.

6.4.1. Current tests

6.4.1.1. Meniscometry (tear meniscus assessment).
Meniscometry describes assessment of the tear meniscus and may take the form of a height, or a cross-sectional volume metric. The tear menisci serve as reservoirs, supplying tears to the precorneal tear film [196]. The majority of tear fluid is contained within the menisci [197], formed by the tears lying at the junctions of the bulbar conjunctiva and the margins of both the upper and lower eyelids. The quantitative assessment of the tear menisci is, at present, the most direct approach to study the tear film volume. Slit-lamp techniques to study tear meniscus height (TMH), curvature (TMR), and cross-sectional area (TMA) are widely used in clinical practice and show good diagnostic accuracy and correlations with other DED tests [198,199]. However, this approach is operator-dependent and has important limitations, mainly related to fluorescein instillation and dependence on time-from-blink, which have potential impact on the tear film characteristics. The simplest type of slit-lamp meniscometry, based on judging the meniscus height by comparison to the variable slit-lamp beam height, has shown poor inter-visit repeatability [39]. Specialized meniscometry systems, equipped with a rotatable projection system that includes a target comprising a series of black and white stripes, a half-silvered mirror, and a digital video recorder, have been developed to facilitate simple and dynamic visualization of the tear meniscus, without the need for fluorescein instillation [200–202]. Meniscometry can be influenced by time after a blink, measurement locus along the lid margin, time of day, temperature, humidity, air speed, and illumination [2,66,203].

Application software for the iPod touch (Apple Inc., Cupertino,CA) has been recently developed to create a portable digital meniscometer that generates a grating of parallel black and white bands on the display, and which is reflected from the tear film at a working distance of 50 cm. This new slit-lamp mounted digital meniscometer exhibits good reproducibility, good agreement with both conventional video-meniscometry [204] and optical coherence tomography (OCT) meniscometry [205], and an ability to facilitate detection of tear meniscus changes following the instillation of artificial tears [206].

OCT assessment of the tear meniscus, described as an emerging technology in the TFOS DEWS 2007 report [2], has been extensively studied in the last ten years [207–226]. Upper and lower TMH, TMA, TMR and tear meniscus depth are, at present, the most commonly studied parameters. Spectral-domain OCT meniscometry has shown good intra-observer and inter-observer repeatability [212,219,224], that is superior to time-domain OCT [213,226]. The measurements are instrument-dependent [213,216], and can be biased by conjunctivochalasis, LIPCOF, disorders of lid margin congruity, and apposition between the lid and ocular surface [218,227]. The main advantages of OCT meniscometry are that it is non-invasive and image acquisition is rapid and simple, however analysis of the image may be complex, time-consuming and operator-dependent [224]. The development of validated measurement software is needed, ideally allowing dynamic image analysis to minimize interfering factors related to head, eye and eyelid movements [211,224].

6.4.1.2. Phenol red thread test.
The phenol red thread (PRT) test that received brief mention in TFOS DEWS report [2], and was removed more than 10 years ago from the Japanese DED diagnostic criteria, consists of a thin cotton thread soaked with phenol red, a pH-sensitive dye. When dry, the thread assumes a yellow color, but when moistened by tears the thread turns red as a consequence of

the slightly alkaline physiological pH of tears (between pH 7 and 8) [228]. The test is performed by hooking the folded end of the thread, within the temporal one-third of the eyelid, over the eyelid margin for 15 s. The small dimensions of the cotton thread should limit the chance of eliciting substantial reflex tearing [229], and the minimal amount of pH indicator soaked on the thread should minimize the irritating effect of the test, as shown by the repeatability of multiple PRT tests performed during the same session [230]. These elements suggest that the PRT test provides an indirect but realistic measure of the resting tear volume [231,232]. However, some authors have reported no significant correlation between the PRT test and tear volume determined with previously established methods such as TMH measurement or fluorophotometry [230], and poor correlation between PRT test results and DED symptoms [17]. Conflicting data, from weak [46], to strong [233], agreement, have been recently published on the correlation between the PRT test and Schirmer test. In clinical practice, an arbitrary cut-off value of 20 mm has been adopted to differentiate DED with and without aqueous deficiency using the PRT test [234]. A cut-off of 10 mm gives a sensitivity of 25% and specificity of 93% [235]. Doughty et al. reported small and not statistically significant differences between PRT performed with open or closed eyes [236].

6.4.1.3. Schirmer test. The Schirmer test is performed by folding the Schirmer paper strip (5 × 35 mm) at the notch and hooking the folded end over the temporal one-third of the lower lid margin. The score is the measured length of wetting from the notch, after a period of 5 min. The Schirmer test without anesthesia is a well-standardized test, providing an estimation of stimulated reflex tear flow. Although some authors have reported that the Schirmer test with topical anesthesia or nasal stimulation might be more objective and reliable in DED detection [237,238], there is a lack of high level evidence data on repeatability, sensitivity and specificity for these approaches [39]. Administering the test with the patient's eyes closed may minimize the variability of results [239], reducing the influence of the vertical gaze position [240], and horizontal eye movements [241]. Several diagnostic cut-off values have been proposed, from ≤5 mm/5 min [2], to ≤ 10 mm/5 min [228], and a range of sensitivity (77% [88] − 85% [242]) and specificity (70% [88] - 83% [242]) values have been reported. The combination of Schirmer and PRT tests has been proposed to improve the diagnostic accuracy, at least in patients with ADDE [228].

A variation of this test, termed strip meniscometry, involves dipping a strip (made of a 25-mm polyethylene terephthalate covered with a urethane-based material with a 0.4 mm central ditch containing a nitrocellulose membrane filter paper strip impregnated in natural blue dye reservoir) for 5 s into the tear meniscus [243]. Strip meniscometry with a cut-off of ≤4 mm has a sensitivity of 84% and specificity of 58% used in isolation and up to 81% sensitivity and 99% specificity when combined with TBUT [244].

6.4.2. Diagnostic test recommendation and technique

Meniscometry (volume or height) provides a non-invasive method to indirectly assess tear volume for DED sub-classification, with moderate repeatability especially if digital imaging rather than observational techniques are adopted. It is traditional to image the meniscus in the centre of the lower eyelid without lid manipulation shortly after a blink [66]. The Schirmer test without anaesthetic remains a diagnostic test recommended for confirming severe aqueous deficiency (such as in Sjögren syndrome) [245], but its variability and invasiveness, precludes it use as a routine diagnostic test of tear volume, especially in cases with evaporative dry eye secondary to MGD where tear quality rather than quantity is predominantly affected and any subtle reduction in resident tear volume in the interpalpebral space will likely be masked by the reflex tearing response on insertion of the strip.

6.5. Tear film composition

6.5.1. Current tests
6.5.1.1. Tear film osmolarity. A recent review of the literature identified 163 articles published since the year 2000 relevant to the use of tear osmolarity in the diagnosis of DED [246]. Hyperosmolarity of the tear film on the ocular surface causes a significant increase in interferon gamma, in the absence of large increases from other Th1, Th2 and Th17 cytokines, which can induce epithelial cell apoptosis through the JAK/STAT signalling pathway to induce cell death [247]. Tear osmolarity has been demonstrated to have the highest correlation to disease severity of clinical DED tests [11], and has been frequently reported as the single best metric to diagnose and classify DED [12,13,246]. However, other studies have indicated current measurement techniques to be highly variable [248]. Osmolarity generally increases with disease severity [174], classified as normal (302.2 ± 8.3 mOsm/L), mild-to-moderate (315.0 ± 11.4 mOsm/L) and severe (336.4 ± 22.3 mOsm/L). More severely affected subjects exhibit both an increased average and increased variability between eyes and between visits [11,170], making the marker heteroscedastic [170]. Various cut-off values for DED have been proposed in the literature, from 305 mOsm/L [249], to 316 mOsm/L [12], with reported sensitivities ranging from 64% to 91% [15,23,249,250], specificities from 78% to 96% [249,251], and positive predictive values ranging from 85% to 98.4% [249,252]. These data support the 316 mOsm/L cut-off as a specific threshold to better differentiate moderate to severe DED, or when used in parallel with other specific tests, while the 308 mOsm/L cut-off has become a widely accepted, more sensitive, threshold for use in routine practice to help diagnose mild to moderate subjects [13,15].

6.5.1.2. Tear film ferning. Ferning occurs when the tear film is dried, typically on a glass plate. As the pattern of the tear fern depends on the composition of the tear sample, tear ferning may be a simple test for tear film quality at a gross biochemical level. The process requires a slow crystal growth rate, low solution viscosity and low impurity levels to permit free-solute diffusion. Seven to 10 min under normal room temperature (20–26 °C) and room humidity (RH up to 50%) has been recommended [253]. The crystallisation begins with the formation of a nucleus, due to the supersaturation of ions with solvent evaporation at the peripheral edge of the drop. When the sample solute is able to diffuse into areas with a lower solute concentration, normal crystals can form [253].

Healthy tear samples produce compact, dense ferning patterns, while in dry eye samples, the pattern is fragmented or absent [254]. Electrolytes may play a role in ferning as hyperosmolarity has been found to result in deteriorated ferns [249,255]. It has a high reported sensitivity and specificity in Sjögren's syndrome [256–258]. and rheumatoid-induced keratoconjunctivitis sicca [259], but the results are more variable in DED [249,254]. Tear ferning is correlated with tear film volume and weakly with tear film stability, but seems to be independent of individual tear proteins [260]. Tear ferning changes with contact lens wear have been found to have a moderately high sensitivity (78.4%) and specificity (78.4%) for predicting contact lens tolerance in a clinical setting [261]. However, other studies have found that the tear ferning test had a poor correlation with tear film stability and symptoms in contact lens wearers [262].

6.5.2. Diagnostic test recommendation and technique

Despite some potential diagnostic ability, the underlying mechanisms responsible for producing tear ferning and their

interaction with dry eye sub-types are still poorly understood and hence this cannot currently be recommended as a diagnostic test [253].

With respect to osmolarity measurement, recent data have reinforced that two values are important to note in tear testing: the higher value of the two eyes, which is considered more indicative of the DED process, and the difference in value between the two eyes, which provides insight about the instability of the tear film [13]. Using the maximum value between both eyes has been shown to provide a higher dynamic range and larger observable change after effective therapy than using the average or single eye [10,170], and this approach is approved by the FDA for commercially available tests [263]. Normal subjects have little or no diurnal change, with repeat testing at time intervals of 1 min, 15 min, 1 day, and 5 days, demonstrating variation largely indistinguishable from the analytical precision of a commercial instrument ($\approx \pm 3-6$ mOsm/L) [10,170,264]. Moreover, a longitudinal study showed that tear osmolarity is the least variable of all the common signs for DED over clinically relevant time scales [10], which might seem counterintuitive, since tear osmolarity has the highest frequency of variation, changing blink-to-blink depending on the stability of the tear film and severity of disease, however the actual amplitude of variation is strongly dependent on disease severity. Inter-eye differences of normal, mild to moderate and severe DED patients were 6.9 ± 5.9 mOsm/L, 11.7 ± 10.9 mOsm/L, and 26.5 ± 22.7 mOsm/L, respectively [13]. The low variation of normal subjects contributes to the high specificity of the marker and makes it a good candidate for parallelization and therapeutic monitoring. Accordingly, normal subjects don't display elevated osmolarity, so a value over 308 mOsm/L in either eye or a difference between eyes ≥ 8 mOsm/L are good indicators of a departure from tear film homeostasis and represent a diseased ocular surface [265].

6.6. Damage to ocular surface

6.6.1. Current tests
6.6.1.1. Ocular surface staining. Punctate staining of the ocular surface is a feature of many ocular diseases and instilled dyes are used extensively in the diagnosis and management of DED. In addition, the distribution of micropunctate staining may provide an etiological clue [266]. The most frequently used dyes are sodium fluorescein, rose bengal, and lissamine green. The clinical appearance of fluorescein staining occurs whenever viable cells experience a compromise to their integrity such as a disruption in superficial cell tight junctions or defective glycocalyx [266,267]. It is suggested that there is some weak background fluorescence of healthy corneal epithelial cells [268]. Rose bengal stains ocular surface epithelial cells that are unprotected by mucin or glycocalyx, as well as dead or degenerated cells [269,270]. However, it stings on instillation and induces reflex tearing. In addition, it has been shown to suppress human corneal epithelial cell viability in vitro [271]. On the other hand, lissamine green is less toxic to the ocular surface and consequently is as well tolerated as fluorescein [272]; it stains epithelial cells only if the cell membrane is damaged (a vital dye), irrespective of the presence of mucin, whereas rose bengal, because of its cytotoxicity, produces staining irrespective of the state of cell health, once mucin is absent [273,274]; therefore lissamine green has largely replaced the use of rose bengal in evaluating ocular surface disorders [13,275]. There have been also several reports using mixtures of these dyes for simultaneous staining of the cornea and conjunctiva [272,276,277]. A solution of 2% fluorescein and 1% lissamine green has been found to be optimal in terms of comfort and staining efficacy, but is not commercially available [272]. Sequential staining and/or using more than one paper strip will increase the likelihood of observing ocular surface

damage [277,278]. Fluorescein has a peak excitation wavelength of 495 nm, whereas the commonly used 'cobalt blue' light filters of slit lamp biomicroscopes have a peak of around 450 nm [279]. The fluorescence peak is around 515 nm within the pH range of the tear film, so the yellow barrier filter required for optimum observation should band pass at around 500 nm [279]. For lissamine green, a red filter (567—634 nm) to enhance contrast against the sclera may enhance staining visibility [280]. For consistent recording of staining severity of the ocular surface, there are various grading systems including the van Bijsterveld system [242], the National Eye Institute/Industry Workshop guidelines [281], the Collaborative Longitudinal Evaluation of Keratoconus (CLEK) schema [282], the Oxford Scheme [283], the area—density combination index [284], and the Sjögren's International Collaborative Clinical Alliance ocular staining score (Table 3) [285]. Corneal and conjunctival staining have been shown to be informative markers of disease severity in severe DED, however, staining of the ocular surface in mild/moderate DED showed poor correlation with disease severity [11]. Therefore, observing staining of the cornea and conjunctiva is considered an important aspect in the clinical analysis of severe DED.

6.6.1.2. Impression cytology. Impression cytology is a relatively simple and practical technique that has been used in the diagnosis of the ocular surface disorders such as DED, limbal stem-cell deficiency, ocular surface neoplasia, and specific viral infections [287]. During the past decade, it has become standard to study squamous metaplasia and goblet cell density of the conjunctiva for the diagnosis and monitoring of DED [288]. Cells from the first to third most superficial layers of the epithelium are removed by application of cellulose acetate filters or biopore membranes, and the cells can be subsequently analyzed by various methods including microscopy, immunocytochemistry, immunoblotting analysis, polymerase chain reaction, and flow cytometry, depending on the objective of the investigation [289]. Specific examination procedures for impression cytology are described elsewhere [290]. For analyzing conjunctival impression cytology, several squamous metaplasia grading systems based on qualitative or quantitative cytological criteria are applied. The best-known methods include the systems by Nelson [291], Tseng [292], and Blades [293]. Among them, the Nelson classification system, considering the density, morphology, cytoplasmic staining affinity and nucleus/cytoplasm ratio of conjunctival epithelial and goblet cells, remains widely used [294].

6.6.1.3. Lid Parallel Conjunctival Folds (LIPCOF). Lid-parallel conjunctival folds (LIPCOF) are folds in the lateral, lower quadrant of the bulbar conjunctiva, parallel to the lower lid margin. LIPCOF may represent the first mild stages of conjunctivochalasis and thus share the same aetiology [295], but they display slightly different characteristics clinically. The cross-sectional area of LIPCOF is much smaller than that of conjunctivochalasis [227,296]. LIPCOF [235,297—299] do not occur centrally as does conjunctivochalasis, and appear unrelated to age [297]. While conjunctivochalasis can be induced or increased by forceful blinks or digital pressure towards the lid margin or gaze [300], this does not appear to happen in the case of LIPCOF.

LIPCOF occur behind the temporal and nasal tear meniscus along 2/3 of the total length of the inferior tear meniscus [203], and may cause tear meniscus height measurements to be underestimated [227]. Decreased mucin production is associated with the severity of LIPCOF [298], and LIPCOF are significantly correlated with lid wiper epitheliopathy [235,298]. LIPCOF may be related to completeness of the blink [301], blink speed and tear film viscosity [295].

Table 3
Grading scales for ocular surface staining.

Scale	Cornea	Conjunctiva	Features
van Bijsterveld system [242]	1: few separated spots 2: many separated spots 3: confluent spots	Nasal and temporal zones: 1: few separated spots 2: many separated spots 3: confluent spots	Focus of Sjögren syndrome Out of 9
National Eye Institute/Industry Workshop guidelines [281]	divided into five sectors (central, superior, inferior, nasal and temporal), each scored 0–3	divided into superior paralimbal, inferior paralimbal & peripheral area both nasally & temporally, each scored 0–3	Total 15 corneal and 9 conjunctival
Collaborative Longitudinal Evaluation of Keratoconus (CLEK) schema [282]	divided into five sectors (central, superior, inferior, nasal and temporal), each scored 0–4 in 0.5 steps	divided into four sectors (superior, inferior, nasal and temporal), each scored 0–4 in 0.5 steps	Fluorescein ICC = 0.76 Rose bengal ICC = 0.40 [39]
area–density combination index [284]	area (A0: no punctate staining; A1: >1/3; A2: 1/3 to 2/3; A3 >2/3) & density (D0: no punctate staining; D1: sparse; D2: moderate; D3: high with lesion overlap).	NA	combined in single index e.g. A2D3
Oxford staining score [283]	Fluorescein, lissamine or rose bengal can be used; 0 to V grade dependent on intensity of punctate staining displayed pictorially across a combination of the cornea and conjunctiva.		Dots increase on a log scale between grades
ocular staining score [285]	Fluorescein 0: 0 dots 1: 1–5 dots 2: 6–30 dots 3: >30 dots	Lissamine green 0: 0–9 dots 1: 10–32 dots 2: 33–100 dots 3: >100 dots	Fluorescein extra points +1 for confluent patches, staining within pupil or filaments Out of 12 ICC ~0.90 [286]

ICC, intra-class correlation.

Patients with increased LIPCOF grades are likely to suffer from DED [297,302–304]. One study, showed that combining nasal LIP-COF and NIBUT using an algorithm appeared to be the most predictive DED test combination [235]. Sensitivity of LIPCOF Sum (nasal + temporal LIPCOF) to discriminate between normal and symptomatic DED patients was reported to be 70%, and specificity was reported to be 91%, for a cut-off value of 2, using a revised LIPCOF grading scale (Table 4), where the LIPCOF score is derived from the number of folds rather than the height of the folds [298,299,302]. Another group evaluated a medium predictive ability of temporal LIPCOF using the Höh et al. fold height-based grading scale [297], and defined the cut-off value as 2, giving a sensitivity of 52% and specificity of 64% [303].

LIPCOF are observed, without fluorescein, on the bulbar conjunctiva in the area perpendicular to the temporal and nasal limbus, above the lower lid (temporal and nasal LIPCOF, respectively), with a slit-lamp microscope using ~25× magnification (Fig. 2) [298,299,302]. LIPCOF can be classified by different grading scales, such as a recent scale counting the number of folds [235]. Care must be taken to differentiate between parallel, permanent, conjunctival folds (LIPCOF, single folds height ~0.08 mm) and disrupted micro-folds (height~0.01 mm) [298,299,302,305–307].

Researchers have also used OCT [308,309] and Scheimpflug photography to observe LIPCOF [297]. Using these instruments, additional criteria such as cross-sectional area of LIPCOF or LIPCOF coverage by the tear meniscus can be evaluated [297,308,309]. Conjunctival shrinkage has been proposed as a diagnostic feature of dry eyes [310],and has been shown to occur more in patients with

dry eye symptoms, less stable tears and with ocular surface staining, but not those with MGD [311].

6.6.1.4. In vivo confocal imaging. In-vivo confocal microscopy (IVCM) is a non-invasive technique that allows the evaluation of signs of ocular surface damage in DED at a cellular level [312,313], including decreased corneal (apex and lower periphery) [314], and conjunctival epithelial cell density [315–317], conjunctival squamous metaplasia (increased mean individual epithelial cell area, decreased nucleocytoplasmic ratio and goblet cell density) [318], and corneal nerve changes (decreased sub-basal nerve density, increased tortuosity and increased number of bead-like formations) [315,317,319–325]. Laser scanning IVCM allows easy identification of conjunctival goblet cells (although some concerns have been reported with regard to evaluation of the tarsal region [326]), suggesting it may be a valuable tool in assessing and monitoring DED-related ocular surface damage [312,327–330]. The IVCM approach seems less invasive, but as effective as impression cytology [318,331], however it has not yet been widely adopted in clinical practice and its predictive ability in the diagnosis of DED is unknown.

6.6.1.5. Ocular surface sensitivity. Cochet-Bonnet or non-contact air-jet esthesiometers have been employed to evaluate ocular surface sensitivity. Loss of corneal sensation can give rise to severe corneal epithelial disorders such as neurotrophic keratopathy [332,333]. The palpebral conjunctival sensitivity appears to be more critical than corneal sensitivity when assessing DED [334]. Corneal esthesiometry is weakly correlated to other DED tests, but increases with severity of the disease and has achieved a reported specificity of 96%, but a sensitivity of just 19% [249].

6.6.2. Diagnostic test recommendation and technique

While corneal staining is perhaps a later stage feature of DED, combination staining with fluorescein and lissamine green instilled by a moistened and saturated filter paper strip to highlight corneal and conjunctival/eyelid margin tissue damage, respectively [272,276,277,279], is recommended as the most appropriate

Table 4
Example of a LIPCOF grading scale [235].

	Grade
No conjunctival folds	0
One permanent and clear parallel fold	1
Two permanent and clear parallel folds, (normally <0.2 mm)	2
More than two permanent and clear parallel folds, (normally 0.2 mm)	3
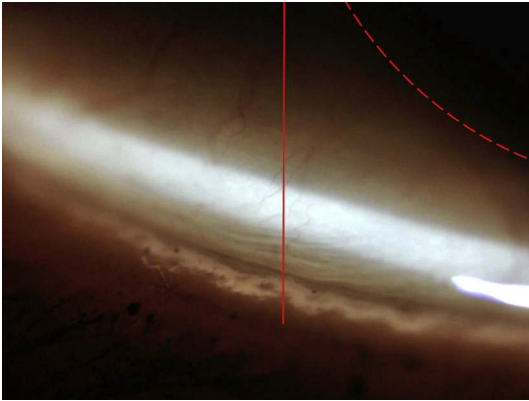
Fig. 2. LIPCOF degree 2 (Pult Scale) - dashed line indicates the corneal limbus - vertical perpendicular line indicates the appropriate area for observation.

diagnostic technique for evaluating ocular surface damage. Ophthalmic stain strips are registered as medical devices rather than pharmaceuticals in some countries and lissamine green is not a licensed product in other countries so it is acknowledged that access to these dyes in a clinical setting can be challenging; however, there have been no reported adverse effects, therefore, the benefits would appear to outweigh any risks.

6.7. Inflammation of the ocular surface

Inflammation is a recognized component of the pathophysiological mechanism of DED [4] and has been proposed to offer a stable indicator of DED severity [335]. However, inflammation is not specific to DED and can occur in other ocular or systemic disease [336,337]. In autoimmune disease, inflammation occurs in the eye as well as specific sites around the body, for example, the joints in rheumatoid arthritis. Autoimmune serum markers, including SSA and SSB are most often evaluated in Sjögren syndrome. Other relevant autoimmune diseases include systemic lupus erythematosis, mixed connective tissue disease, chronic hepatitis, Stevens Johnson syndrome and chronic graft versus host disease (GVHD). Systemic investigations used to differentially diagnose the cause of inflammation might involve biopsy of salivary glands, flow cytometry of peripheral mononuclear blood cells, radiology or imaging of joints.

6.7.1. Current tests

6.7.1.1. Ocular/conjunctival redness. The most common clinical sign that is suggestive of ocular surface inflammation is conjunctival redness [338–340]. This is a consistent sign of conjunctival vascular dilatation and reactive change to pathological stimuli. It can occur in any disease with inflammation, not just DED, for example, in response to chemical injury, infective conjunctivitis or allergic conjunctivitis. Ocular redness can be easily detected with a pen torch or standard slit lamp biomicroscopic examination. For the purpose of diagnosis and documentation of treatment effects, more quantitative documentation methods using digital imaging analysis have been developed [341–344].

6.7.1.2. Matrix metalloproteinases. The matrix metalloproteinases (MMPs) are one of many classes of proteases secreted into the tears in DED [174,345–347]. The level of MMPs reflect the loss of ocular surface barrier function, since MMPs can destroy tight junctions in the ocular surface epithelium. MMPs are produced as inactive proenzymes and can be cleaved to become active enzymes. It is

therefore important for an MMP diagnostic test to detect enzyme activity levels and not just total tear protein levels. One development is the availability of a commercial 'point of care' diagnostic device (InflammaDry®, Rapid Pathogen Screening, Inc, Sarasota, FL, USA) which assays tear MMP-9 levels in 10 min [348]. In its current form, this assay produces a dichotomous outcome, with levels above 40 ng/ml producing a positive result, and is non-specific to the source of ocular surface inflammation.

6.7.1.3. Cytokines and chemokines. The levels of tear cytokines and chemokines are important and reflect the level of epithelial disease. Certain cytokines can highlight a specific disease process, for example, elevation of Th1 and Th17 subclasses of cytokines suggest involvement of particular T lymphocyte differentiation pathways in the disease [349]. Elevation of tear Th2 cytokines, on the other hand, may suggest a more allergic-based disease, although recent evidence suggests various aspects of T cell Th1, Th2 and Th17 exist across aqueous deficient, evaporative and mixed forms of DED, with a propensity towards Th1 type T cell responses as a more global indicator of DED [350]. Since collection of tear fluid is relatively non-invasive compared to biopsies or venipuncture for serum assays, it is an attractive idea to include these as diagnostic tools [174,351].

A recent report on standard operating instructions for the tear assay of tumor necrosis factor alpha, interferon gamma, IL-1 beta and IL-6 has been published [352]. This refers to the collection, storage and repeatability of the tear assay, but with laboratory testing rather than a 'point of care' device. The operating instructions could be made even more cost-effective by reducing the need for reagents [353]. However, it has been found that tear IL-10 and IL-1β levels had significant inter-day variations, while epidermal growth factor, fractalkine, IP-10 and vascular endothelial growth factor were consistently higher in the evening compared to the midday measurements [354]. Such issues will affect how readily these tests are adopted in routine clinical practice. Tear chemokines such as CXCL9, -10, -11, and CXCR3 are important in the tear fluid, as they serve as ligands for specific chemokine receptors on immune cells [29,355–357]. The elevation of specific ligands may therefore imply the involvement of the specific lymphocytes in the ocular surface, without actually measuring the presence of these lymphocytes.

6.7.1.4. Ocular surface immune markers. The most commonly used ocular surface immune marker is HLA-DR expression, a Class-II MHC antigen, which indicates a loss of the normally immune-suppressed environment of the ocular surface. Epstein has recently published standard operating instructions for impression cytology, for use in clinics and in clinical trials [358]. It was reported that sufficient conjunctival epithelial cells could be harvested for the quantification of HLA-DR using a suitable impression membrane, for example, the commercially available Eyeprim™ membrane (Opia Technology, Paris, France). The precision/repeatability of HLA-DR expression was studied and it was noted that collection, storage and shipment of specimens from distant sites were successful and storage of specimens for up to 30 days (with refrigeration) before processing did not affect results. Since the centralized laboratory was able to track large number of masked samples reliably, the authors suggest that this tool is suitable for use in randomized controlled trials of DED.

Although the authors found increased expression of HLA-DR associated with increased clinical severity of DED [358], a comparison with six other studies showed that the normal levels of HLA-DR expression are very variable (ranging from 5% to 54%), and its correlation with traditional clinical signs of DED is weak [359]. This may suggest that not all DED cases are equally inflammatory,

or that the marker is non-specific for DED and indeed can reflect any ocular surface inflammation. Nevertheless. use of impression cytology can be helpful in the documentation of particular immune cells in specific contexts of DED. For example, the quantification of neutrophil involvement in Stevens Johnson syndrome has been published [360]. Other relevant markers of apoptosis include CAM-1, CD14[+], CD8[+] and CD4[+] cells [361,362].

6.7.1.5. In vivo confocal imaging.
Corneal sub-epithelial and stromal IVCM signs of inflammation have been hypothesized and studied in DED for over 10 years [363,364]. More recently, IVCM has allowed examination to be extended to a number of components of the ocular surface morpho-functional unit [315,365]. Recent literature has shown significant differences between patients with DED and controls, and among different types of DED, for many presumed inflammatory parameters, including corneal dendritic cells (DC), stromal hyper-reflective (activated) cells [317,320,321,366], conjunctival hyper-reflective roundish or ovoidal (inflammatory) cells [316,326,367], and meibomian gland acinar wall and inhomogeneous appearance between 'slices' (inflammatory infiltration) [329,368]. Some of these parameters have shown good repeatability and correlate with tear film inflammatory mediators, and other signs of DED [369]. Inflamed ocular surfaces, in immune-mediated diseases and in DED, show not only increased DC density, but also morphological DC changes, which may indicate cell maturation [364,370]. In recent research, IVCM imaging of DCs in DED has been able to predict, as well as monitor, the response to anti-inflammatory drugs [370,371], and to detect sub-clinical ocular surface inflammation [372].

6.7.2. Diagnostic test recommendation and technique
As described, practitioners need to be aware that the ocular inflammation tests mentioned are not specific for DED. For a clinical test to be acceptable, it should be readily performed without excessive demands on technical manpower or time [373]. For this reason, research techniques such as mass spectrometry [374,375], have not been included in this section. The technical challenges involved in assessment of tear protein levels should not be underestimated. Only a very minute amount of tears can be sampled from DED patients, and since the linear range of many analytes is different, differential dilution of the collected tears may be necessary. Some of the tests may be problematic when used in a population without normal reference values. For example, many tear cytokines and even MMPs tend to increase with age [376], and age specific upper limits of the normal values have not been published, thereby potentially limiting the usefulness of the tools as diagnostic devices. However, multiplexed cytokine systems are increasingly becoming available [353]. Currently most practitioners do not include one of these tests for inflammation as a prerequisite for clinical diagnosis of DED. Certain clinical tools have been available for a long time, but the recent availability of a standard commercial platform, such as the ocular redness index within the Oculus Keratograph 5M software suggests that tools for measuring inflammation may now be within reach of many clinicians [340].

With the availability of newer immunosuppressive medications and trials concerning these drugs [377,378] it is logical that inflammation should be assessed. The exact modality used may need to be varied depending on the pathway or target cell upon which the immunosuppressive drug acts, and such diagnostic tools should be used for refining patient selection as well as monitoring after commencement of treatment. Costs of these diagnostic tests should be considered, but these should be calculated from a holistic standpoint. For example, if the tests can assist the channelling of patients to appropriate healthcare services there may be cost savings for reduced referrals.

6.8. Eyelid aspects

6.8.1. Current tests
6.8.1.1. Anterior.
Anterior eyelid features, such as anterior blepharitis and demodex blepharitis, are differential diagnoses and comorbidities of DED rather than diagnostic criteria and therefore are discussed in Section 9.

6.8.1.2. Posterior
6.8.1.2.1. Lid wiper epitheliopathy (LWE).
A small portion of the marginal conjunctiva of the upper and lower lid acts as a wiping surface to spread the tear film over the ocular surface [379,380]. This contacting surface at the lid margin has been termed the 'lid wiper' [379]. The normal lid wiper is rich in goblet cells [381], and appears to be the most sensitive conjunctival tissue of the ocular surface [382]. Lid wiper staining with dyes such as fluorescein and lissamine green, which occurs principally in DED patients [298,299,379,383,384], has been termed lid wiper epitheliopathy (LWE) or upper lid margin staining [379,385,386]. It has been proposed that LWE is related to increased friction (direct contact between surfaces) throughout blinks [298,379,383,384], although modelling of the tribology suggests that tear film viscosity-induced hydrodynamic forces at the start of each blink are the principal cause [295]. Boundary lubrication may therefore play a key role in reducing dry eye [387]. LWE occurs on the upper and lower lids, but most studies report only upper LWE. Lower LWE in contact lens wearers has been found to be associated with DED symptoms in some studies [388], but not others [298,299].

Korb and colleagues reported that 88% of symptomatic patients had LWE but only 16% of asymptomatic patients presented with LWE [45]. Shiraishi et al. reported a higher prevalence of LWE in younger than older contact lens wearers [389]. The predictive ability of upper LWE is reported to be 48% (sensitivity) and 96% (specificity) in non-lens wearers using a cut-off value of grade 1 (based on the Korb grading scale; Table 5). In their protocols, Korb et al. recommend the use of fluorescein and lissamine green in combination to stain LWE with repeated instillation of lissamine green before the evaluation of LWE [45,278]. However in another study, LWE increased following repeated lid eversion, but not dual instillation [390]. LWE can be observed immediately adjacent to the lid margin of the everted eyelid using a slit lamp biomicroscope and is most commonly classified by combining the extent of its staining, in terms of length in mm, and width relative to the lid margin width [45,235,278,298,302,379,391]. Another grading system has proposed observing the area and staining pattern [392]. A more advanced method may be to use confocal microscopy, where small hyperreflective dots, assumed to highlight inflammation, have been observed in lens wearers wearing high coefficient of friction lenses [384].

6.8.1.2.2. Interferometry.
Oily substances spread to form a thin layer on the surface of water. Exposure of such an oily layer to adequate light results in the generation of an interferometric fringe pattern from interference from the front and back surface refractive index change reflections (from the interface with the air and the muco-aqueous tear film phase respectively). The superficial oily layer of the tear film is thought to retard evaporation of the tears, and, with the rest of the tear film, provides an optically smooth surface over the cornea [148,175,393]. The lipids produced by the meibomian glands usually distribute dynamically from the inferior to the superior region over the ocular surface and then stabilize shortly thereafter [394].

In conjunction with the surface reflection pattern and dynamics, interferometry can allow the thickness of the lipid layer of the tear film to be estimated [148]. Using slit lamp photometry to measure reflectivity, Olsen first estimated the thickness of the