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.

Illuminance-stratified calibration of open-field eccentric photorefraction: model development and independent four-device evaluation

Yan Huang[1]
Beijing Tongren Hospital
Hong Chang
He Eye Hospital
Gao Jie[1]
Beijing Tongren Hospital
Zhen Yi
dr_zhenyi@163.com

Beijing Tongren Hospital

Research Article

Keywords: eccentric photorefraction, ambient illuminance, refractive error, calibration, open-field autorefraction, method comparison

Posted Date: August 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10746211/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

To develop an illuminance-stratified calibration for open-field eccentric photorefraction and assess measurement availability and refractive stability at 50 and 500 lux. i.Profiler plus was the primary independent comparator and NIDEK AR-1 the secondary comparator.

Methods

The study comprised model development and independent clinical evaluation. The development dataset contained 1,065 eyes. Single-frame infrared images and NIDEK AR-1 measurements were obtained at 50 and 500 lux. Data were divided by eye into training (70%), tuning (20%) and internal test (10%) sets, with paired records kept in one partition. After model locking, 50 newly recruited adults underwent randomised four-device testing at both illuminances; three readings were averaged. Co-primary outcomes were measurement success at 500 lux and reference-adjusted illuminance-related drift in power vector M.

Results

At 500 lux, the calibrated system, NIDEK AR-1 and i.Profiler plus returned measurements in all 50 participants; Spot VS100 returned results in 35 (70%; exact McNemar $p = 0.000061$). Relative to i.Profiler plus, adjusted M drift was − 0.021 D (95% CI − 0.306 to 0.264) for the calibrated system and + 0.574 D (0.264 to 0.885) for Spot VS100. Relative to NIDEK AR-1, estimates were − 0.149 D (− 0.407 to 0.109) and + 0.450 D (0.144 to 0.756). In 35 complete pairs, M mean absolute error relative to i.Profiler plus was 0.609 and 1.191 D, respectively ($p < 0.001$). Individual agreement remained poorer than with tabletop autorefraction.

Conclusion

Illuminance-stratified calibration preserved measurement availability and reduced systematic illuminance-related displacement at 500 lux. The findings support wider usable room-lighting conditions, not interchangeability with tabletop autorefraction or wavefront aberrometry.

Key Points

• Open-field eccentric photorefraction can lose measurement availability or develop systematic displacement under routine room lighting; Spot VS100 returned a result in 70% of participants at 500 lux.

• An illuminance-stratified calibration developed in 1,065 eyes maintained 100% measurement availability and showed no detectable reference-adjusted spherical-equivalent drift in an independent cohort.

• Individual agreement remained poorer than with tabletop autorefraction or wavefront aberrometry; the calibrated system extends usable lighting conditions rather than replacing conventional refraction.

Introduction

Uncorrected refractive error remains a leading cause of avoidable visual impairment. In settings where trained eye-care personnel and conventional equipment are scarce, access to spectacle prescribing and dispensing remains limited[1–3]. Portable objective refraction that is rapid and less dependent on operator expertise may therefore be useful in school screening, community eye care, primary care and teleophthalmology[4,5].

Tabletop autorefractors and wavefront aberrometers control the measurement environment through short working distances, enclosed or semi-enclosed viewing channels, active infrared illumination, alignment systems and dedicated detectors. These instruments use different optical principles and should not be treated as equivalent, but both are less directly exposed to room lighting than open-field distance photorefraction. Eccentric photorefraction uses an off-axis infrared source to produce a spatial luminance distribution across the pupil after fundus reflection. The direction and magnitude of the reflex contain information about refractive state[6,7]. The method is rapid and non-contact, but image formation is susceptible to ambient illuminance, pupil constriction, eyelid and eyelash obstruction, off-axis positioning, detector exposure and changes in signal-to-noise ratio.

Commercial photoscreeners are used to identify amblyopia risk factors and clinically relevant refractive abnormalities, particularly in children and other patients for whom conventional testing is difficult[8–10]. Most evaluations, however, have been performed under controlled or relatively dim conditions, and manufacturers commonly recommend low room illumination and an adequate pupil diameter. In routine clinics, classrooms and community settings, room lights are usually kept on for safe working and communication. Requiring dim illumination can increase repeat measurements, prolong testing and lead to failure to obtain a result.

The effect of ambient illuminance can be considered an input-domain shift. Brighter room lighting alters pupil size and the geometry, contrast and intensity distribution of the infrared reflex. Algorithms based on fixed crescent geometry, threshold segmentation or a calibration curve may then receive images that differ systematically from those used to establish the original mapping. Two failures are possible: no result is returned, or a result is returned with systematic displacement. These outcomes require separate assessment. Agreement among successful measurements alone does not capture selective failure and does not establish stability after the measurement environment changes.

Convolutional neural networks have been used to improve pupil localisation and refractive prediction from photorefraction images[11–13]. The previously reported Tongren Digital Sight platform used open-field eccentric photorefraction and image-based prediction of refractive state[14]. That work focused on general clinical performance and did not develop an illuminance-stratified calibration. The present study retained the optical platform but used separate development data and a new clinical cohort to examine measurement availability and systematic displacement across two room-lighting conditions.

The study had two stages. First, paired images from 1,065 independent eyes were collected at 50 and 500 lux, with NIDEK AR-1 measurements at the matched illuminance used as supervised labels. Second, the locked system was evaluated in 50 newly recruited participants using a randomised, within-participant, four-device design. Because i.Profiler plus was not used during model development, it was prespecified as the primary independent optical comparator. NIDEK AR-1 was retained as a secondary comparator to examine agreement with the device that provided the calibration labels. We hypothesised that, compared with Spot VS100, the calibrated system would show greater measurement availability and less illuminance-related M displacement at 500 lux, with a consistent direction of inference across the two comparators.

Methods

Study design and reporting framework

The study comprised (1) retrospective model development using an independent paired-measurement dataset and (2) prospective independent clinical evaluation using a controlled, within-participant repeated-measures design. The development and clinical cohorts were separate and neither overlapped with the participants in the previously published evaluation of the hardware platform. Comparator roles were prespecified: i.Profiler plus served as the primary independent optical comparator, whereas NIDEK AR-1 served as the source of supervised labels and as a secondary clinical comparator. An overview of the two-stage model-development and independent clinical-evaluation workflow is provided in Fig. 1. Reporting was informed by TRIPOD + AI, DECIDE-AI, PROBAST + AI and STARD-AI[15–18].

Ethics and participants

The clinical evaluation protocol was approved by the Ethics Committee of Beijing Tongren Hospital, Capital Medical University (TREC2024-KYS032), and followed the Declaration of Helsinki. All participants gave written informed consent. Adults aged 18–60 years were eligible if they had no history of ocular disease or ocular surgery, best-corrected visual acuity of 20/20 or better in each eye, and absolute cylinder not exceeding 3.00 D. Only the right eye was included in the clinical analysis to preserve statistical independence. The 50-eye clinical cohort was newly recruited and was not used for model development, tuning, internal testing or the earlier clinical publication.

Image-based refraction and illuminance-stratified calibration

The open-field eccentric photorefraction system retained the optical hardware and the base infrared image-prediction framework described in the preceding hardware-platform study[14]. An eccentric near-infrared source produced a fundus reflex across the pupil, and a convolutional neural network was used to extract image information related to pupil boundaries, the spatial distribution of the reflex, luminance gradients and local texture. These features were mapped to an initial estimate of sphere, cylinder and axis.

The base image-prediction model had been developed previously using 362,000 infrared eccentric-photorefraction images acquired under a range of refractive states, pupil sizes, ambient-lighting conditions and degrees of eyelid or eyelash obstruction. In the published implementation, individual infrared frames were resized to 256 × 256 pixels. Random rotation and horizontal and vertical flipping were used during training. Model optimisation used stochastic gradient descent with a momentum of 0.95, an initial learning rate of 0.01, a step-decay factor of 0.9, a batch size of 8 and a maximum of 300 epochs. Mean squared error was used as the regression loss, and the checkpoint with the lowest validation loss was retained. The optical configuration and convolutional image-feature extractor were not redesigned for the present study.

The new component evaluated in the present study was a downstream illuminance-stratified calibration procedure. The calibration-development dataset comprised 1,065 eyes from 1,065 participants, with only one eye included from each participant. Each eye contributed measurements under two prespecified room-lighting conditions, 50 lux and 500 lux. At each illuminance, the open-field system acquired a single-frame infrared eccentric-photorefraction image, and the NIDEK AR-1 provided the corresponding objective refraction under the same room-lighting condition. Images acquired at 50 lux were therefore paired only with 50-lux NIDEK AR-1 measurements, whereas images acquired at 500 lux were paired only with 500-lux measurements. Measurements obtained under one illuminance were not used as labels for images obtained under the other illuminance.

The 1,065 eyes were divided at participant level into a training set of 745 eyes (70%), a tuning set of 213 eyes (20%) and a held-out internal test set of 107 eyes (10%). Both illuminance records from the same eye were retained in the same partition. This prevented information from one participant appearing in both model fitting and model assessment. The independent 50-participant clinical cohort was not included in any of these three partitions.

The initial sphere (S), cylinder (C) and axis (A) estimates produced by the base image-prediction network were converted to the power-vector components M, J_0 and J_{45} before calibration:

$$M = S + \frac{C}{2}$$

$$J_0 = -\frac{C}{2}\cos(2A)$$

$$J_{45} = -\frac{C}{2}\sin(2A)$$

The NIDEK AR-1 results were converted using the same equations and served as the supervised calibration targets. Power-vector representation allowed sphere and astigmatism to be treated as continuous numerical outcomes and avoided direct regression of an axis variable with 180° periodicity.

For eye i measured under illuminance l, where l was either 50 or 500 lux, the initial image-derived vector was denoted by

$$\widehat{\mathbf{y}}_{i,l} = \left(\widehat{M}_{i,l}, \widehat{J}_{0i,l}, \widehat{J}_{45i,l}\right)$$

and the matched NIDEK AR-1 vector by

$$\mathbf{y}_{i,l}^{AR} = \left(M_{i,l}^{AR}, J_{0i,l}^{AR}, J_{45i,l}^{AR}\right).$$

Separate supervised neural-network regression calibrations were fitted for the two illuminance strata. The 50-lux calibration function was estimated only from 50-lux prediction-reference pairs, and the 500-lux function was estimated only from 500-lux pairs. Each calibration function mapped the initial three-component power vector to a calibrated vector:

$$\widetilde{\mathbf{y}}_{i,l} = g_l\left(\widehat{\mathbf{y}}_{i,l}\right),$$

where g_{50} and g_{500} denote the locked calibration functions for 50 and 500 lux, respectively. Calibration was performed on the numerical outputs of the base network rather than by reprocessing the raw image or altering the optical hardware.

The calibration objective was to minimise the mean squared difference between the calibrated vector components and the matched NIDEK AR-1 vector components within each illuminance stratum. The training set was used to estimate the calibration parameters. The tuning set was used to monitor prediction loss, select the final calibration configuration and determine the stopping point. After model selection had been completed, the internal test set was evaluated once and was not used for further parameter adjustment.

During the independent clinical evaluation, the prespecified room-illuminance condition determined whether the locked 50-lux or 500-lux calibration function was applied. The method therefore represents stratified calibration at two discrete illuminance levels rather than continuous adaptation across an unrestricted range of lighting conditions. No interpolation between the two calibration functions was used. The calibrated M, J_0 and J_{45} values were subsequently converted back to sphere, cylinder and axis for device reporting.

Both calibration functions were finalised before the independent clinical evaluation began. None of the images or measurements from the 50-participant clinical cohort was used to retrain the base image-

prediction network, refit the calibration functions, modify calibration coefficients, select thresholds or choose the reported model version. The i.Profiler plus and Spot VS100 measurements were used only for clinical performance evaluation and were not used during model development. Thus, NIDEK AR-1 served as the supervised calibration reference, whereas i.Profiler plus provided an optical comparison independent of the model-development labels.

Clinical instruments

Four instruments were evaluated. The index instrument was the open-field eccentric photorefractor incorporating the locked illuminance-stratified calibration. Spot VS100 (Welch Allyn, Skaneateles Falls, NY, USA) served as an active comparator from the same broad photorefraction class. NIDEK AR-1 (NIDEK Co., Gamagori, Japan) was the source of supervised labels during development and a prespecified secondary comparator in the independent cohort. ZEISS i.Profiler plus (Carl Zeiss Vision GmbH, Aalen, Germany), which combines objective autorefraction with wavefront aberrometry, was not involved in model development and was prespecified as the primary independent optical comparator. Neither comparator was defined as an absolute clinical gold standard.

Illuminance and measurement procedure

All measurements were performed in the same examination room at target ambient illuminances of 50 and 500 lux. Ambient illuminance was measured using a CL-200A Chroma Meter (Konica Minolta Sensing, Inc., Osaka, Japan), operated in illuminance mode. Room lighting was adjusted and verified with a lux meter, with measurements maintained within ± 5% of the target. The order of the four instruments was randomised to reduce order and accommodative effects. Three consecutive readings were obtained with each instrument at each illuminance, and their arithmetic mean was used for analysis. A valid measurement required sphere, cylinder and axis values. Failure to obtain a complete result within the routine measurement sequence was recorded as measurement failure and was not imputed.

Refractive representation

All refractions were expressed in minus-cylinder form and converted to M, J0 and J45 using the equations above. Axis difference was calculated as the smallest circular difference within a 180° period. Because axis becomes unstable at low cylinder powers, axis analyses were exploratory and restricted to eyes meeting the prespecified cylinder threshold.

Outcomes

Two co-primary outcomes were prespecified. The first was paired measurement success for the calibrated system and Spot VS100 at 500 lux. The second was illuminance-related M drift adjusted for the primary comparator: [(device M at 500 lux − i.Profiler plus M at 500 lux) − (device M at 50 lux −

i.Profiler plus M at 50 lux)]. This difference-in-differences quantity separates device-specific change from change occurring concurrently in the optical comparator.

Secondary outcomes included mean bias, mean absolute error (MAE), root mean square error (RMSE), Bland–Altman 95% limits of agreement, the proportion of errors within 0.25, 0.50 and 1.00 D, and corresponding analyses for J0 and J45. Prespecified sensitivity analyses repeated agreement and drift calculations using NIDEK AR-1 as the secondary comparator. Because NIDEK AR-1 provided the supervised development labels, these analyses assessed retention of calibration in new participants rather than independent external validity. Exploratory outcomes included circular axis error and participant characteristics associated with Spot VS100 failure at 500 lux.

Statistical analysis

The participant was the statistical unit. Continuous variables were summarised with means, standard deviations, 95% confidence intervals (CI), medians and interquartile ranges as appropriate. Measurement success was analysed as a paired binary outcome using the exact McNemar test; individual device success rates were accompanied by Wilson 95% CI. Paired continuous differences were tested with paired t-tests, with Wilcoxon signed-rank tests used as distributionally robust sensitivity analyses.

Method agreement was assessed using mean bias, Bland–Altman 95% limits of agreement, MAE, RMSE and clinically interpretable error thresholds rather than correlation alone. The primary analyses used i.Profiler plus as comparator. The same quantities were recalculated relative to NIDEK AR-1, and the two comparator instruments were compared directly to describe their own measurement differences. The clinical database retained the mean of three readings but not the individual readings; repeatability coefficients and intraclass correlation coefficients could therefore not be calculated.

A linear mixed-effects sensitivity model included a participant-specific random intercept and fixed effects for instrument, illuminance and instrument-by-illuminance interaction. The primary model used i.Profiler plus and 50 lux as reference levels; the analysis was repeated with NIDEK AR-1 as comparator. All available continuous observations were retained, so failure of Spot VS100 did not remove valid measurements from the other instruments. To address selective failure, direct comparisons between the two photorefractors at 500 lux were also performed in the 35 participants measured successfully by both instruments, using each comparator in turn.

M was the primary continuous refractive outcome. J0 and J45 were key secondary outcomes; sphere, cylinder and axis were supportive. Co-primary analyses used two-sided $p < 0.05$. Secondary analyses emphasised effect estimates and confidence intervals and were interpreted as supportive evidence.

Results

Clinical cohort and measurement completeness

Fifty participants (50 right eyes) were included; 21 were women. Age ranged from 21 to 39 years (mean 26.5 ± 5.1 years). All four instruments returned valid measurements in all participants at 50 lux. At 500 lux, the calibrated system, NIDEK AR-1 and i.Profiler plus returned results in all 50 participants, whereas Spot VS100 returned a result in 35 (70%; Wilson 95% CI 56.2%–80.9%). All 15 discordant pairs were calibrated-system successes and Spot failures; there were no calibrated-system failures with Spot success. The exact McNemar p value was 0.000061. Measurement success at 500 lux is summarised in Fig. 2a.. See Table 1.

Table 1
Measurement completion by instrument and ambient illuminance

Ambient illuminance	i.Profiler plus	NIDEK AR-1	Calibrated system	Spot VS100
50 lux	50/50 (100%)	50/50 (100%)	50/50 (100%)	50/50 (100%)
500 lux	50/50 (100%)	50/50 (100%)	50/50 (100%)	35/50 (70%)

Within-instrument change with illuminance

i.Profiler plus showed no systematic change in M between 50 and 500 lux (mean change − 0.020 D; 95% CI − 0.129 to 0.088; p = 0.710). NIDEK AR-1 showed a small positive change (+ 0.108 D; 0.023 to 0.192; p = 0.014). The calibrated system showed no systematic mean displacement (− 0.041 D; −0.310 to 0.228; p = 0.760), although individual absolute changes remained dispersed. Among the 35 participants with paired Spot measurements, the mean M change was + 0.593 D (0.295 to 0.890; p < 0.001). The within-instrument changes in M from 50 to 500 lux are summarised in Table 2.

Table 2
Within-instrument change in M from 50 to 500 lux

Instrument	n	Mean change (D)	95% CI (D)	Mean absolute change (D)	p value
i.Profiler plus	50	−0.020	−0.129 to 0.088	0.258	0.710
NIDEK AR-1	50	+ 0.108	0.023 to 0.192	0.228	0.014
Calibrated system	50	−0.041	−0.310 to 0.228	0.743	0.760
Spot VS100	35	+ 0.593	0.295 to 0.890	0.793	< 0.001

Agreement with i.Profiler plus

At 50 lux, mean M bias relative to i.Profiler plus was + 0.035 D for NIDEK AR-1, − 0.087 D for the calibrated system and + 0.482 D for Spot VS100; MAE was 0.335, 0.587 and 0.638 D, respectively. At 500 lux, bias was + 0.163 D for NIDEK AR-1 and − 0.108 D for the calibrated system. Among the 35 successful Spot measurements, bias was + 1.107 D. Corresponding MAE values were 0.330, 0.671 and 1.191 D. The calibrated system had mean bias close to zero at both illuminances, but wider limits of

agreement than NIDEK AR-1. Mean M bias and mean absolute error relative to i.Profiler plus are shown in Fig. 2b and Fig. 2c, respectively. See Tables 3 and 4.

Table 3
Agreement in M with i.Profiler plus at 50 lux

Instrument	Bias (D)	95% limits of agreement (D)	MAE (D)	RMSE (D)	Error ≤ 0.50 D
NIDEK AR-1	+ 0.035	−0.826 to 0.896	0.335	0.436	76.0%
Calibrated system	−0.087	−1.584 to 1.411	0.587	0.761	52.0%
Spot VS100	+ 0.482	−1.224 to 2.189	0.638	0.988	62.0%

Table 4
Agreement in M with i.Profiler plus at 500 lux

Instrument	n	Bias (D)	95% limits of agreement (D)	MAE (D)	RMSE (D)	Error ≤ 0.50 D
NIDEK AR-1	50	+ 0.163	−0.743 to 1.068	0.330	0.485	78.0%
Calibrated system	50	−0.108	−1.897 to 1.682	0.671	0.910	56.0%
Spot VS100	35	+ 1.107	−1.115 to 3.329	1.191	1.573	28.6%

Agreement with NIDEK AR-1

With NIDEK AR-1 as comparator, mean M bias for the calibrated system was − 0.122 D at 50 lux and − 0.270 D at 500 lux; MAE was 0.633 and 0.809 D. Spot VS100 bias was + 0.448 and + 0.957 D, with MAE of 0.593 and 1.143 D. MAE between i.Profiler plus and NIDEK AR-1 was 0.335 D at 50 lux and 0.330 D at 500 lux, indicating small average differences but not identity between the two comparator instruments. See Table 5.

Table 5
Agreement in M with NIDEK AR-1

Illuminance	Instrument	n	Bias (D)	95% limits of agreement (D)	MAE (D)	Error ≤ 0.50 D
50 lux	i.Profiler plus	50	−0.035	−0.896 to 0.826	0.335	76.0%
50 lux	Calibrated system	50	−0.122	−1.775 to 1.532	0.633	52.0%
50 lux	Spot VS100	50	+ 0.448	−1.437 to 2.332	0.593	72.0%
500 lux	i.Profiler plus	50	−0.163	−1.068 to 0.743	0.330	78.0%
500 lux	Calibrated system	50	−0.270	−2.400 to 1.860	0.809	46.0%
500 lux	Spot VS100	35	+ 0.957	−1.599 to 3.514	1.143	45.7%

Table 6
Reference-adjusted M drift from 50 to 500 lux

Comparator	Instrument evaluated	n	Adjusted drift (D)	95% CI (D)	p value
i.Profiler plus	NIDEK AR-1	50	+ 0.128	−0.009 to 0.264	0.066
i.Profiler plus	Calibrated system	50	−0.021	−0.306 to 0.264	0.883
i.Profiler plus	Spot VS100	35	+ 0.574	0.264 to 0.885	< 0.001
NIDEK AR-1	i.Profiler plus	50	−0.128	−0.264 to 0.009	0.066
NIDEK AR-1	Calibrated system	50	−0.149	−0.407 to 0.109	0.253
NIDEK AR-1	Spot VS100	35	+ 0.450	0.144 to 0.756	0.005

Direct comparison of the two photorefractors at 500 lux

Among the 35 participants measured successfully by both photorefractors, M MAE relative to i.Profiler plus was 0.609 D for the calibrated system and 1.191 D for Spot VS100, a paired difference of −0.582 D ($p<0.001$). Relative to NIDEK AR-1, MAE was 0.721 and 1.143 D, respectively, a paired difference of −0.422 D (95% CI −0.735 to −0.108; $p=0.010$). These analyses quantify numerical differences only among participants in whom Spot returned a result and do not incorporate the additional 15 Spot failures.

Bland–Altman analyses

At 500 lux, M bias for the calibrated system versus i.Profiler plus was −0.108 D, with 95% limits of agreement from −1.897 to 1.682 D (n=50). For Spot VS100, bias among 35 successful measurements

was +1.107 D, with limits from −1.115 to 3.329 D. Relative to NIDEK AR-1, the calibrated-system bias was −0.270 D (limits −2.400 to 1.860 D), whereas Spot VS100 bias was +0.957 D (limits −1.599 to 3.514 D). The limits for both photorefractors were wider than those between NIDEK AR-1 and i.Profiler plus; a small group mean bias should therefore not be interpreted as individual interchangeability. The corresponding Bland–Altman plots at 500 lux are presented in Fig. 3a and Fig. 3b.

Astigmatic vectors, measurement failure and mixed-effects sensitivity analyses

Illuminance-related effects for J0 and J45 were less consistent than for M. In the mixed-effects model with i.Profiler plus as comparator, the calibrated-system-by-500-lux interaction was +0.024 D for J0 (95% CI −0.116 to 0.164; p=0.735), and the Spot-by-500-lux interaction was −0.088 D (−0.236 to 0.060; p=0.244). For J45, the corresponding estimates were −0.050 D (−0.127 to 0.027; p=0.204) and +0.019 D (−0.063 to 0.100; p=0.654). The confidence intervals crossed zero and did not provide a consistent astigmatic-vector effect corresponding to that observed for M.

Axis analysis was restricted to eyes with absolute cylinder ≥0.50 D by i.Profiler plus and was exploratory. At 500 lux, the mean circular axis difference from i.Profiler plus was 13.8° for the calibrated system, with 72.4% of eyes within 10°. For Spot VS100, the mean difference was 24.6°, with 27.3% within 10°. These estimates were not treated as confirmatory because of the smaller sample and the dependence of axis stability on cylinder magnitude.

At 500 lux, Spot VS100 succeeded in 35 eyes and failed in 15. According to i.Profiler plus, mean M was −3.311±2.900 D in the success group and −1.534±2.756 D in the failure group, a difference of −1.777 D (Welch t-test p=0.049; Mann–Whitney U p=0.017). Failure was therefore associated with lower myopia and was unlikely to be completely random. Measurement failure was retained as a performance outcome, and valid observations from the other instruments were not deleted when Spot failed.

The mixed-effects model based on all available continuous observations supported the paired analyses. Relative to i.Profiler plus, the calibrated-system-by-500-lux interaction for M was −0.021 D (95% CI −0.359 to 0.317; p=0.904), whereas the Spot-by-500-lux interaction was +0.623 D (0.264 to 0.981; p<0.001). Repeating analyses with NIDEK AR-1 as comparator and restricting direct comparisons to the 35 complete pairs did not change the direction of inference.

Discussion

This two-stage study addressed a practical limitation of open-field eccentric photorefraction: sensitivity to ordinary indoor lighting. A separate dataset of 1,065 paired eyes was used to establish illuminance-stratified calibration, followed by four-device evaluation in a new 50-participant cohort. At 500 lux, Spot VS100 showed two distinct problems: failure to return a measurement in 30% of participants and positive displacement in M among those measured successfully. The calibrated system returned a result

in every participant and showed no detectable reference-adjusted systematic drift. The finding concerns environmental stability, not equivalence to enclosed clinical instruments[19].

Measurement availability is a clinically relevant device property. Failure to return a result can lead to repeat acquisition, movement to a darker room, use of another instrument or referral. Restricting analysis to successful measurements hides this operational burden and may introduce selection bias when failure is related to refractive state, pupil size or image quality. In the present data, deleting the 15 Spot failures would also remove valid observations from the other instruments and would retain a group with greater mean myopia[20,21].

The optical basis for an illuminance effect is plausible. Increased room illuminance constricts the pupil and alters the geometry and contrast of the eccentric infrared reflex[6,19]. A smaller usable pupil area increases the relative contribution of eyelid and eyelash obstruction, while fixed exposure or threshold rules encounter a different image distribution. The change may affect both the extraction of a valid signal and the mapping between image features and refractive error.

The calibration procedure was developed before the clinical evaluation and did not use the 50-participant cohort to derive a post hoc correction. Images and NIDEK AR-1 labels were paired within each illuminance stratum, and the locked model was then applied without updating. The procedure is therefore best described as supervised, illuminance-specific calibration. It has been evaluated at two discrete room-lighting conditions and should not be interpreted as continuous adaptation to arbitrary illuminance[22-24].

The use of two comparators served different purposes. NIDEK AR-1 provided practical supervised labels during model development and allowed concordance with the development reference to be examined in new participants. i.Profiler plus was not used during development and provided an independent optical comparison. Its mean M changed little between 50 and 500 lux, supporting its use for an environmental method comparison. The difference-in-differences analysis then separated device-specific drift from small concurrent changes in the comparator. Similar inference with both comparator devices reduced dependence on a single reference system.

The performance boundary remains important. The calibrated system was more available than Spot VS100 and had lower absolute error at 500 lux, but its individual limits of agreement with i.Profiler plus remained wider than those of NIDEK AR-1. The system should not be presented as a replacement for wavefront aberrometry, subjective refraction or routine tabletop autorefraction. Its current role is to broaden the lighting conditions in which an open-field, lower-cost platform can provide an objective refractive estimate[25].

Effects for J0 and J45 were smaller and less consistent than for M. Direction-specific information may have a lower signal-to-noise ratio, and axis becomes unstable at low cylinder powers. The clinical sample also provided limited precision for interaction effects. Future development should assess vector-

specific loss functions, uncertainty estimation and quality-control thresholds rather than assuming that improvement in M will extend automatically to astigmatic components.

The study used a within-participant design, randomised instrument order, three readings per condition, prospective verification of illuminance, explicit retention of measurement failure and participant-level mixed-effects analysis. These choices increase efficiency in a small cohort and avoid treating 400 instrument–illuminance observations as independent participants. Agreement was assessed with bias, limits of agreement, absolute error, threshold accuracy and interaction estimates rather than correlation alone.

The findings suggest that open-field photorefraction may be used in normally lit rooms without reproducing a dark-room workflow. Potential settings include schools, community clinics, primary-care rooms and optical retail environments. The current evidence is limited to 50 and 500 lux, however, and does not establish robustness under intermediate illuminances, mixed spectra, outdoor light or direct glare.

The study has limitations. The independent cohort was small, single-centre and restricted to young adults, giving limited precision for subgroup and astigmatic analyses. Measurements were non-cycloplegic and objective; agreement with subjective or cycloplegic refraction was not evaluated. Pupil diameter, image-exposure variables, spectral composition and adaptation time were not available for mechanistic modelling. Individual repeated readings were not retained, so repeatability could not be quantified. Finally, complete internal test-set performance and the final model audit should be documented before submission. Further evaluation should include additional illuminance levels, image-quality variables, children, older adults, broader refractive ranges and external centres, together with prospective comparison against subjective refraction.

Conclusion

An illuminance-stratified calibration developed from 1,065 paired eyes preserved measurement availability and reduced systematic illuminance-related displacement of open-field eccentric photorefraction in an independent clinical cohort. At 500 lux, the calibrated system returned a result in all participants and performed more consistently than Spot VS100. Individual agreement remained poorer than with tabletop autorefraction or wavefront aberrometry. The principal contribution is therefore an extension of usable room-lighting conditions, subject to external validation and evaluation over a wider range of illuminance.

Declarations

Funding

This study was funded by the Beijing Municipal Science and Technology Project (Z201100005520042).

Competing interests

The authors declare no competing interests. Any patents, patent applications, ownership interests or commercial relationships related to the index instrument must be disclosed before submission.

Ethics approval

The study was approved by the Ethics Committee of Beijing Tongren Hospital, Capital Medical University (TREC2024-KYS032), and was conducted in accordance with the Declaration of Helsinki and its later amendments.

Consent to participate

Written informed consent was obtained from all participants.

Author contributions

Z.Y. and C.H. conceived the study. Y.H. and H.C. accessed and verified the data. Z.Y. developed the method and performed the statistical analyses. Y.H. and G.J. collected the data. Z.Y. drafted the manuscript. All authors critically revised the manuscript, had access to the study data and approved the decision to submit.

Data and code availability

De-identified clinical evaluation data may be made available by the corresponding author on reasonable request, subject to institutional and ethics requirements. Model weights and proprietary calibration coefficients are not publicly available because of intellectual-property restrictions. The authors should provide sufficient documentation of data preprocessing, participant-level partitioning, model selection and internal test performance for independent methodological assessment. Analysis code for the clinical evaluation may be shared where permitted.

Use of AI-assisted tools

A large language model was used to assist with English translation and editorial organisation. All scientific content, statistical results, interpretations and final wording were reviewed and approved by the authors, who take full responsibility for the manuscript.

References

1. Bourne R, Steinmetz JD, Flaxman S, et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. *The Lancet global health*. 2021;9(2):e130-e143.
2. Steinmetz JD, Bourne RR, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *The Lancet global health*. 2021;9(2):e144-e160.
3. Dhablania N, Torres M, Burkemper B, McKean-Cowdin R, Varma R. Uncorrected Refractive Error in the African American Eye Disease Study. *JAMA ophthalmology*. 2024;142(4):292-298.
4. Muma S, Naidoo KS, Hansraj R. Proposed task shifting integrated with telemedicine to address uncorrected refractive error in Kenya: Delphi study. *BMC Health Services Research*. 2024;24(1):115.
5. McCormick I, Tong K, Abdullah N, et al. Strategies to address inequity of uncorrected refractive error in the Western Pacific: A modified Delphi process. *Ophthalmic and Physiological Optics*. 2024;44(6):1148-1161.
6. Bharadwaj SR, Manzanera S, Artal P. Eccentric Infrared Photorefraction: A Review of Evolution, Optical Design, Features and Applications. *Journal of Bio-optics*. 2025;1(1):2507000945.
7. Howland HC. Photorefraction of eyes: history and future prospects. *Optometry and Vision Science*. 2009;86(6):603-606.
8. Horwood AM, Griffiths HJ, Carlton J, et al. Scope and costs of autorefraction and photoscreening for childhood amblyopia—a systematic narrative review in relation to the EUSCREEN project data. *Eye*. 2021;35(3):739-752.
9. Ferreira A, Vieira R, Maia S, Miranda V, Parreira R, Menéres P. Photoscreening for amblyopia risk factors assessment in young children: A systematic review with meta-analysis. *European Journal of Ophthalmology*. 2023;33(1):92-103.
10. Pophal CJ, Trivedi RH, Bowsher JD, et al. Effectiveness of the spot tm vision screener with variations in ocular pigments. *American Journal of Ophthalmology*. 2024;264:99-103.
11. Cao Y-T, Che D-Y, Pan Y-L, et al. Artificial intelligence improves accuracy, efficiency, and reliability of a handheld infrared eccentric autorefractor for adult refractometry. *International Journal of Ophthalmology*. 2022;15(4):628.
12. Xu D, Ding S, Zheng T, et al. Deep learning for predicting refractive error from multiple photorefraction images. *BioMedical Engineering OnLine*. 2022;21(1):55.
13. Linde G, Chalakkal R, Zhou L, et al. Automatic refractive error estimation using deep learning-based analysis of red reflex images. *Diagnostics*. 2023;13(17):2810.
14. Yan H, Yi Z, Qilin S, Hong C, Yan H, Wei T. Deep learning–empowered low-cost portable automated refraction system: A solution to the inadequate effective correction rate of refractive errors in resource-limited areas. *Ophthalmic and Physiological Optics*. 2025;45(7):1676-1688.
15. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *bmj*. 2022;377

16. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *bmj*. 2024;
17. Moons KG, Damen JA, Kaul T, et al. PROBAST+ AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *bmj*. 2025;388
18. Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nature medicine*. 2025;31(10):3283-3289.
19. Howland HC. Photorefraction of eyes: history and future prospects. *Optom Vis Sci*. Jun 2009;86(6):603-606. doi:10.1097/OPX.0b013e3181a523c9
20. Ferreira A, Vieira R, Maia S, Miranda V, Parreira R, Menéres P. Photoscreening for amblyopia risk factors assessment in young children: A systematic review with meta-analysis. *Eur J Ophthalmol*. Jan 2023;33(1):92-103. doi:10.1177/11206721221099777
21. Pophal CJ, Trivedi RH, Bowsher JD, et al. Effectiveness of the Spot (tm) Vision Screener With Variations in Ocular Pigments. *Am J Ophthalmol*. Aug 2024;264:99-103. doi:10.1016/j.ajo.2024.03.018
22. Cao YT, Che DY, Pan YL, et al. Artificial intelligence improves accuracy, efficiency, and reliability of a handheld infrared eccentric autorefractor for adult refractometry. *Int J Ophthalmol*. 2022;15(4):628-634. doi:10.18240/ijo.2022.04.17
23. Xu D, Ding S, Zheng T, et al. Deep learning for predicting refractive error from multiple photorefraction images. *Biomed Eng Online*. Aug 8 2022;21(1):55. doi:10.1186/s12938-022-01025-3
24. Linde G, Chalakkal R, Zhou L, et al. Automatic Refractive Error Estimation Using Deep Learning-Based Analysis of Red Reflex Images. *Diagnostics (Basel)*. Aug 30 2023;13(17)doi:10.3390/diagnostics13172810
25. Rao DP, Negiloni K, Gurunathan S, et al. Validation of a simple-to-use, affordable, portable, wavefront aberrometry-based auto refractometer in the adult population: A prospective study. *BMC Ophthalmol*. Dec 19 2022;22(1):498. doi:10.1186/s12886-022-02684-5

Tables

Table 1 Measurement completion by instrument and ambient illuminance

Ambient illuminance	i.Profiler plus	NIDEK AR-1	Calibrated system	Spot VS100
50 lux	50/50 (100%)	50/50 (100%)	50/50 (100%)	50/50 (100%)
500 lux	50/50 (100%)	50/50 (100%)	50/50 (100%)	35/50 (70%)

Table 2 Within-instrument change in M from 50 to 500 lux

Instrument	n	Mean change (D)	95% CI (D)	Mean absolute change (D)	p value
i.Profiler plus	50	−0.020	−0.129 to 0.088	0.258	0.710
NIDEK AR-1	50	+0.108	0.023 to 0.192	0.228	0.014
Calibrated system	50	−0.041	−0.310 to 0.228	0.743	0.760
Spot VS100	35	+0.593	0.295 to 0.890	0.793	<0.001

Table 3 Agreement in M with i.Profiler plus at 50 lux

Instrument	Bias (D)	95% limits of agreement (D)	MAE (D)	RMSE (D)	Error ≤0.50 D
NIDEK AR-1	+0.035	−0.826 to 0.896	0.335	0.436	76.0%
Calibrated system	−0.087	−1.584 to 1.411	0.587	0.761	52.0%
Spot VS100	+0.482	−1.224 to 2.189	0.638	0.988	62.0%

Table 4 Agreement in M with i.Profiler plus at 500 lux

Instrument	n	Bias (D)	95% limits of agreement (D)	MAE (D)	RMSE (D)	Error ≤0.50 D
NIDEK AR-1	50	+0.163	−0.743 to 1.068	0.330	0.485	78.0%
Calibrated system	50	−0.108	−1.897 to 1.682	0.671	0.910	56.0%
Spot VS100	35	+1.107	−1.115 to 3.329	1.191	1.573	28.6%

Table 5 Agreement in M with NIDEK AR-1

Illuminance	Instrument	n	Bias (D)	95% limits of agreement (D)	MAE (D)	Error ≤0.50 D
50 lux	i.Profiler plus	50	−0.035	−0.896 to 0.826	0.335	76.0%
50 lux	Calibrated system	50	−0.122	−1.775 to 1.532	0.633	52.0%
50 lux	Spot VS100	50	+0.448	−1.437 to 2.332	0.593	72.0%
500 lux	i.Profiler plus	50	−0.163	−1.068 to 0.743	0.330	78.0%
500 lux	Calibrated system	50	−0.270	−2.400 to 1.860	0.809	46.0%
500 lux	Spot VS100	35	+0.957	−1.599 to 3.514	1.143	45.7%

Analyses using NIDEK AR-1 assess concordance with the device used to derive supervised calibration labels and are not an independent external validation.

Table 6 Reference-adjusted M drift from 50 to 500 lux

Comparator	Instrument evaluated	n	Adjusted drift (D)	95% CI (D)	p value
i.Profiler plus	NIDEK AR-1	50	+0.128	−0.009 to 0.264	0.066
i.Profiler plus	Calibrated system	50	−0.021	−0.306 to 0.264	0.883
i.Profiler plus	Spot VS100	35	+0.574	0.264 to 0.885	<0.001
NIDEK AR-1	i.Profiler plus	50	−0.128	−0.264 to 0.009	0.066
NIDEK AR-1	Calibrated system	50	−0.149	−0.407 to 0.109	0.253
NIDEK AR-1	Spot VS100	35	+0.450	0.144 to 0.756	0.005

Figures

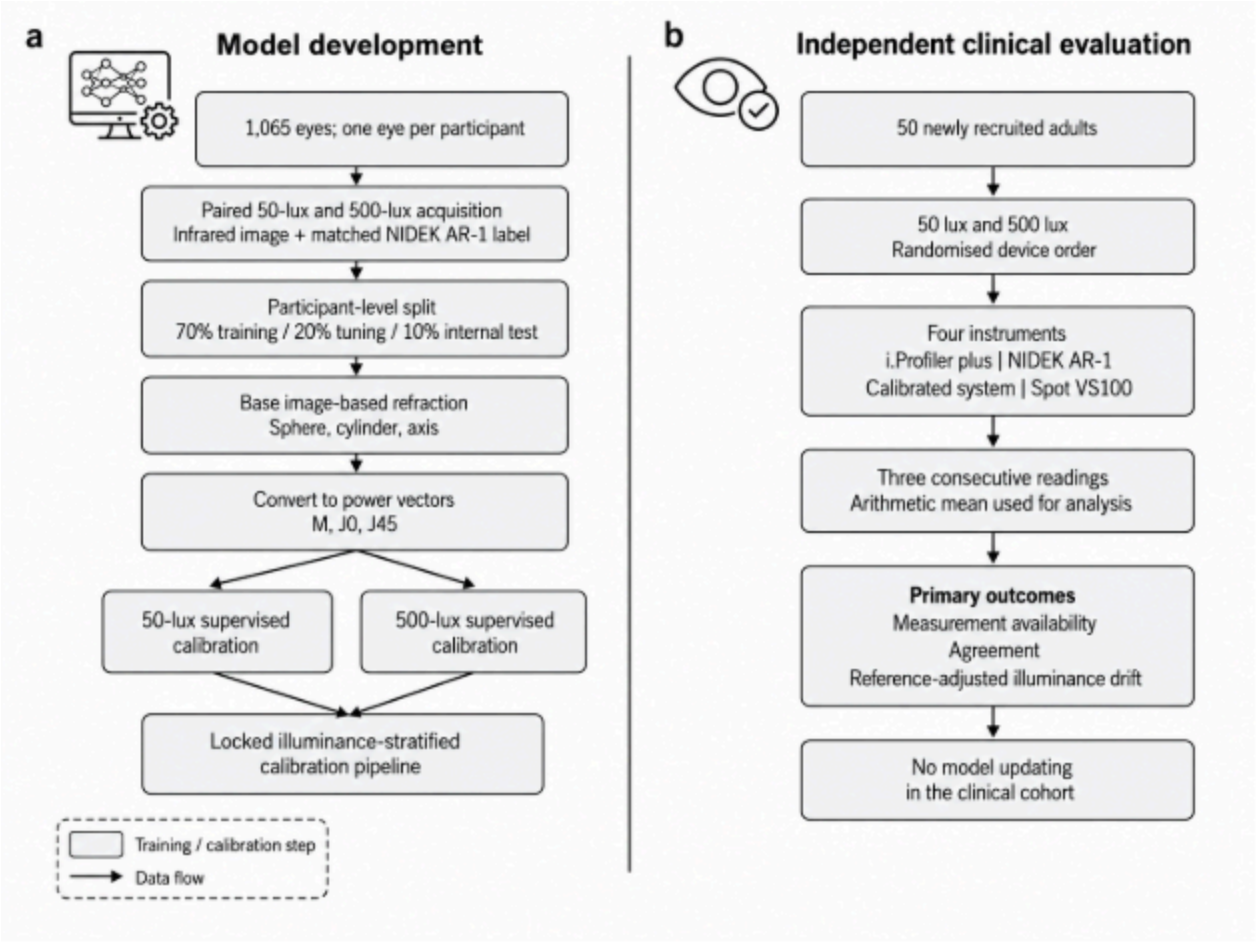


Figure 1

Two-stage design. (a) Model development in 1,065 eyes with participant-level partitioning and separate 50- and 500-lux calibration. (b) Independent four-device evaluation in 50 adults without model updatin

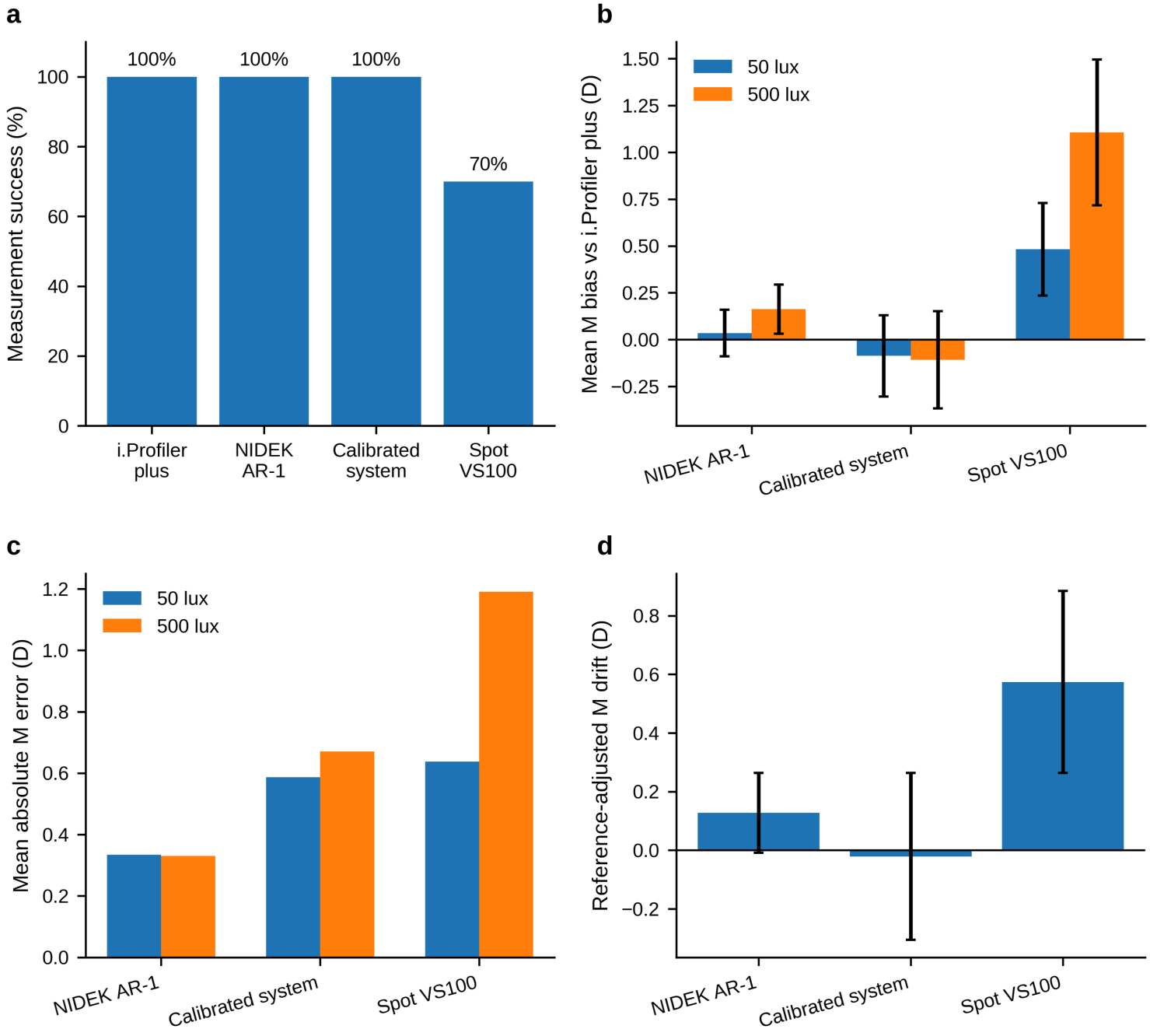


Figure 2

Clinical performance summary. (a) Measurement success at 500 lux. (b) Mean M bias relative to i.Profiler plus at 50 and 500 lux; error bars are 95% confidence intervals. (c) Mean absolute M error relative to i.Profiler plus. (d) Reference-adjusted M drift from 50 to 500 lux using i.Profiler plus as comparator; error bars are 95% confidence intervals. Spot VS100 estimates at 500 lux are based on 35 successful measuremen

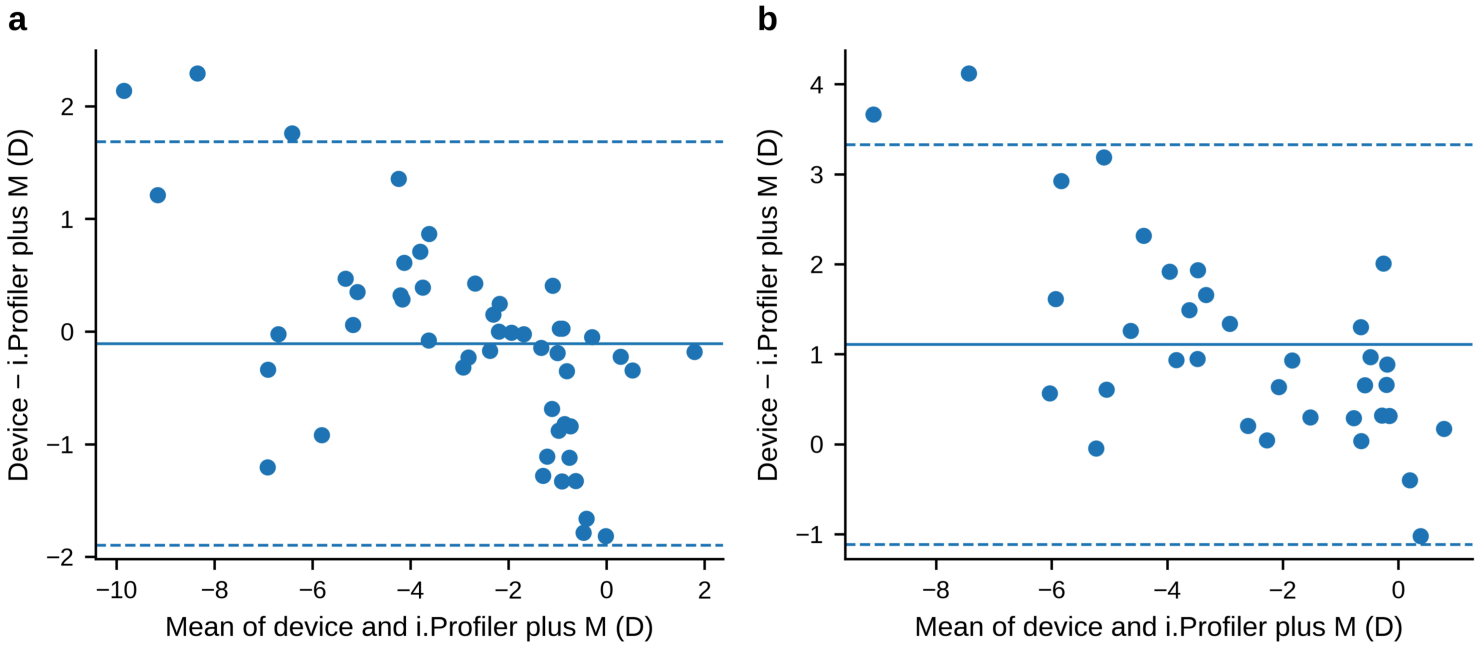


Figure 3

Bland–Altman analyses at 500 lux. (a) Calibrated system versus i.Profiler plus (n=50; bias −0.108 D; 95% limits of agreement −1.897 to 1.682 D). (b) Spot VS100 versus i.Profiler plus among successful measurements (n=35; bias +1.107 D; 95% limits of agreement −1.115 to 3.329 D)