Challenges in Using Signs and Symptoms to Evaluate New Biomarkers of Dry Eye Disease

Benjamin Sullivan, PhD

INTRODUCTION

The lack of association between signs and symptoms in dry eye disease (**DED**)[1-4] presents a challenge in designing clinical trials to objectively evaluate the performance of new biomarkers. When making a diagnosis based on multiple signs, one must balance the sensitivity of therapeutic response, temporal variability, the *a priori* probability distribution of each sign, and the patient's history of therapy against the practical considerations of patient recruitment. Proper classification will almost entirely determine the success or failure of a clinical trial, yet there is little agreement between the constituent signs in DED.

Tear osmolarity, a test that has been proposed as the "gold standard" of DED and has recently become widely available, represents an example of such a conundrum that clinicians and researchers must learn how to resolve. Many decades of basic research have confirmed the role of tear hyperosmolarity as a causative mechanism and clinical endpoint in the pathogenesis of DED.[5-10] Current literature reveals that tear osmolarity demonstrated the highest correlation with disease severity and was found to be the single best metric to diagnose and classify DED.[11,12] Osmolarity has been found to be superior in overall accuracy to any other single test for dry eye diagnosis, even when the other test measures were applied to a diagnosis within the sample groups from which they were derived,[9] and recent data suggest that patients with severe keratoconjunctivitis sicca have a significantly higher tear film osmolarity than healthy controls.

Testing tear film osmolarity can be a very effective objective diagnostic tool in the diagnosis of DED.[13] Further, with tear osmolarity >305 mOsm/L selected as the cut-off value for dry eye, the likelihood ratio was 10.99, higher than that used for the other tests. The positive predictive value was also very high (98.4). Osmolarity performed well in dry eye diagnosis, better than the other tests considered, mainly in severe dry eye. Tear osmolarity values should be interpreted as an indicator of the DED evolutionary progression to severity.[14] Basic studies have revealed that tear osmolarity may be more relevant to the clinician in the diagnosis of ocular surface diseases with an increased level of inflammatory mediators.[15] Finally, more recent data demonstrated that tear osmolarity was the least variable of all the common signs across a clinically relevant timeframe, and was the only sign to reduce its variation upon application of effective therapy.[16]

Despite the demonstrated success of tear osmolarity as a marker of dry eye, critics suggest that it provides little clinical value. Messmer et al[17] and Szalai et al[18] reported that osmolarity measurements were not able to distinguish between healthy volunteers and patients with dry eye. However, these early papers failed to 1) take into account patient therapeutic status, 2) use the maximum of two eyes in diagnosis, 3) report validated quality controls, and 4) recognize the heteroscedasticity of the marker as an indicator of tear film instability. Probably the most important aspect of these studies is that they uniformly relied upon a series of threshold-limited, uncorrelated variables (corneal staining, tear film breakup time [**TFBUT**], etc.) to determine which patients were and were not classified as dry eye subjects.

Such inconsistent results highlight the dilemma of using combinations of signs to inform patient diagnosis or assess disease severity. This editorial explores the various implications and approaches of using multiple uncorrelated signs for diagnosis and severity assessment in DED.

STATISTICAL INDEPENDENCE AND SEVERITY ASSESSMENT

Hypothetically, if tear osmolarity were perfectly correlated with staining, Schirmer test results, or symptom questionnaires, there would be no reason to measure it — the information would already be available. Correlation with individual signs is therefore a poor choice for evaluating the clinical utility of new biomarkers, especially in a disease such as dry eye in which no consensus "gold standard" measurement exists. If a new marker gives new information, one must presuppose that clinical markers will routinely disagree

Accepted for publication October 2013.

From TearLab Corporation, San Diego, CA.

The author is Chief Scientific Officer of TearLab Corp., which manufactures and distributes the TearLab Osmolarity System.

Corresponding author: Benjamin Sullivan, PhD, Chief Scientific Officer, TearLab Corp., 9980 Huennekens St., Suite 100, San Diego, CA 92121. Tel: 760-224-4595. E-mail address: bdsulliv@TearLab.com

OUTLINE

and that conflicting information is the norm rather than the exception (Table 1). For instance, the absence of staining gives the clinician no information as to whether that same patient will have normal TFBUT, yet we find few papers that are critical of staining due to a lack of correlation to TFBUT. Correlation across small subsets of patients is certainly possible and has been observed,[19,20] but in large cohorts where there are diverse underlying etiologies, signs tend to diverge.[21]

The implications of statistical independence have a profound impact on how combinations of signs of dry eye influence diagnosis and severity assessment. A particularly good example can be found in a paper by Chotikavanich et al, detailing the production and activity of MMP-9 on the ocular surface.[22] Patients were included based on a threshold of Ocular Surface Disease Index (**OSDI**) symptoms >20, and one or more of the following signs: TFBUT ≤7, punctate corneal fluorescein staining, or Schirmer I score <10 mm. Computerized videokeratoscopy provided a measure of ocular surface regularity. Subjects were then classified into dysfunctional tear syndrome severity levels (DTS1−DTS4) based on increases in corneal staining, while the other signs were unconstrained and allowed to vary. Predictably, there was no discernable pattern to the symptoms, surface regularity index, TFBUT, or conjunctival staining for the 35 subjects classified as DTS1−DTS3. Despite the attempt to sort patients into increasing disease severity categories, symptoms *decreased* from DTS1 to DTS3, the ocular surface regularity actually improved from DTS1 to DTS3, and the TFBUT and conjunctival staining were less severe in DTS2 than in DTS3. While all of the signs were much worse in DTS4, that cohort was populated with patients with graft vs host disease, Stevens-Johnson syndrome, or Sjögren syndrome (**SS**), whereas DTS1−DTS3 were representative of common dry eye patients, featuring meibomian gland disease, blepharitis, and history of LASIK.

So what happened? Chotikavanich and co-authors are world-class experts at identifying dry eye, so these outcomes cannot be an indictment of the clinical technique. Rather, the distributions are a direct result of the statistical independence of dry eye signs. When uncorrelated signs are sorted based on a small subset of markers, the signs that were not used in the stratification will become randomly distributed. This phenomenon is also observed in the critiques of tear osmolarity listed above; patients stratified based on thresholds of uncorrelated variables produced random distributions in tear osmolarity. Chotikavanich et al do not comment on the value of symptoms, TFBUT, or ocular surface regularity in DED any more than Szalai[18] and Messmer[17] provide information regarding the clinical utility of tear osmolarity. Instead, the random distributions are entirely a consequence of the choices made when stratifying patients due to the underlying independence of dry eye signs.

A similar investigation by Huang et al, also performed by highly qualified and experienced researchers, measured biomarkers in 102 subjects.[23] Dry eye severity was determined by symptoms and stratified by increasing corneal staining. As expected, DE1−DE3 groups exhibited linear increases in staining with low intragroup variation, while Schirmer tests and tear osmolarity exhibited noisy, random behavior with large spreads within each group. If severity classifications based on staining and symptoms were accurate, we would expect Schirmer scores to be lower in patients with severe dry eye compared to those with more moderate disease, yet DE3 (11.3±9.3 mm) showed higher wetting than DE2 (8.8±6.7 mm). We also expect that tear osmolarity should rise when progressing from the moderately affected DE2 patients to the more severely affected DE3 patients, yet this is not what is observed when staining and symptoms alone estimate disease severity (DE2=306.8±11.6 mOsm/L to DE3=302.6±10.2 mOsm/L). Clearly, the methods used to formally stratify disease severity in a clinical trial setting can have a variety of unintended consequences that disagree with routine clinical experience.

SENSITIVITY OF THERAPEUTIC RESPONSE

One of the most significant confounding variables in classifying patients in a trial is the varying sensitivity each marker has to concomitant therapy. For instance, several artificial tear formulations have been shown to significantly reduce tear osmolarity *prior* to the response from other signs or symptoms, which tend to be less sensitive to therapeutic intervention than osmolarity.[16,24,25] If the relative sensitivities and dynamics are not taken into consideration in forming inclusion criteria, a cross-sectional study of SS subjects who are heavily treated would be expected to have a lower incidence of hyperosmolarity but moderate-to-high symptoms. Indeed, that is what has been reported in the literature.[18,20] Compare this to a case report of tear osmolarity readings on untreated SS patients presenting to a general ophthalmologist[26]:

> "I really didn't know how [tear osmolarity] would impact my practice. I figured I was pretty good at telling if someone had dry eye, and I wasn't sure what a tear osmolarity number would do for my treatment algorithm. Then, a patient we'll call Tammy walked in to my office. She had bounced from ophthalmologist to ophthalmologist and had even been recommended to see a psychiatrist. She had severe ocular surface discomfort, dryness, pain, and intermittent blurred vision. Her story was classic for dry eye. The only problem was that she had no corneal staining whatsoever. She seemed to have an adequate tear lake and lacked the