Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Analytical performance, spatial dynamics, and clinically meaningful change thresholds for automated non-invasive tear film assessment using the Oculus Keratograph 5M

Daniela Oehring

daniela.oehring@plymouth.ac.uk

University of Plymouth

Research Article

Keywords:

Posted Date: March 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9181006/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

Non-invasive tear film metrics from automated topographers are increasingly used, yet device-specific analytical performance remains incompletely characterised. This study quantified the analytical performance of the Oculus Keratograph 5M for non-invasive break-up time (NIKBUT) and tear meniscus height (NIKTMH), evaluated agreement with conventional methods, and characterised the spatial dynamics of tear film break-up.

Methods

Thirty-five participants (18 symptomatic, 17 asymptomatic) attended three visits on consecutive days. NIKBUT (first and average), NIKTMH, fluorescein break-up time (FBUT), and slit-lamp tear meniscus height (TMH) were each measured three times per eye per visit. Precision (CV), reliability (ICC 3,1), standard error of measurement (SEM), and minimum detectable change (MDC_{95}) were calculated. Method agreement was assessed using random-effects Bland-Altman analysis. Spatial distribution of break-up events was analysed by corneal zone.

Results

NIKTMH demonstrated excellent precision (CV = 8.8%) and moderate-to-good reliability (ICC = 0.727), with an MDC_{95} of 0.173 mm. NIKBUT showed poor precision (CV = 53.6% for First, 42.8% for Average) and symptom-dependent reliability. FBUT required a change exceeding 9.28 s to surpass its MDC_{95}. Bland-Altman analysis confirmed systematic bias between NIKBUT and FBUT with limits of agreement exceeding ± 19 s and proportional bias. Spatial analysis revealed that NIKBUT break-up occurred predominantly paracentrally (53–63%), while FBUT events concentrated centrally (86–97%), indicating the methods capture fundamentally different tear film phenomena. Intra-subject repeatability of break-up location was poor (Krippendorff's α = 0.115–0.308).

Conclusions

NIKTMH is the most analytically robust Keratograph metric, suitable for longitudinal monitoring when changes exceed its MDC_{95} of 0.173 mm. NIKBUT shows poor precision; only large changes exceed noise. Spatial analysis confirms that NIKBUT and FBUT interrogate distinct biophysical processes - these methods are not interchangeable. These benchmarks should inform clinical interpretation and study design.

Key Points

• The minimum detectable change for FBUT is 9.28 seconds and for NIKTMH is 0.173 mm; smaller changes cannot be distinguished from measurement noise and should not be over-interpreted clinically.

• NIKTMH demonstrated the best analytical performance of all Keratograph metrics (CV = 8.8%, ICC = 0.727), making it the most reliable automated endpoint for longitudinal tear film monitoring.

• Spatial analysis reveals NIKBUT break-up occurs predominantly paracentrally (53–63%) while FBUT concentrates centrally (86–97%), confirming these methods capture fundamentally different tear film phenomena and are not interchangeable.

1. Introduction

Dry eye disease (DED) is a multifactorial condition of the ocular surface characterised by loss of tear film homeostasis, with a global prevalence estimated between 5% and 50% (1). Tear film instability is recognised as a core mechanism, making its accurate and reliable assessment fundamental to both diagnosis and management (1, 2).

The fluorescein tear film break-up time (FBUT) test has served as the clinical standard for evaluating tear film stability for decades. However, its limitations are well documented: the instilled dye alters tear film osmolarity and structure, paradoxically destabilising the parameter being measured, while reliance on clinician observation introduces substantial inter- and intra-observer variability (1, 3, 4). Similarly, manual slit-lamp assessment of tear meniscus height (TMH) is subject to observer variability and limited resolution.

Automated non-invasive instruments, such as the Oculus Keratograph 5M, address these limitations by projecting Placido rings onto the cornea and using software algorithms to detect distortions corresponding to tear film break-up, yielding metrics including the non-invasive Keratograph break-up time (NIKBUT) and non-invasive tear meniscus height (NIKTMH) (5). These systems offer objectivity, repeatability in principle, and removal of the confounding effects of fluorescein instillation.

However, a large body of evidence now confirms that while NIKBUT and FBUT are often correlated, they are not interchangeable (6–9). The DREAM study, a large multicentre randomised clinical trial, found only weak correlations between Keratograph NIKBUT and FBUT (Spearman $\rho = 0.18–0.26$) (10). Similarly, automated NIKTMH shows variable agreement with other objective measures including anterior segment optical coherence tomography (11, 12). These findings underscore that non-invasive metrics cannot simply be substituted for their conventional counterparts.

Yet despite the growing clinical adoption of these instruments, a critical gap persists: the device-specific analytical performance - precision, day-to-day reliability, and inherent measurement error - that determines whether an observed change in a patient represents a genuine clinical change or merely measurement noise, remains insufficiently characterised for the Keratograph 5M. While several studies have reported repeatability coefficients for NIKBUT (13–16), few have provided the full suite of clinically interpretable metrics - intra-session coefficient of variation (CV), inter-session intraclass correlation coefficient (ICC), standard error of measurement (SEM), and minimum detectable change (MDC_{95}) - that clinicians need to interpret sequential measurements with confidence.

The MDC_{95} is particularly important for clinical practice. It defines the smallest change in a measured value that exceeds the inherent measurement error at the 95% confidence level. Any observed change smaller than the MDC_{95} cannot be reliably attributed to a true biological change and may simply reflect the instrument's noise floor. Without this benchmark, clinicians risk over-interpreting small fluctuations in NIKBUT or NIKTMH as evidence of treatment response or disease progression.

Furthermore, automated and conventional methods may differ not only in the timing of tear film break-up but in its spatial location on the cornea. Guarnieri, Carnero (17) used the Keratograph 5M to characterise the spatial distribution and progression of automated tear film break-up in glaucoma patients, demonstrating that break-up location and area differed between glaucomatous and healthy eyes, but did not compare the spatial patterns of automated and conventional methods. If NIKBUT and FBUT consistently identify break-up in different corneal regions, this would suggest they capture distinct biophysical processes rather than merely providing different measurements of the same phenomenon. Characterising these spatial differences could provide mechanistic insight into the non-interchangeability of automated and conventional methods and has implications for understanding the pathophysiology of different DED subtypes (18).

The primary aim of this study was therefore to comprehensively characterise the analytical performance of the Oculus Keratograph 5M for NIKBUT and NIKTMH measurement, including precision, day-to-day reliability, and clinically meaningful change thresholds (SEM and MDC_{95}). Secondary aims were to quantify inter-method agreement between the automated non-invasive metrics and their conventional clinical counterparts (FBUT and slit-lamp TMH) using rigorous Bland-Altman methodology, and to characterise the spatial dynamics of tear film break-up - including the distribution, timing, and intra-subject repeatability of break-up location - to determine whether NIKBUT and FBUT interrogate the same or distinct biophysical processes.

2. Methods

2.1 Study design and participants

This prospective, single-centre, repeated-measures study was conducted at the Centre for Contact Lens Research (CCLR), University of Waterloo, Ontario, Canada, between February and April 2012. The study adhered to the tenets of the Declaration of Helsinki and was approved by the Office of Research Ethics at the University of Waterloo (ORE #17216). All participants provided written informed consent.

Participants were recruited consecutively from an existing patient database. Inclusion criteria required participants to be at least 18 years of age with sufficient understanding of English or German to complete study questionnaires. Key exclusion criteria included active ocular pathology, use of ocular medications, contact lens wear within 24 hours of a study visit, and pregnancy or lactation. The sample size of 35 participants was determined to provide acceptable precision for the primary ICC analyses (see Sensitivity and power analysis).

A composite symptom severity score (SSS) was derived from three validated dry eye questionnaires (OSDI, McMonnies, and Schein). Raw scores were independently standardised to z-scores, and the arithmetic mean was calculated. Participants were dichotomised into symptomatic and asymptomatic groups based on the median SSS, enabling stratified analysis of analytical performance by symptom status.

2.2 Visit protocol

Three study visits were conducted on consecutive days, each scheduled at the same time of day (±10 minutes) to minimise diurnal variation. A randomisation sequence determined the starting eye (OD or OS) and the initial

NIKBUT illumination modality (white or infrared light), maintained consistently across all visits.

At each visit, measurements were performed in a fixed sequence designed to minimise the influence of invasive procedures on subsequent non-invasive measurements. The sequence was: (1) NIKTMH under infrared illumination; (2) NIKBUT under the first randomised illumination modality; (3) completion of the OSDI questionnaire; (4) a mandatory washout period of at least 10 minutes; (5) NIKBUT under the alternate illumination modality; (6) completion of the McMonnies questionnaire; (7) slit-lamp examination for TMH assessment; (8) a second washout of at least 10 minutes; (9) fluorescein instillation for FBUT measurement.

Each objective endpoint was measured three times per eye at each visit, yielding up to 18 observations per participant per metric across the study (3 repeats × 2 eyes × 3 visits).

2.3 Measurements

Non-invasive Keratograph break-up time (NIKBUT): Measured using the Oculus Keratograph 5M (Oculus GmbH, Wetzlar, Germany). The instrument projects Placido rings onto the corneal surface and uses automated software to detect distortions in the reflected image corresponding to tear film break-up. Two metrics were recorded: NIKBUT First (time to the first detected distortion) and NIKBUT Average (mean of all break-up times detected across the cornea during the recording period). Additional parameters recorded included the break-up area (percentage of the measurement area exhibiting break-up) and the total measurement period.

Non-invasive Keratograph tear meniscus height (NIKTMH): Measured under infrared illumination using the Keratograph 5M's built-in software.

Fluorescein break-up time (FBUT): Measured at the slit lamp following standardised fluorescein instillation. A fluorescein sodium strip (1.0 mg) was moistened with one drop of non-preserved saline, shaken to remove excess fluid, and gently touched to the inferior temporal bulbar conjunctiva for approximately one second to deliver a minimal, consistent volume (estimated ≤1 µL) (3). The participant was instructed to blink naturally two to three times to distribute the dye, then to keep the eyes open as naturally as possible without squeezing. Observation was performed at ×10 magnification using broad cobalt blue illumination with a Wratten #12 yellow barrier filter to enhance tear film visualisation. The time in seconds from the last complete blink to the appearance of the first dark spot, streak, or discontinuity in the fluorescein layer was recorded using a stopwatch. Three consecutive measurements were obtained per eye, with a minimum 30-second rest period with eyes closed between measurements to allow tear film recovery (19, 20). The mean of the three recordings was used for analysis. The same examiner performed all FBUT measurements throughout the study.

Slit-lamp tear meniscus height (TMH): Measured manually at the slit lamp using a reticule measurement scale, expressed in millimetres.

2.4 Spatial allocation for sectoral analysis

For the automated NIKBUT, the Keratograph 5M maps the ocular surface to a two-dimensional matrix of 192 areas, defined by 24 angular segments and eight concentric radial zones. Radial zones were grouped into three

categories: central, paracentral, and peripheral. Each matrix area was allocated to one of twelve sectors (e.g., central-superior, paracentral-nasal), enabling spatially resolved analysis of break-up frequency and timing.

For the manual FBUT, a distinct allocation system was applied. The central corneal region was defined as a 3 mm diameter circle. The remaining surface was divided into four anatomical sectors (superior, inferior, nasal, temporal) using axes at 0°–180° and 90°–270°, with intermediate boundaries at 45° and 135°.

2.5 Statistical analysis

All analyses were performed using Python (version 3.13.7) with pandas, scipy, statsmodels, pingouin, and simpledorff libraries. A p-value of < 0.05 was considered statistically significant. This study was reported in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (21).

Missing data handling: A hybrid imputation strategy was employed. Missing NIKBUT values due to right-censoring (no break-up within the 25-second recording limit) were imputed at 25 seconds using interval-censoring (Tobit approach). Remaining missing values, presumed Missing at Random, were handled using Multiple Imputation by Chained Equations (MICE) to generate five complete datasets. Primary analyses used Rubin's rules for pooled estimates; sensitivity analyses confirmed robustness across all imputed datasets.

Preliminary analyses: Inter-eye differences were assessed using paired t-tests. Visit-to-visit stability was evaluated using repeated measures ANOVA. Illumination effects (white vs. infrared light) on NIKBUT were assessed both by paired comparisons and by linear mixed-effects models (LMM) with participant as a random effect.

Analytical performance metrics:

Intra-session precision was quantified using the coefficient of variation (CV), calculated from the three repeated measurements within each visit and averaged across visits and eyes.

Inter-session (day-to-day) reliability was assessed using the intraclass correlation coefficient (ICC 3,1), derived from a linear mixed-effects model with eye nested within participant to account for the within-subject between-eye correlation. The ICC evaluates consistency of single measurements across the three study visits. Ninety-five per cent confidence intervals for the ICC were computed and are reported alongside point estimates.

Standard error of measurement (SEM) was calculated as $SEM = SD \times \sqrt{(1 - ICC)}$, providing a metric of absolute measurement error in the original units.

Minimum detectable change at 95% confidence (MDC_{95}) was calculated as $MDC_{95} = SEM \times 1.96 \times \sqrt{2}$, representing the smallest change that exceeds measurement noise with 95% confidence.

Method comparison: Agreement between automated and conventional methods was evaluated using Bland-Altman analysis with a subject-level random intercept to account for repeated measures. In all method comparisons, the difference was calculated as conventional minus automated method (FBUT − NIKBUT; TMH − NIKTMH), such that negative bias indicates the automated method yielded higher values. Mean bias, 95%

limits of agreement (LoA), and between- and within-subject variance components were reported. Proportional bias was assessed by regressing the difference on the mean.

Spatial analysis: The frequency and location of the first tear film break-up event were analysed for both NIKBUT and FBUT, stratified by corneal zone (central, paracentral, peripheral) and by angular quadrant (superior, inferior, nasal, temporal). Mean break-up times were compared across zones using one-way ANOVA with Tukey's HSD post-hoc tests. Between-group differences within each zone were assessed using independent t-tests. Intra-subject repeatability of break-up location was quantified using Krippendorff's alpha, calculated on the corneal zone of first break-up across the three within-visit repeated measurements for each method and symptom group combination.

3. Results

3.1 Participant characteristics

The study included 35 participants, of whom 26 (74.3%) were habitual soft contact lens wearers. Based on the composite symptom severity score, 18 participants (51.4%) were classified as symptomatic and 17 (48.6%) as asymptomatic. The two groups did not differ significantly in age, sex distribution, or ethnicity (Table 1). The effect size for the symptom severity score between groups was large (r = 0.851), confirming the validity of the classification.

TABLE 1: Baseline demographic and clinical characteristics. values are presented as mean ± standard deviation [95% Confidence Interval], median [Interquartile Range], or n (%).

Characteristic	Asymptomatic (n=17)	Symptomatic (n=18)	p-value
Age (years)	32.9 ± 11.5	28.1 ± 7.4	0.145
Sex (female)	10 (58.8%)	9 (50.0%)	0.854
Contact Lens Wearers	12 (71.4%)	14 (79.1%)	0.921
Schein Score	3.0 [2.0 - 4.0]	6.0 [3.2 - 8.0]	**0.003**
OSDI Score	4.0 [2.0 - 5.0]	11.0 [6.5 - 17.0]	**< 0.001**
McMonnies Score	2.0 [1.0 - 4.0]	9.0 [6.0 - 11.0]	**< 0.001**
Symptom Severity Score (SSS)	-0.544 [-0.764 - -0.359]	0.267 [0.176 - 0.807]	**< 0.001**
FBUT (s)	9.69 ± 5.56 [6.83, 12.55]	8.53 ± 4.99 [6.05, 11.01]	0.525
Slit Lamp TMH (mm)	0.24 ± 0.05 [0.21, 0.27]	0.22 ± 0.04 [0.20, 0.24]	0.306
NIKBUT First BUT (s)	10.51 ± 4.52 [8.18, 12.83]	12.23 ± 4.58 [9.95, 14.51]	0.232
NIKBUT Average (s)	12.13 ± 4.19 [9.98, 14.29]	13.78 ± 4.02 [11.78, 15.78]	0.250
NIKBUT Area (%)	2.71 ± 1.02 [2.19, 3.24]	3.09 ± 1.66 [2.26, 3.91]	0.494
NIKBUT Max measuring time (s)	17.06 ± 4.06 [14.97, 19.15]	18.36 ± 3.59 [16.57, 20.14]	0.378
NIKTMH (mm)	0.34 ± 0.11 [0.28, 0.39]	0.31 ± 0.11 [0.26, 0.37]	0.507

3.2 Preliminary analyses

Inter-eye symmetry: Paired t-tests revealed no statistically significant differences between right and left eyes for any endpoint (smallest $p = 0.47$). Data from both eyes were therefore pooled for all subsequent analyses.

Visit-to-visit stability: Repeated measures ANOVA detected no systematic differences across the three study visits for any endpoint (smallest $p = 0.33$), excluding learning effects or fatigue as confounders and supporting aggregation across visits.

Illumination effects: Pairwise comparisons between white and infrared illumination revealed a statistically significant difference only for NIKBUT break-up area, which was larger under white light (Cohen's $d = 0.130$ in the total cohort). However, the linear mixed-effects model, which accounts for inter-subject variability and repeated measures, found no significant main effect of illumination on any NIKBUT endpoint (all $p > 0.05$). Given the absence of a systematic effect, data from both illumination modalities were pooled for all primary analyses.

3.3 Right-censoring and descriptive statistics

Of the 1,404 individual NIKBUT First measurements across all participants, visits, and eyes, 331 (23.6%) reached the 25-second recording limit without a detected break-up event and were imputed at 25 seconds (see

Missing data handling). The same proportion applied to NIKBUT Average, as both metrics share the same recording epoch. Censoring was more frequent in the symptomatic group (29.2%) than the asymptomatic group (22.2%), a counterintuitive finding that may reflect greater blink interference or recording instability in symptomatic participants, leading to more recordings that fail to capture a break-up event within the device's time window.

Mean (±SD) values for the total cohort were: NIKBUT First 10.86 ± 6.81 s, NIKBUT Average 12.50 ± 6.22 s, NIKTMH 0.32 ± 0.12 mm, FBUT 9.20 ± 6.20 s, and slit-lamp TMH 0.23 ± 0.07 mm. Symptomatic participants had lower mean FBUT (8.52 ± 6.09 vs. 9.79 ± 6.44 s) but comparable NIKTMH (0.31 ± 0.13 vs. 0.34 ± 0.12 mm). Full descriptive statistics are provided in Table 2.

3.4 Analytical performance

The analytical performance metrics are summarised in Table 2. Substantial differences in performance were observed between the measurement techniques.

Intra-session precision (CV): NIKTMH demonstrated excellent precision with a CV of 8.8% for the total cohort (where CV < 10% was considered excellent, 10–30% acceptable, and > 30% poor). In contrast, the automated NIKBUT metrics were highly variable: NIKBUT First showed a CV of 53.6%, and NIKBUT Average 42.8%. The conventional methods showed intermediate variability: slit-lamp TMH had a CV of 19.1%, and FBUT was 31.1%. Precision was generally comparable between the symptomatic and asymptomatic groups.

Inter-session reliability (ICC): Day-to-day consistency followed a similar pattern (classified per Koo and Li (22): poor < 0.50, moderate 0.50–0.75, good 0.75–0.90, excellent > 0.90). NIKTMH was the most reliable automated metric (ICC = 0.727, 95% CI: 0.580–0.840; moderate-to-good). Slit-lamp TMH demonstrated poor reliability (ICC = 0.401, 95% CI: 0.200–0.600). A notable finding was that tear stability measures showed symptom-dependent reliability: the asymptomatic group consistently demonstrated higher ICC values than the symptomatic group (e.g., NIKBUT First: asymptomatic ICC = 0.655 [0.390–0.840] vs. symptomatic ICC = 0.600 [0.330–0.810]), indicating symptom-dependent measurement variability (heteroscedastic error), whereby participants with greater symptom burden exhibited less stable measurements. Confidence intervals for the stratified ICCs were wide, reflecting the modest subgroup sample sizes (n = 17–18), and several lower bounds crossed interpretive thresholds (e.g., NIKBUT First symptomatic lower CI = 0.330, within the 'poor' range).

Measurement error (SEM and MDC_{95}): The clinically most revealing metrics were the SEM and MDC_{95} values. For FBUT, the MDC_{95} was 9.28 seconds - meaning that a change in a patient's fluorescein break-up time must exceed 9.28 seconds before it can be confidently attributed to a true clinical change rather than measurement variability. For NIKTMH, the MDC_{95} was 0.173 mm. NIKBUT First had an MDC_{95} of 8.83 seconds and NIKBUT Average 8.37 seconds.

TABLE 2: Analytical performance metrics (CV, ICC, SEM, MDC_{95}) for all endpoints, stratified by total cohort and symptom group.

Measurement	Group	Intra-session Precision (CV %)	Inter-session Reliability (ICC, 95% CI)	SEM	MDC95
FBUT	Total	31.1%	0.674 (0.510 to 0.810)	3.348	9.281
	Asymptomatic	32.8%	0.744 (0.530 to 0.890)	3.041	8.428
	Symptomatic	31.6%	0.611 (0.350 to 0.810)	3.575	9.908
Slit Lamp TMH	Total	19.1%	0.401 (0.200 to 0.600)	0.049	0.135
	Asymptomatic	18.4%	0.376 (0.070 to 0.670)	0.056	0.155
	Symptomatic	21.2%	0.431 (0.150 to 0.700)	0.040	0.112
NIKBUT First	Total	53.6%	0.629 (0.450 to 0.770)	3.186	8.830
	Asymptomatic	56.1%	0.655 (0.390 to 0.840)	2.984	8.271
	Symptomatic	52.9%	0.600 (0.330 to 0.810)	3.340	9.258
NIKBUT Average	Total	42.8%	0.602 (0.420 to 0.760)	3.019	8.367
	Asymptomatic	45.5%	0.656 (0.400 to 0.840)	2.757	7.643
	Symptomatic	41.9%	0.543 (0.260 to 0.780)	3.226	8.941
NIKTMH	Total	8.8%	0.727 (0.580 to 0.840)	0.062	0.173
	Asymptomatic	7.7%	0.746 (0.530 to 0.890)	0.058	0.161
	Symptomatic	10.4%	0.719 (0.500 to 0.870)	0.065	0.181

3.5 Method comparison

NIKBUT vs. FBUT: Bland-Altman analysis revealed poor agreement between the automated and conventional tear stability measures (Figure 1a, 1b). For NIKBUT First, the mean bias evaluated at the grant mean was −0.44 seconds in the asymptomatic group and −3.42 seconds in the symptomatic group, indicating that NIKBUT consistently recorded longer break-up times than FBUT. The 95% limits of agreement were exceptionally wide: −19.14 to 18.27 seconds (asymptomatic) and −20.10 to 13.25 seconds (symptomatic). Significant proportional bias was detected (slope = −0.39 asymptomatic; −0.52 symptomatic), indicating that the direction and magnitude of disagreement varied systematically across the measurement range, with the discrepancy between methods increasing at longer mean break-up times. For NIKBUTAverage, the pattern was similar with a larger systematic bias (−2.12 s asymptomatic; −5.02 s symptomatic), comparably wide limits of agreement, and proportional bias (slope = −0.25 asymptomatic; −0.37 symptomatic).

NIKTMH vs. TMH: Agreement between the automated and manual tear volume measures was also limited (Figure 1c). A consistent systematic bias was observed, with NIKTMH measuring approximately 0.09 mm higher than slit-lamp TMH across both groups. While the mean bias was small, the 95% limits of agreement were wide relative to the measurement scale (e.g., −0.32 to 0.13 mm for the asymptomatic group). Substantial proportional bias was evident (slope = −0.36 asymptomatic; −0.79 symptomatic), indicating that the methods diverged systematically, with NIKTMH increasingly exceeding TMH at higher tear meniscus values.

TABLE 3: Summary of Bland-Altman analysis results: bias (at grant mean), 95% LoA, between-subject SD, within-subject SD, and proportional bias slope for each method comparison, stratified by symptom group.

Comparison	Group	Bias (at Grant Mean)	95% LoA Lower	95% LoA Upper	Between-Subject SD	Within-Subject SD	Proportional Bias Slope	Units
FBUT vs NIKBUT First	Asymptomatic	-0.44	-19.14	18.27	7.71	5.62	-0.39	s
	Symptomatic	-3.42	-20.10	13.25	5.68	6.33	-0.52	s
FBUT vs NIKBUT Average	Asymptomatic	-2.12	-19.93	15.69	7.36	5.33	-0.25	s
FBUT vs NIKBUT Average	Symptomatic	-5.02	-20.82	10.78	5.29	6.08	-0.37	s
Slit Lamp TMH vs NIKTMH	Asymptomatic	-0.09	-0.32	0.13	0.09	0.07	-0.36	mm
Slit Lamp TMH vs NIKTMH	Symptomatic	-0.09	-0.29	0.12	0.07	0.07	-0.79	mm

3.6 Spatial dynamics of tear film break-up

Analysis of tear film break-up location revealed that NIKBUT and FBUT capture fundamentally different spatial phenomena (Figures 2 and 3; Table 4).

Break-up location frequency: For NIKBUT, break-up events were distributed predominantly across the paracentral cornea, with a higher incidence in the symptomatic group (62.8%) compared to the asymptomatic group (53.1%). Central break-up accounted for 46.9% (asymptomatic) and 35.3% (symptomatic) of events, with negligible peripheral involvement. In stark contrast, FBUT events were overwhelmingly concentrated in the central zone: 96.7% in asymptomatic and 85.9% in symptomatic participants. This difference in spatial distribution was consistent across angular quadrants. For NIKBUT, events were evenly distributed across inferior (28.8–28.9%), nasal (27.9–30.0%), and temporal (28.8–32.1%) sectors. For FBUT, the few non-central events were sparse and scattered, with no quadrant exceeding 5% of total events.

Break-up time by zone: Significant spatial gradients in break-up timing were observed for both methods, but with opposing patterns. For NIKBUT, central zones had the shortest break-up times, while paracentral zones (particularly nasal and temporal) had the longest (ANOVA $p < 0.001$ in both groups). In the asymptomatic group, central NIKBUT was significantly shorter than paracentral inferior (mean difference: 3.11 s, $p = 0.003$), nasal (4.09 s, $p < 0.001$), and temporal (4.33 s, $p < 0.001$). In symptomatic patients, paracentral nasal NIKBUT was significantly longer than both central (3.97 s, $p < 0.001$) and paracentral inferior (4.04 s, $p < 0.001$). For FBUT, spatial differences were significant only in the symptomatic group (ANOVA $p < 0.001$), where the

superior zone exhibited markedly shorter break-up times than the central zone (mean difference: −6.47 s, $p < 0.001$), and central break-up time was significantly longer than temporal (mean difference: 3.91 s, $p = 0.003$).

Intra-subject repeatability of break-up location: Despite these clear group-level patterns, the intra-subject repeatability of the break-up location was very poor. Krippendorff's α ranged from 0.115 to 0.308 across methods (NIKBUT and FBUT) and symptom groups (asymptomatic and symptomatic), where α was calculated on the corneal zone of first break-up across the three within-visit repeated measurements. All values fell well below the conventional threshold of 0.667 for acceptable reliability, indicating that the specific corneal zone of first break-up is not a stable individual characteristic but rather a stochastic event, highly sensitive to micro-variations in blink completeness and local tear composition.

TABLE 4: Spatial distribution of tear film break-up by corneal zone and symptom group.

Values are presented as mean ± SD. NIKBUT frequency represents the percentage of segments within each zone that exhibited first break-up per eye, averaged across eyes (zone sizes: Central = 72 segments, Paracentral = 72, Peripheral = 48, each angular quadrant = 48, out of 192 total).

Zone Category	Radial			Angular				Total
Zone	Central	Paracentral	Peripheral	Superior	Inferior	Nasal	Temporal	
Asymptomatic NIKBUT n (%)	6.0 ± 3.1	6.8 ± 3.1	0	3.0 ± 3.3	6.2 ± 3.7	4.7 ± 3.4	5.3 ± 4.1	6.0 ± 3.1
Asymptomatic NIKBUT (s)	8.11 ± 4.87	11.74 ± 6.34	-	9.71 ± 6.95	10.12 ± 6.02	9.21 ± 7.21	9.26 ± 6.71	8.11 ± 4.87
Asymptomatic FBUT n (%)	100	5.1 ± 1.2 *	0	8.8 ± 2.9	8.8 ± 2.9 *	2.9 ± 1.7 *	0 *	100
Asymptomatic FBUT (s)	9.68 ± 5.52	10.00 ± 6.71	-	11.97 ± 7.06	11.95 ± 10.93	7.96		9.68 ± 5.52
Symptomatic NIKBUT n (%)	5.2 ± 2.6	7.8 ± 3.7	3.0 ± 9.0	1.9 ± 2.0	6.2 ± 4.0	5.8 ± 4.5	6.0 ± 4.1	5.2 ± 2.6
Symptomatic NIKBUT (s)	9.95 ± 6.30	12.88 ± 6.66	23.51 ± 2.98	10.73 ± 8.91	11.61 ± 7.40	10.73 ± 6.92	11.57 ± 6.93	9.95 ± 6.30
Symptomatic FBUT n (%)	100	18.8 ± 28.3 *	0	11.1 ± 31.9	27.8 ± 45.4 *	19.4 ± 40.1 *	16.7 ± 37.8 *	100
Symptomatic FBUT (s)	8.51 ± 5.21	6.52 ± 2.95	-	6.64 ± 3.13	7.53 ± 6.04	7.88 ± 7.03	4.82 ± 1.43	8.51 ± 5.21

**FBUT (n) Paracentral - Asymptomatic vs. Symptomatic: p = 0.041*

FBUT (n) Inferior - Asymptomatic vs. Symptomatic: p = 0.043

FBUT (n) Nasal - Asymptomatic vs. Symptomatic: p = 0.032

FBUT (n) Temporal - Asymptomatic vs. Symptomatic: p = 0.014

Each eye contributed up to 9 measurements (3 visits × 3 repetitions), with unique break-up segments identified from f-ring/s-segment coordinates. FBUT frequency represents the percentage of eyes in which the clinician-observed first break-up occurred in that zone at least once across all measurements. Break-up time (s) is the mean first break-up time for eyes with data in that zone, computed as the per-eye average then averaged across eyes. Radial zones classify break-up by distance from the corneal apex; angular quadrants classify by meridian (nasal/temporal mirrored for OD/OS). P-values are from Mann-Whitney U tests comparing asymptomatic versus symptomatic groups. NIKBUT: non-invasive keratograph break-up time (n = 32 asymptomatic, 36 symptomatic eyes); FBUT: fluorescein break-up time (n = 34 asymptomatic, 36 symptomatic eyes).

3.7 Sensitivity and power analysis

The sample size of 35 participants, each completing three visits with three measurements per eye per visit (up to 18 observations per participant per metric), was determined to provide acceptable precision for the primary ICC analyses. With k = 3 measurement occasions and n = 35, the expected width of the 95% confidence interval for a true ICC of 0.70 is approximately ±0.16 (23). For the primary ICC estimates, a post-hoc sensitivity calculation confirms that this design provides 80% power to reject the null hypothesis that the ICC ≤ 0.40 when the true ICC is ≥ 0.65, using a one-sided test at α = 0.05. Stratified analyses by symptom group (n ≈ 17–18 per group) have reduced precision, with expected CI widths of approximately ±0.25 for a true ICC of 0.70; these should therefore be interpreted as hypothesis-generating rather than definitive. For the Bland-Altman method-comparison analyses, with approximately 210 paired observations per group (3 visits × 2 eyes × 35 participants), the study had sufficient power to estimate mean bias with a standard error of less than 1 second for NIKBUT and FBUT comparisons, and less than 0.01 mm for TMH comparisons. Sensitivity analyses were conducted to assess the robustness of the primary findings to the imputation strategy for right-censored NIKBUT values: results were consistent across all five MICE-generated datasets and when censored observations were excluded entirely, confirming that the imputation approach did not materially influence the reported precision or reliability estimates.

4. Discussion

This study provides a comprehensive characterisation of the analytical performance of the Oculus Keratograph 5M for non-invasive tear film assessment. The results reveal a striking divergence between volume- and stability-oriented endpoints, with direct implications for how clinicians should interpret these measurements in practice.

Volume assessment: NIKTMH as a monitoring candidate - Among all endpoints evaluated, NIKTMH demonstrated the strongest analytical profile. Its intra-session precision was excellent (CV = 8.8%), and its inter-session reliability was moderate-to-good (ICC = 0.727), substantially outperforming both NIKBUT metrics

and slit-lamp TMH. The resulting MDC_{95} of 0.173 mm provides a concrete clinical benchmark: changes in NIKTMH smaller than this value should not be interpreted as evidence of treatment response or disease progression. This degree of precision is compatible with longitudinal monitoring of tear volume, provided clinicians apply the MDC_{95} threshold when interpreting sequential measurements. These findings align with recent evidence indicating better repeatability for Keratograph-derived TMH than for many stability indices (13-15), and are consistent with the growing consensus that NIKTMH provides clinically useful volumetric information (12). Recent work by Wolffsohn, Ayaz (24) has further demonstrated that TMH measurement using the Keratograph is sensitive to illumination type and timing post-blink, emphasising the importance of standardised measurement protocols.

However, the method comparison against slit-lamp TMH revealed a systematic bias (~0.09 mm) and wide limits of agreement with proportional bias, confirming that the two methods are not interchangeable. Clinicians should not substitute NIKTMH for slit-lamp TMH (or vice versa) when absolute values inform clinical decisions. It is also important to note that the gold standard for objective TMH assessment has shifted toward anterior segment OCT over the past decade (11). Early comparisons suggested that the Keratograph underestimated TMH relative to OCT, although a recent study reported improved agreement, likely reflecting software refinements (25). The present data, collected in 2012, should be interpreted with awareness that subsequent Keratograph software updates may have altered measurement characteristics.

Stability assessment: the precision problem - The NIKBUT metrics presented a fundamentally different analytical profile. Both NIKBUT First (CV = 53.6%) and NIKBUT Average (CV = 42.8%) showed poor intra-session precision, and their inter-session reliability was moderate at best and dependent on symptom status. The resulting MDC_{95} values - 8.83 seconds for NIKBUT First and 8.37 seconds for NIKBUT Average - indicate that large changes are required before clinicians can attribute a shift in NIKBUT to biology rather than noise.

The symptom-dependent heteroscedasticity observed - where reliability was lower in the symptomatic group - warrants particular attention. A plausible mechanistic explanation is that altered blink behaviour and sensory gain in symptomatic DED produce inconsistent post-blink tear film spread and unstable inter-trial conditions, inflating measurement variance (26, 27). This means that the patients in whom clinicians most need reliable monitoring - those with active symptoms - are precisely the group in which NIKBUT measurements are least stable. Paradoxically, this variability may itself carry clinical information: when patients report substantial symptoms yet exhibit inconsistent NIKBUT measurements across closely spaced sessions, this instability may serve as an indirect clinical cue warranting further evaluation rather than a signal to repeat the test (24, 28).

For FBUT, the MDC_{95} of 9.28 seconds has immediate practical implications. A clinician observing a change from 4 to 8 seconds in a patient's FBUT - which might intuitively seem meaningful - cannot confidently attribute this to clinical improvement, as it falls within the measurement noise. This finding accords with TFOS DEWS III guidance, which cautions that invasive methods can perturb the construct under measurement and that tear film tests differ substantially in repeatability (2, 24).

It is important to distinguish the MDC_{95} from the minimum clinically important difference (MCID). The MDC_{95} is a measurement property reflecting the instrument's noise floor - the smallest change detectable above random error. The MCID, by contrast, is a patient-centred threshold representing the smallest change that patients or clinicians perceive as meaningful. A treatment effect must exceed the MDC_{95} to be considered real, and must

additionally exceed the MCID to be considered clinically worthwhile. The present study establishes only the former; determination of MCIDs for these metrics would require anchoring against patient-reported outcomes in longitudinal treatment studies. For illustration, if future research determines that the MCID for FBUT is approximately 3 seconds - the threshold at which patients or clinicians perceive a meaningful change - this would be substantially smaller than the MDC_{95} of 9.28 seconds reported here, implying that FBUT may be fundamentally unable to detect individually meaningful changes at the single-patient level. Conversely, a change in NIKTMH from 0.20 to 0.40 mm (a 0.20 mm increase) would exceed the MDC_{95} of 0.173 mm and could therefore be interpreted as a genuine change rather than noise.

Non-interchangeability: method comparison evidence and spatial mechanism - The Bland-Altman analyses provide definitive evidence against the interchangeability of automated and conventional tear film metrics. For tear stability, the limits of agreement between NIKBUT and FBUT exceeded ±19 seconds in both symptom groups, with significant proportional bias indicating that the magnitude of discrepancy between methods varied systematically across the measurement range. The overall direction of bias (NIKBUT longer than FBUT) is consistent with the established mechanism whereby fluorescein instillation destabilises the tear film, artificially shortening the measured break-up time (3). These findings add to the growing consensus from multiple devices and populations that non-invasive and invasive break-up time methods provide related but distinct information (7, 8, 29, 30).

The spatial analysis provides a mechanistic explanation for this quantitative non-agreement and extends the work of Guarnieri, Carnero (17), who demonstrated that the Keratograph 5M can characterise the spatial distribution of automated tear break-up but examined only non-invasive patterns in a glaucoma population. The present study adds the critical comparison between automated and conventional spatial patterns in a dry-eye-relevant cohort. FBUT events were overwhelmingly concentrated in the central cornea (85.9–96.7%), while NIKBUT events were distributed predominantly across the paracentral zones (53–63%). This spatial divergence is not simply a measurement artefact - it reflects the two methods sampling fundamentally different aspects of tear film physiology. Central FBUT may reflect aqueous-deficient or mucin-deficient thinning, where fluorescein pooling in the thinnest central region produces an early visible dry spot (3). In contrast, the paracentral predominance of NIKBUT events could indicate detection of localised wettability deficits or lipid spreading irregularities - features that would be more consistent with evaporative mechanisms (19, 31). However, the present study did not include lipid layer interferometry, meibography, or DED subtype classification, so these mechanistic interpretations remain speculative. Studies incorporating lipid layer assessment alongside spatial break-up mapping are needed to test this hypothesis directly.

The observation that symptomatic participants showed a greater proportion of paracentral NIKBUT events (62.8% vs. 53.1%) and reduced central FBUT concentration (85.9% vs. 96.7%) compared to asymptomatic participants suggests that symptom burden may be associated with more diffuse tear film instability. Furthermore, the significant zonal gradients in break-up timing - with NIKBUT breaking up fastest centrally but FBUT (in symptomatic patients) breaking up fastest superiorly - reinforce that these methods are interrogating distinct biophysical processes rather than providing redundant measurements of the same construct.

However, the clinical applicability of spatial analysis at the individual level is currently limited. Intra-subject repeatability of break-up location was very poor (Krippendorff's $\alpha = 0.115–0.308$), indicating that the location of first break-up is a stochastic event. While spatial analysis is therefore a valuable research tool for

understanding disease heterogeneity at the group level, its high variability makes it unsuitable as a biomarker for individual patient monitoring.

Taken together, the Bland-Altman data demonstrate that NIKBUT and FBUT cannot be substituted for one another, and the spatial analysis reveals why: these methods are not measuring the same construct. Clinicians should treat them as complementary sources of information about tear film health, and should not compare NIKBUT values against FBUT-derived diagnostic cut-offs or vice versa.

Strengths and limitations: The principal strength of this study is its prospective, repeated-measures design with three visits on consecutive days at standardised times, enabling robust quantification of both intra- and inter-session variability while controlling for diurnal effects. The use of multiple complementary analytical metrics (CV, ICC, SEM, MDC_{95}) provides a more complete picture than studies reporting only repeatability coefficients or ICC alone.

The temporal context of data collection (2012) warrants careful interpretation. For the spatial dynamics findings, this limitation is less consequential than it might appear: spatial break-up patterns are determined primarily by Placido ring geometry and corneal surface optics, which are hardware-dependent properties unchanged between 2012 and the present. The precision benchmarks (CV, ICC, SEM, MDC_{95}), which depend on both hardware resolution and software-based break-up detection algorithms, may be more susceptible to software evolution. Recent evidence suggests improving agreement between NIKTMH and OCT with contemporary software (25), and Su, Yu (32) reported higher intrasession ICC values for NIBUTf (0.89) using current Keratograph software than those observed in the present study (0.629), consistent with the hypothesis that software evolution has improved measurement precision. The MDC_{95} thresholds reported here should therefore be regarded as conservative upper-bound estimates; contemporary software may yield tighter precision. Validation studies using current Keratograph software are recommended to confirm or update these benchmarks.

The 23.6% right-censoring rate for NIKBUT measurements is a methodological consideration that warrants acknowledgement. Because censored values were imputed at the 25-second device cap, the upper tail of the NIKBUT distribution was compressed, potentially underestimating the true variance and therefore the MDC_{95}. A Tobit regression approach was used within the imputation framework, but readers should be aware that the true measurement variability for NIKBUT may be larger than reported here.

The sample size (n = 35), while modest, provided acceptable precision for the primary ICC analyses (expected CI width ±0.16 for ICC = 0.70 with k = 3 (23)), and the repeated-measures design (up to 18 observations per participant per metric) yields stable variance estimates. However, the single-centre design and high proportion of habitual contact lens wearers (74%) represent important limitations. Contact lens wear is associated with altered tear film stability, modified blink dynamics, and potentially different spatial break-up characteristics (33). The MDC_{95} thresholds reported here may therefore not be directly transferable to non-lens-wearing populations, and replication in a predominantly non-CL cohort is recommended; future studies should establish separate benchmarks for CL and non-CL populations. Additionally, the FBUT precision estimates reflect a single trained examiner; in multi-examiner clinical settings, inter-observer variability would likely increase the MDC_{95} beyond the 9.28 seconds reported here. Future studies should include larger, more diverse cohorts from different clinical settings and age groups.

Finally, the symptom classification used a median-split approach on composite z-scores, which oversimplifies the continuous nature of DED symptomatology and may obscure gradations in the sign-symptom relationship. This limitation primarily affects the stratified analyses rather than the overall analytical performance estimates.

5. Conclusions

This study establishes device-specific analytical performance benchmarks for the Oculus Keratograph 5M and provides a mechanistic explanation for the non-interchangeability of automated and conventional tear film methods. NIKTMH is the most analytically robust automated metric, with excellent precision and moderate-to-good day-to-day reliability, making it a viable candidate for longitudinal tear volume monitoring when changes are interpreted against the MDC_{95} of 0.173 mm. In contrast, NIKBUT metrics show poor precision and symptom-dependent reliability, meaning only large changes can be confidently attributed to biology rather than measurement noise. FBUT, despite its long clinical history, has an MDC_{95} of 9.28 seconds, indicating that commonly observed clinical changes may fall within the noise floor. Spatial analysis confirms that NIKBUT and FBUT capture fundamentally different phenomena - paracentral versus central break-up - reinforcing that these methods are complementary rather than substitutable. While spatial patterns offer mechanistic insight at the group level, the poor intra-subject repeatability of break-up location limits its clinical utility as an individual biomarker. Clinicians should interpret sequential tear film measurements in the context of these benchmarks and avoid over-interpreting changes that fall within the MDC_{95}.

Declarations

Ethics approval: All procedures involving human participants were performed in accordance with relevant institutional guidelines and with the principles of the Declaration of Helsinki. Ethical approval for the study was obtained from the Office of Research Ethics of the University of Waterloo Office of Research Ethics (ORE #17216). Written informed consent was obtained from all participants prior to participation. The privacy rights of participants were fully observed.

Funding: This work was supported by Oculus GmbH, Wetzlar, Germany.

Conflicts of interest: The author declares the following competing interests. D.O. was a consultant with Oculus Optikgeräte GmbH at the time of data collection. The funder had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to ethical and privacy considerations.

Consent to Participate declaration: Informed consent was obtained from all individual participants included in the study.

Consent to Publish declaration: Not applicable.

Author contributions: D.O. was responsible for Conceptualisation, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing - Original Draft Preparation, Writing - Review and Editing, Visualisation, Supervision, Project Administration, and Funding Acquisition. The author has read and agreed to the published version of the manuscript.

References

1. Wolffsohn JS, Benitez-Del-Castillo JM, Loya-Garcia D, Inomata T, Iyer G, Liang L, et al. TFOS DEWS III: Diagnostic Methodology. Am J Ophthalmol. 2025;279:387–450.
2. Stapleton F, Argueso P, Asbell P, Azar D, Bosworth C, Chen W, et al. TFOS DEWS III: Digest. Am J Ophthalmol. 2025;279:451–553.
3. Mooi JK, Wang MTM, Lim J, Muller A, Craig JP. Minimising instilled volume reduces the impact of fluorescein on clinical measurements of tear film stability. Cont Lens Anterior Eye. 2017;40(3):170–4.
4. Tian L, Qu JH, Zhang XY, Sun XG. Repeatability and Reproducibility of Noninvasive Keratograph 5M Measurements in Patients with Dry Eye Disease. J Ophthalmol. 2016;2016:8013621.
5. Oehring D, Sickenberger W, editors. Prospective Study to Compare Two Different Kinds of Illuminations by Measuring the Non-Invasive Tear Film Break-Up Time by Means of a Novel Video Topographer. American Academy of Optometry Annual Meeting; 2014 2014–11; Denver, CO.
6. Cox SM, Nichols KK, Nichols JJ. Agreement between Automated and Traditional Measures of Tear Film Breakup. Optom Vis Sci. 2015;92(9):e257–63.
7. Lim J, Wang MTM, Craig JP. Evaluating the diagnostic ability of two automated non-invasive tear film stability measurement techniques. Cont Lens Anterior Eye. 2021;44(4):101362.
8. Szczesna-Iskander DH, Llorens-Quintana C. Agreement between invasive and noninvasive measurement of tear film breakup time. Sci Rep. 2024;14(1):3852.
9. Wang MTM, Craig JP. Comparative Evaluation of Clinical Methods of Tear Film Stability Assessment: A Randomized Crossover Trial. JAMA Ophthalmol. 2018;136(3):291–4.
10. Sutphin JE, Ying GS, Bunya VY, Yu Y, Lin MC, McWilliams K, et al. Correlation of Measures From the OCULUS Keratograph and Clinical Assessments of Dry Eye Disease in the Dry Eye Assessment and Management Study. Cornea. 2022;41(7):845–51.
11. Chen M, Wei A, Xu J, Zhou X, Hong J. Application of Keratograph and Fourier-Domain Optical Coherence Tomography in Measurements of Tear Meniscus Height. J Clin Med. 2022;11(5):1343.
12. Soares I, Ramalho E, Brardo FM, Nunes AF. Tear meniscus height agreement and reproducibility between two corneal topographers and spectral-domain optical coherence tomography. Clin Exp Optom. 2025;108(4):430–6.
13. Garcia-Marques JV, Martinez-Albert N, Talens-Estarelles C, Garcia-Lazaro S, Cervino A. Repeatability of Non-invasive Keratograph Break-Up Time measurements obtained using Oculus Keratograph 5M. Int Ophthalmol. 2021;41(7):2473–83.
14. Garcia-Montero M, Rico-Del-Viejo L, Lorente-Velazquez A, Martinez-Alberquilla I, Hernandez-Verdejo JL, Madrid-Costa D. Repeatability of Noninvasive Keratograph 5M Measurements Associated With Contact

Lens Wear. Eye Contact Lens. 2019;45(6):377–81.

15. Yin Chan K, Liao X, Guo B, Tse JSH, Li PH, Cheong AMY, et al. Ocular surface parameters repeatability and agreement -A comparison between Keratograph 5M and IDRA. Cont Lens Anterior Eye. 2024;47(6):102281.
16. Dumpati S, Kumar M, Vijay AK, Tan J, Willcox M. Comparison of different methods to image the tear film before and during contact lens wear. Cont Lens Anterior Eye. 2026;49(1):102511.
17. Guarnieri A, Carnero E, Bleau AM, Lopez de Aguileta Castano N, Llorente Ortega M, Moreno-Montanes J. Ocular surface analysis and automatic non-invasive assessment of tear film breakup location, extension and progression in patients with glaucoma. BMC Ophthalmol. 2020;20(1):12.
18. Yokoi N, Georgiev GA, Kato H, Komuro A, Sonomura Y, Sotozono C, et al. Classification of Fluorescein Breakup Patterns: A Novel Method of Differential Diagnosis for Dry Eye. Am J Ophthalmol. 2017;180(9):72–85.
19. King-Smith PE, Begley CG, Braun RJ. Mechanisms, imaging and structure of tear film breakup. Ocul Surf. 2018;16(1):4–30.
20. Speakman S, Wang MTM, Muntz A, Vidal-Rohr M, Menduni F, Dhallu S, et al. Investigating the diagnostic utility of non-invasive tear film stability and breakup parameters: A prospective diagnostic accuracy study. Ocul Surf. 2022;25:72–4.
21. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud. 2011;48(6):661–71.
22. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.
23. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med. 2002;21(9):1331–5.
24. Wolffsohn JS, Ayaz M, Bandlitz S, von der Hoh F, Ebner A, Craig JP. Optimising the methodology for assessing tear meniscus height using digital imaging. Cont Lens Anterior Eye. 2025;48(4):102419.
25. Yik ALP, Barodawala FS. Tear meniscus height comparison between AS-OCT and Oculus Keratograph(R) K5M. Rom J Ophthalmol. 2024;68(4):398–403.
26. Oganov A, Yazdanpanah G, Jabbehdari S, Belamkar A, Pflugfelder S. Dry eye disease and blinking behaviors: A narrative review of methodologies for measuring blink dynamics and inducing blink response. Ocul Surf. 2023;29:166–74.
27. Zheng Q, Wang L, Wen H, Ren Y, Huang S, Bai F, et al. Impact of Incomplete Blinking Analyzed Using a Deep Learning Model With the Keratograph 5M in Dry Eye Disease. Transl Vis Sci Technol. 2022;11(3):38.
28. Belmonte C, Nichols JJ, Cox SM, Brock JA, Begley CG, Bereiter DA, et al. TFOS DEWS II pain and sensation report. Ocul Surf. 2017;15(3):404–37.
29. Pflugfelder SC, Kikukawa Y, Tanaka S, Kosugi T. The utility of software-detected non-invasive tear break-up in comparison to fluorescein tear break-up measurements. Front Med (Lausanne). 2024;11:1351013.
30. Tashbayev B, Badian RA, Chen X, Vitelli V, Lagali N, Dartt D, et al. Comparison of non-invasive and fluorescein tear film break-up time in a 65-year-old Norwegian population: a cross-sectional study. BMJ Open. 2025;15(4):e090305.

31. Willcox MDP, Argueso P, Georgiev GA, Holopainen JM, Laurie GW, Millar TJ, et al. TFOS DEWS II Tear Film Report. Ocul Surf. 2017;15(3):366–403.
32. Su L, Yu T, Chen J, Yang F, Zhang Q, Xu L, et al. Agreement and repeatability of ocular surface function using the S390L Firefly WDR slitlamp compared with Keratograph 5M. Sci Rep. 2025;15(1):34992.
33. Nichols JJ, Nichols KK, Puent B, Saracino M, Mitchell GL. Evaluation of tear film interference patterns and measures of tear break-up time. Optom Vis Sci. 2002;79(6):363–9.

Figures

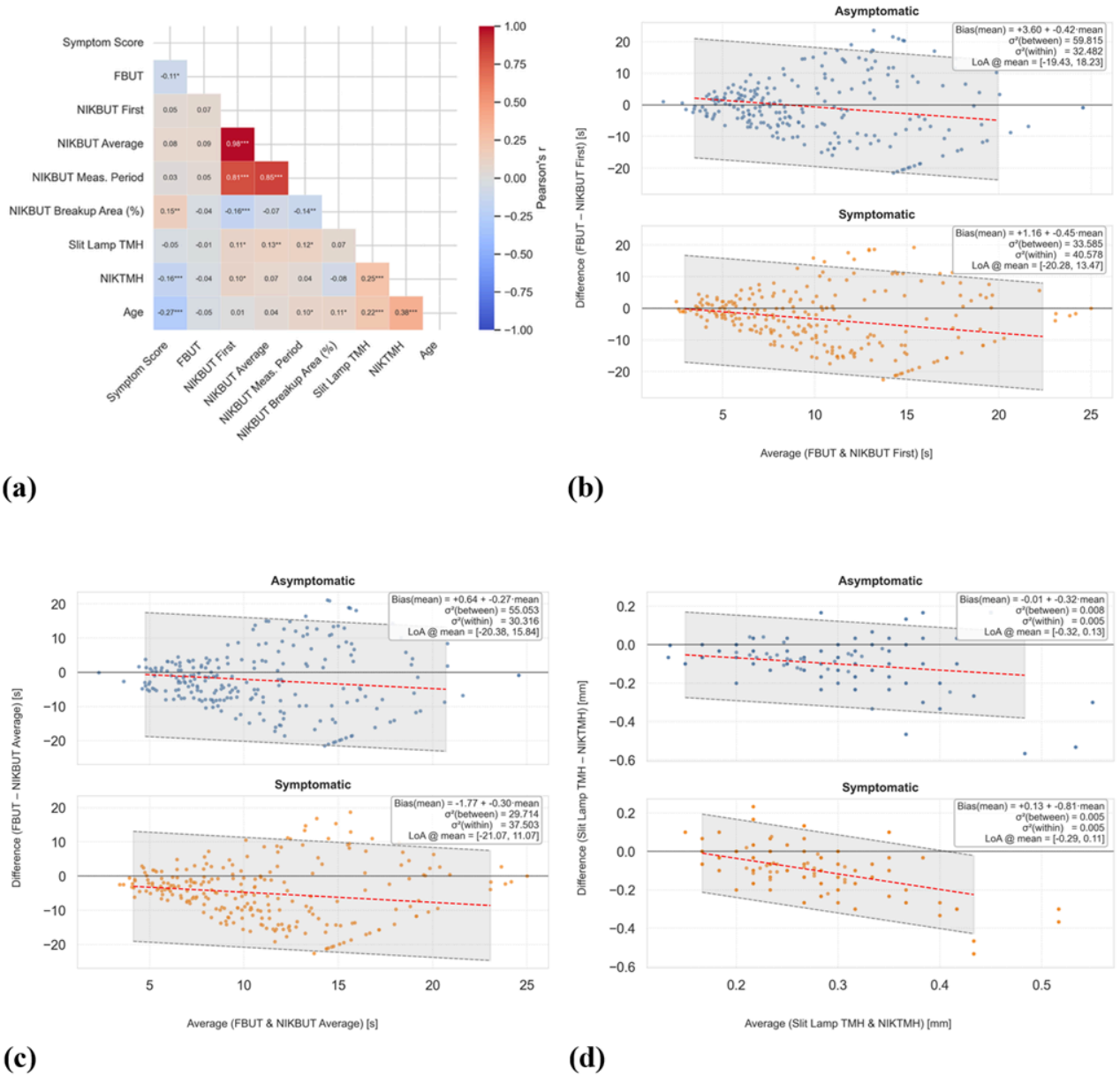


Figure 1

Method comparison demonstrating poor correlation and agreement between automated and manual clinical measures. (a) Correlation matrix of key subjective and objective endpoints. (b–d) Bland-Altman plots assessing agreement between methods. The solid line represents the mean bias, and the dashed lines represent the 95% limits of agreement. (b) Comparison of NIKBUT First and FBUT. (c) Comparison of NIKBUT Average and FBUT. Both plots show that NIKBUT measurements are systematically longer than FBUT, with wide and clinically unacceptable limits of agreement. (d) Comparison of NIKTMH and slit-lamp TMH, showing systematic bias and wide limits of agreement.

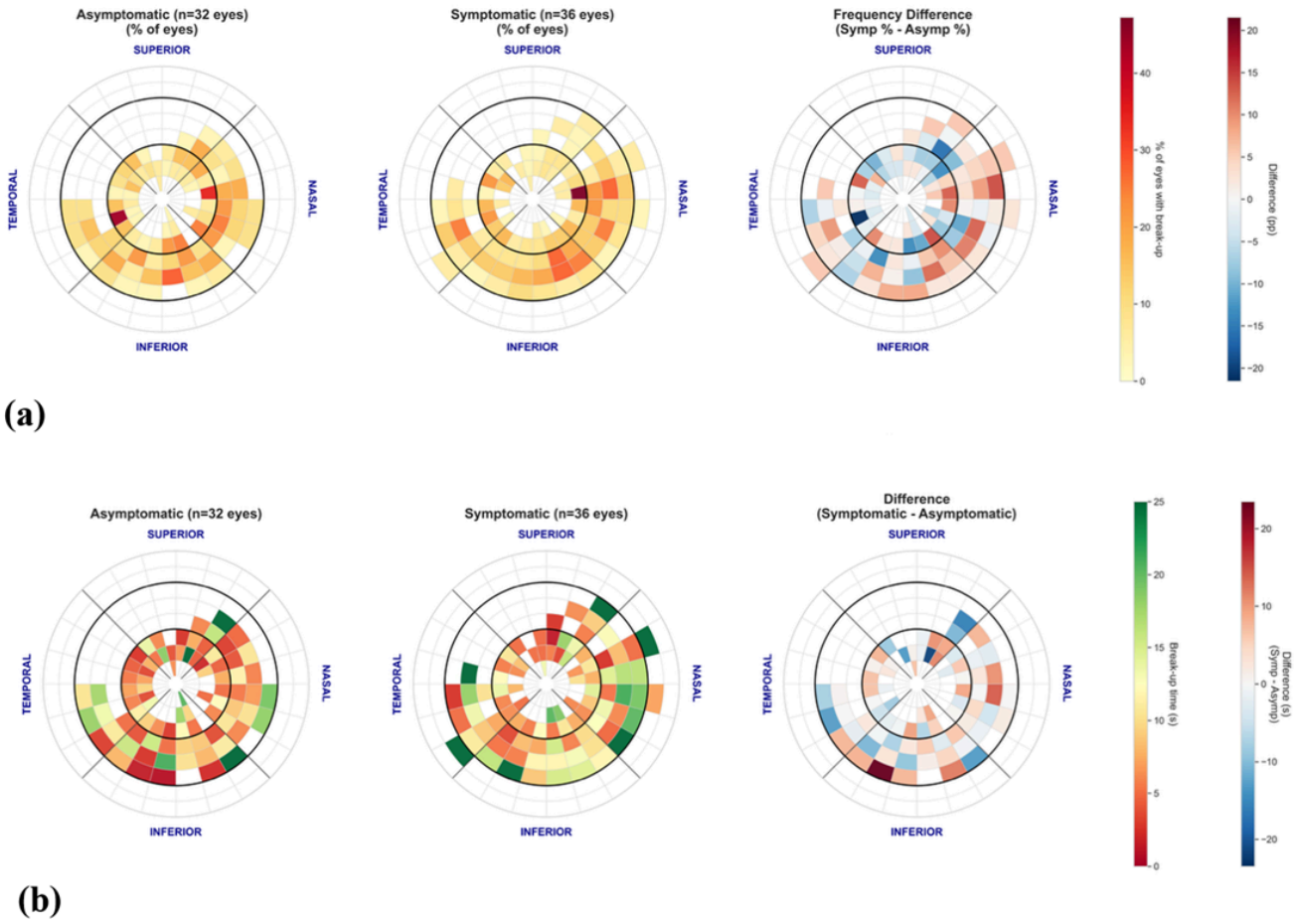


Figure 2

Spatial distribution and duration of tear film break-up events for the automated NIKBUT method. (a) Polar plot showing the frequency and location of the first break-up event. Note the distinct pattern where events are widely distributed in the paracentral cornea. (b) Mean break-up time (seconds) by corneal zone, stratified by symptom group. Break-up times were consistently shortest in the central zone and longest in the paracentral regions (e.g., nasal and temporal).

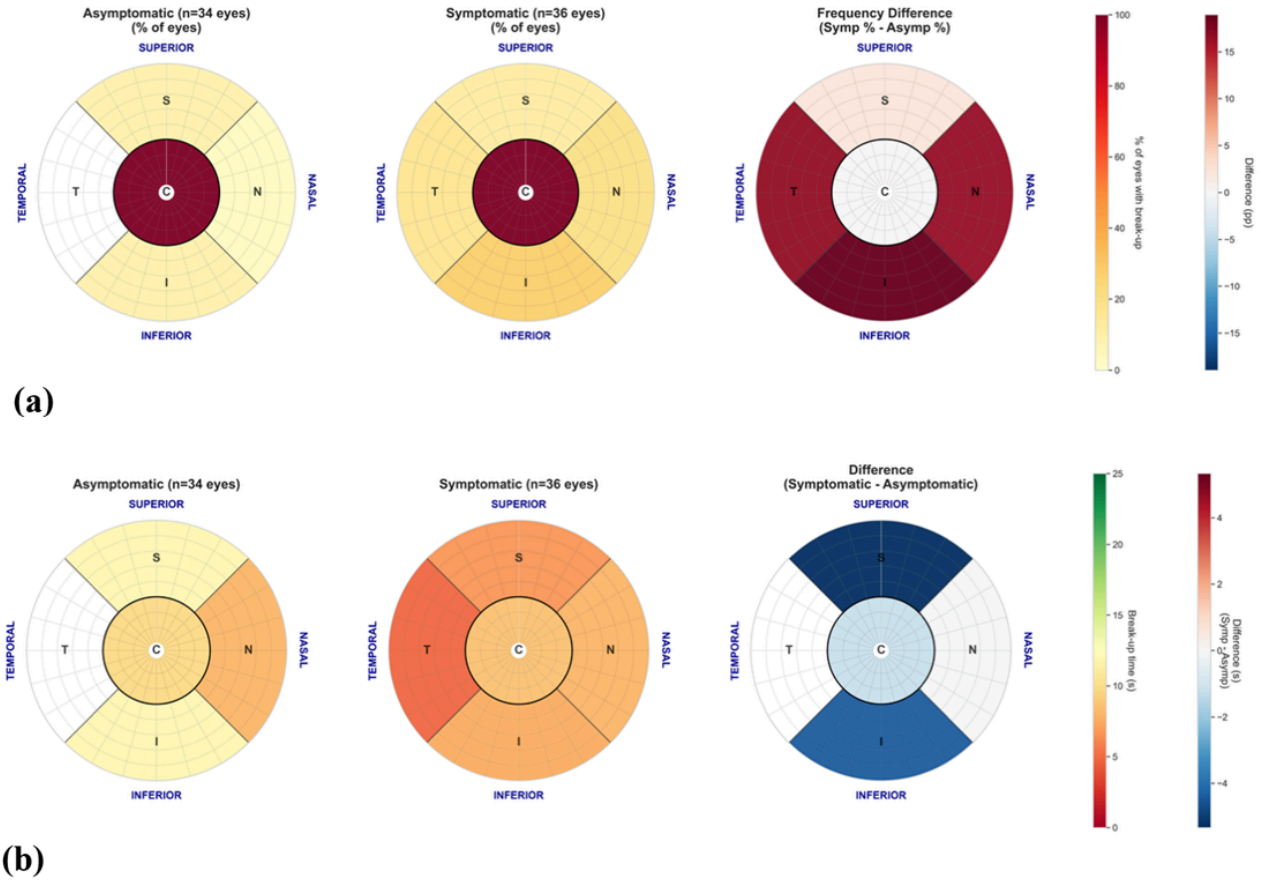


Figure 3

Spatial distribution and duration of tear film break-up events for the manual FBUT method. (a) Polar plot showing the frequency and location of the first break-up event. In contrast to NIKBUT, FBUT events are heavily concentrated in the central zone for both groups. (b) Mean break-up time (seconds) by corneal zone, stratified by symptom group. Symptomatic patients showed significantly shorter break-up times in the superior zone compared to asymptomatic patients.