BMJ Open

Skin tone and diagnostic equity in contactless blood pressure screening: a prospective observational field evaluation of remote photoplethysmography in Nigeria

David Dasa , Philip Davies

To cite: Dasa D, Davies P. Skin tone and diagnostic equity in contactless blood pressure screening: a prospective observational field evaluation of remote photoplethysmography in Nigeria. *BMJ Open* 2026;**16**:e119311. doi:10.1136/bmjopen-2026-119311

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2026-119311).

Received 07 March 2026
Accepted 03 June 2026

Bournemouth University, Poole, UK

Correspondence to
David Dasa;
ddasa@bournemouth.ac.uk

ABSTRACT

Objectives To evaluate diagnostic equity, feasibility and acceptability of a remote photoplethysmography-based blood pressure screening application among adults with darker skin tones in Nigeria.

Design Prospective observational multisite field evaluation.

Setting Three hospitals in Kebbi State, Nigeria.

Participants Adults with Fitzpatrick skin types V–VI.

Outcome measures Feasibility, agreement, diagnostic accuracy, acceptability, and equity relevant factors including facial tribal markings and internet bandwidth, using automated cuff measurements as the reference standard and a 140 over 90 mm Hg hypertension threshold.

Results Among 306 enrolled participants, 249 (81.4%) produced usable readings. Agreement was poor (systolic mean absolute error (MAE) 15.4 mm Hg, root mean square error (RMSE) 19.9; diastolic MAE 10.9 mm Hg, RMSE 13.6). Sensitivity for threshold-based systolic and diastolic blood pressure classification was very low (systolic 0.04; diastolic 0.10), with systolic sensitivity 0.00 in Fitzpatrick type VI. Specificity was high (systolic 0.99; diastolic 0.89). Lower internet bandwidth correlated with reading failure (r = −0.69 to −0.51). While 70% of patients and over 90% of staff rated the tool favourably, technical limitations created a clear perception–performance gap. In an exploratory interaction analysis, Fitzpatrick type VI was associated with higher odds of measurement failure (OR 5.08, 95% CI 2.41 to 10.72), but there was no clear evidence that facial tribal markings modified this association (interaction OR 0.66, 95% CI 0.16 to 2.73; p=0.564).

Conclusions Remote photoplethysmography (rPPG)-based blood pressure screening was feasible but showed inadequate performance in this darker-skinned field cohort, with critically low sensitivity. Without algorithmic recalibration for skin tone diversity and improved offline functionality, cloud-dependent rPPG systems deployed without spectrum-balanced validation may risk exacerbating diagnostic inequities in similar settings.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Prospective field evaluation conducted in routine outpatient settings across three hospitals in Kebbi State, Nigeria.
⇒ Adults with Fitzpatrick skin types V and VI were enrolled using predefined eligibility criteria.
⇒ Index test readings were compared with automated cuff measurements using prespecified 140/90 mm Hg thresholds, following STARD guidance for diagnostic accuracy studies.
⇒ Feasibility, agreement, diagnostic accuracy, acceptability, skin type, facial markings and connectivity were assessed using prespecified data collection procedures.
⇒ Limitations include convenience sampling, no Fitzpatrick I to IV comparator group, no objective lux or melanin measurement, binary recording of facial tribal markings and small hypertensive subgroup counts.

INTRODUCTION

Hypertension affects 31% of adults globally and is a leading cause of cardiovascular morbidity and mortality; nearly half of hypertensive adults are unaware of their condition.[1–3] The UK National Institute for Health and Care Excellence (NICE) defines clinic hypertension as ≥140/90 mm Hg, confirmed by ambulatory/home averages ≥135/85 mm Hg.[4] In Nigeria, prevalence is 30.6% with substantial hospital burden; prevalence across Africa remains among the highest globally.[5 6] Conventional cuff measurement depends on trained staff, calibrated devices and time; resources that can be scarce in low-resource settings.[7–10]

Remote photoplethysmography (rPPG) infers blood-volume pulse from subtle skin reflectance changes captured by commodity cameras and has been proposed for contactless vital signs estimation[11 12] (online

supplemental figure S1). However, training/validation data sets frequently under-represent darker skin, risking performance disparities.[13 14] Large evaluations such as VISION-D/V included very few Fitzpatrick V/VI participants,[15] limiting subgroup inference. Connectivity and deployment constraints in rural clinics further complicate use.[16]

We conducted a prospective field evaluation of Lifelight (Xim), an rPPG app that estimates systolic/diastolic blood pressure (SBP/DBP) from a 40 s facial reflection capture,[17 18] in adults with Fitzpatrick types V–VI in Kebbi State, Nigeria (online supplemental figure S2). The application uses the smartphone camera sensor as an optical detector to measure temporal variations in reflected light intensity from facial skin, from which remote photoplethysmographic signals are derived.[19] We assessed feasibility, agreement with a validated cuff reference, diagnostic accuracy at 140/90 mm Hg thresholds and user acceptability, and examined two equity-relevant factors: facial tribal markings and internet bandwidth.[20 21]

rPPG accuracy and limitations in diverse populations

Foundational work demonstrated that plethysmographic signals can be measured remotely from facial reflected light using consumer cameras under ambient illumination.[19] Data set imbalance and optical differences across skin tones can degrade rPPG performance for underrepresented groups.[13 14] Large Lifelight evaluations reported overall accuracy but had very few V/VI participants, restricting equity conclusions.[15] Reviews note small, controlled cohorts and limited real-world testing.[12]

Clinical validation of rPPG tools

VISION-D/V validated Lifelight against standard measurements with limited dark-skin representation.[15] These gaps motivate independent, spectrum-balanced field testing in settings where lighting and connectivity vary.[18 22]

Impact of facial markings on rPPG accuracy

Facial scarifications (‘tribal markings’), culturally significant in parts of Nigeria, could alter local reflectance and potentially affect rPPG signals; this has not been systematically studied.[23 24] We therefore recorded the presence/absence of facial tribal markings to explore any measurement effects (online supplemental figure S3 shows examples of common patterns).

Identified gaps and study objectives

We addressed four gaps: (1) under-representation of darker skin in rPPG data sets, (2) limited independent field validation, (3) unknown impact of facial markings and (4) acceptability in low-resource settings. Accordingly, we evaluated Lifelight’s performance/acceptability in Fitzpatrick V/VI adults in Nigeria, including effects of markings and rural deployment barriers.

MATERIALS AND METHODS

Design, setting and participants

We performed a prospective observational field study across three hospitals in Kebbi State, Nigeria: Sir Yahaya Memorial Hospital, Birnin Kebbi; Martha Bamaiyi General Hospital, Zuru and Bamaiyi Sundu Memorial Medical Centre, Senchi (13 June to 2 July 2024). Two hospitals were urban and one was rural. Adults (≥18 years) with Fitzpatrick V or VI were eligible; exclusions were inability to consent, pregnancy, relevant facial dermatologic conditions, heavy makeup and inability to remain still for ~40 s. Two cohorts were enrolled: patients (primary feasibility/accuracy) and staff (acceptability). Convenience sampling was used. The study was not prospectively registered because it was designed as an observational field evaluation rather than an interventional trial; no public protocol was available before data collection. The protocol and analysis plan have been deposited post hoc on OSF: Dasa and Davies, 2026, https://doi.org/10.17605/OSF.IO/5CZJM. This does not constitute prospective registration.

Devices and procedure

OMRON M2 automated cuffs (Boots branded) procedures followed ANSI/AAMI/ISO 81 060–2 clinical investigation guidance and the European Society of Hypertension International Protocol revision.[21 25] Lifelight (LLTry v2.0.0; Xim) ran on an iPad (eighth gen; iPadOS 17.5.1; front camera 1.2 MP). Participants rested 5 min seated before measurements.[26] Sequence: two OMRON readings (2 min apart), a 40 s Lifelight capture (up to three attempts), then a final OMRON reading to align temporally with the index test. Distance/lighting were standardised where possible; a ring light was used if ambient light was insufficient. Lifelight is CE-marked software with MHRA registration and requires stable connectivity on supported devices.[18]

Ascertainment of skin type and facial tribal markings

Fitzpatrick skin type was assessed in-person by trained healthcare staff using a printed Fitzpatrick visual chart. No objective colourimetry, spectrophotometry, melanin index or lux measurement was available; skin type and lighting were therefore treated as pragmatic field variables rather than instrumented optical measurements. Visual Fitzpatrick assessment may introduce misclassification because it combines observer judgement, ambient lighting and phenotypic appearance rather than direct measurement of melanin or skin reflectance; this could attenuate or distort apparent subgroup differences, particularly between types V and VI. Facial tribal markings were documented as a binary variable (present/absent) following non-invasive visual inspection during enrolment; no further subtype classification (pattern, location, depth or laterality) was recorded. During the measurement, the camera sensor stream was processed in real time to derive rPPG-based numerical

outputs. No identifiable images or video recordings were retained by the study team; derived rPPG time series were deleted after analysis.[27–29]

Reference standard and analysis sets (STARD)

This diagnostic-accuracy study follows STARD guidance.[20] For hypertension classification, the reference standard was the mean of three OMRON readings (SBP≥140 and/or DBP≥90 mm Hg). For paired continuous comparisons, the final OMRON reading was time-aligned to the Lifelight capture. Participants without a valid Lifelight estimate after ≤3 attempts contributed to feasibility outcomes but were excluded from diagnostic-accuracy analyses.

Outcomes

Primary feasibility outcome: proportion producing an app BP estimate. Agreement outcomes: mean absolute error (MAE), root mean square error (RMSE) and Bland–Altman bias and 95% limits of agreement (LoA) for SBP/DBP.[30] Diagnostic outcomes: sensitivity, specificity, PPV and NPV at 140/90 mm Hg. Secondary outcomes: subgroup performance (Fitzpatrick V vs VI; tribal markings present vs absent), bandwidth effects and acceptability from patient/staff questionnaires.

Connectivity monitoring and site characteristics

Bandwidth (MTN Nigeria) was measured two times per day (12:00, 16:00) at each site using fast.com; daily minima/maxima were linked to session-level success/failure and error metrics.

Participant workflow, questionnaires and staff training

Eligibility was confirmed, written informed consent obtained (English/Hausa) and demographics recorded. Mid–upper-arm circumference was used solely for cuff selection; MUAC/BMI were not recorded. Participants sat ~50 cm from the iPad on a tripod facing natural light; the app prompted stillness. Staff received a 1 day training on consent, skin-type assignment, tribal-mark ascertainment, positioning/lighting and troubleshooting. Fitzpatrick skin typing was performed using standardised visual assessment against reference photographic charts. Patient and staff questionnaires (developed from study objectives and digital-health usability literature; clinician-reviewed and translated/back-translated) captured perceived accuracy, comfort, workflow impact, barriers and willingness to adopt.

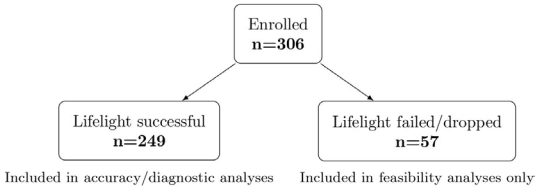

Figure 1 Participant flow for index (Lifelight) and reference (OMRON) testing (STARD).

Data management, privacy and security

Paper CRFs were stored in locked cabinets; data were transcribed to CSV and stored in a password-protected OneDrive accessible only to study staff. Identifiers were replaced with study IDs prior to analysis. In line with protocol and regulation, derived rPPG time series were deleted after analysis.

Statistical analysis

Continuous agreement was summarised with paired differences, MAE/RMSE and Bland–Altman plots with non-parametric bootstrap 95% CIs for bias and LoA. Pearson correlations with the reference were reported where relevant. Diagnostic accuracy used exact (Clopper–Pearson) 95% CIs for sensitivity/specificity/PPV/NPV and Wilson CIs for accuracy. Subgroup estimates were interpreted cautiously given small numerators. Analyses were performed in Python (pandas, NumPy, SciPy, statsmodels); plots were generated using matplotlib and seaborn. We performed an exploratory interaction analysis using logistic regression with Lifelight measurement failure as the outcome and Fitzpatrick type, facial tribal marking status and their product term as predictors. We performed an exploratory connectivity threshold analysis by modelling day-site success rates against Mbps_low, defined as the lower of two daily downlink probes, using attempt-weighted isotonic regression; the threshold was defined as the Mbps_low value associated with 80% per-attempt success.

Sample size and power

Target enrolment (patients n=306; staff n=30) exceeded simple requirements for paired mean comparisons and permitted prespecified subgroup summaries. A paired t-test calculation (two-sided α=0.05, 80% power) indicates n≈34 to detect a medium effect (Cohen's d=0.5); regression guidance of 10–20 observations per predictor suggests n≥120 for modest multivariable models. Precision for subgroup sensitivities was expected to be limited; exact CIs are therefore reported.

STARD flow and handling of indeterminate results

Participants with repeated Lifelight failures (up to three attempts) were classified as failed/dropped for the index test and excluded from diagnostic-accuracy analyses; they contributed to feasibility outcomes. Participant flow is shown in figure 1.

RESULTS

Cohort, demographics and feasibility

We enrolled 306 patients and 30 staff. Of 306 patients, 287 completed questionnaires (response rate 93.8%); not all respondents answered every item. Among those declaring sex, 149 were female and 137 were male. Most participants were 18–30 years (n=133), with 78 aged 31–45 years, 45 aged 46–60 years and 30 aged>60 years. By

Table 1 Participant characteristics and feasibility outcomes

Characteristic	Value
Total enrolled	306
Female/male	149/137 (among 286 respondents reporting sex)
Age 18–30 years	133 (46.5%)
Age 31–45 years	78 (27.3%)
Age 46–60 years	45 (15.7%)
Age>60 years	30 (10.5%)
Fitzpatrick type V	174 (56.9%)
Fitzpatrick type VI	132 (43.1%)
Facial tribal markings present	88 (28.7%)
Lifelight successful	249 (81.4%)
Hypertensive (OMRON; n=249)	71 (28.5%)

clinical assessment, Fitzpatrick type V was more prevalent (174/306; 56.9%) than type VI (132/306; 43.1%). Facial tribal markings were present in 88/306 (28.7%). Participant characteristics are summarised in table 1.

Lifelight produced a BP estimate for 249/306 patients (81.4%); 57 failed/dropped (see figure 1). Urban sites achieved higher success than the rural site, where no valid readings were obtained (online supplemental figure S4).

Among the 71 participants classified as hypertensive by the reference standard (mean of three OMRON readings≥140/90 mm Hg), severity was distributed using NICE clinic blood pressure (BP) staging: 36 (50.7%) had stage 1 hypertension (SBP 140–159 and/or DBP 90–99 mm Hg), 31 (43.7%) had stage 2 hypertension (SBP 160–179 and/or DBP 100–119 mm Hg) and 4 (5.6%) met criteria for severe hypertension (SBP≥180 and/or DBP≥120 mm Hg).

Connectivity and feasibility

Daily bandwidth correlated negatively with failure rates (max speed r=−0.69; min speed r=−0.51; online supplemental figure S5), consistent with the rural site's 100% failure (online supplemental table S1). Correlations between bandwidth and absolute error were weak; connectivity primarily affected the probability of obtaining a reading rather than error magnitude.

Factors associated with measurement success

We examined whether the physiological level of BP itself predicted whether Lifelight would return a reading. Visual inspection of scatter and distribution plots (online supplemental figure S6-S7) showed substantial overlap between retained and dropped attempts, with no clustering of failures at extreme SBP/DBP values. Contingency analyses found no association between elevated BP (by the reference) and success (all p>0.34; negligible Cramer's V; online supplemental table S2).

In contrast, skin type showed a clear association with success (χ^2=22.25, p=2.39×10^{-6}; Cramer's V=0.27), with more failures in type VI (online supplemental table S3). Success rates were 90.8% for type V versus 68.9% for type VI. The presence of facial tribal markings did not affect success (online supplemental table S4; χ^2=0.0012, p=0.9721); success was 80.7% with markings versus 81.7% without.

Agreement (continuous BP)

Having established which factors were linked to success, we evaluated continuous agreement between Lifelight and the reference. Overall error was substantial (SBP MAE 15.37 mm Hg; DBP MAE 10.91 mm Hg), and Bland–Altman analyses showed near-zero mean SBP bias but very wide LoA (LoA~±40 mm Hg), with a small positive bias and wide LoA for DBP. Figures 2 and 3 display subgroup patterns: agreement appeared poorest for type VI, and participants with tribal markings showed greater SBP underestimation than those without.

The diastolic plots similarly indicate wide LoA and a small positive bias overall (figure 3).

Diagnostic accuracy

We next assessed threshold-based classification. Sensitivity for threshold-based SBP and DBP classification was very low, whereas specificity was higher (table 2). For systolic BP, only 2 of 52 hypertensives were correctly identified (sensitivity 0.04), while for diastolic BP, 6 of 60 were detected (sensitivity 0.10). Detailed confusion matrices and full operating characteristics are provided in the supplementary material (online supplemental table S5-S8).

Subgroups

We explored whether performance varied by facial phenotype and skin tone. The critically low sensitivity for Fitzpatrick type VI (systolic 0.00, 95% CI 0.00 to 0.16) reflects complete failure to detect any of the 21 hypertensive cases (by SBP criterion) identified by OMRON in this subgroup. Sensitivity for type V was also poor (SBP 0.06; DBP 0.11). These subgroup estimates should be interpreted as exploratory because hypertensive case counts within each subgroup were small, resulting in wide exact CIs; the type VI finding should be viewed as a safety signal requiring adequately powered validation rather than a precise estimate. Full subgroup analyses are provided in online supplemental table S9-S11.

Exploratory interaction analysis: skin type and facial tribal markings

In an exploratory crossed subgroup analysis, Lifelight failure rates were 9.0% among Fitzpatrick type V participants without facial tribal markings (12/134), 10.0% among type V participants with markings (4/40), 33.3% among type VI participants without markings (28/84), and 27.1% among type VI participants with markings (13/48). In logistic regression including Fitzpatrick type, facial marking status and their interaction, Fitzpatrick type VI was strongly associated with higher odds of failure (unadjusted OR 5.08, 95% CI 2.41 to 10.72; p<0.001), but there was no clear evidence of an interaction between

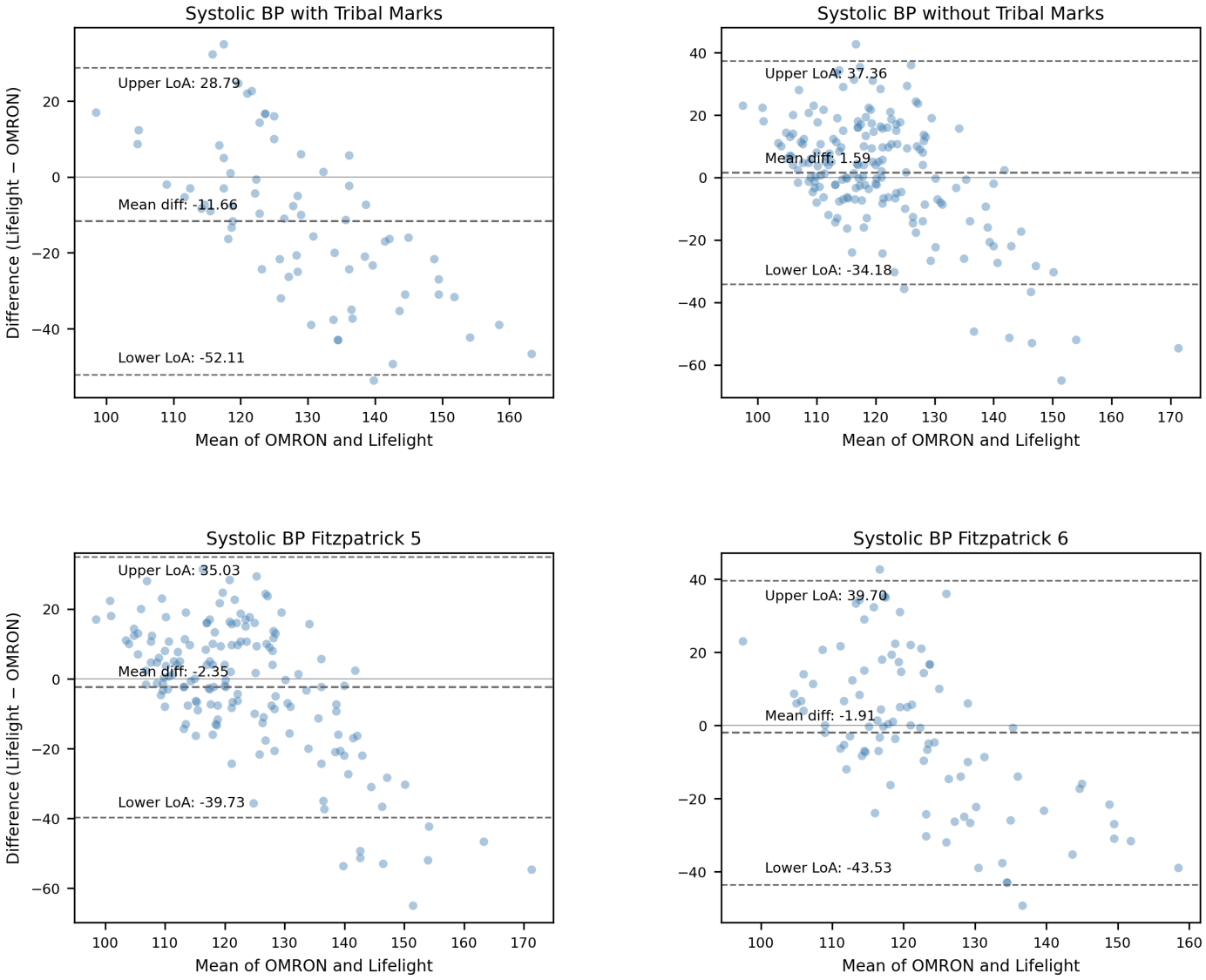

Figure 2 Bland–Altman for SBP by subgroup. Agreement poorest for Fitzpatrick VI. Tribal markings were not associated with overall success/failure. BP, blood pressure; LoA, limits of agreement; SBP, systolic blood pressure.

Fitzpatrick type VI and facial tribal markings (interaction OR 0.66, 95% CI 0.16 to 2.73; p=0.564). This analysis was exploratory and should be interpreted cautiously because crossed subgroup counts were limited. The crossed failure strata are shown in online supplemental table S15.

Bandwidth and measurement error

Because bandwidth governed feasibility, we checked whether it also shifted error magnitudes. Correlations with absolute differences were small and inconsistent (online supplemental table S12), consistent with a threshold effect: adequate speed prevents failure, but extra bandwidth adds little accuracy.

Exploratory bandwidth threshold analysis

To estimate a pragmatic deployment threshold, we modelled day-site success rates against Mbps_low (the lower of two daily downlink probes) using weighted isotonic regression. The estimated Mbps_low associated with 80% per-attempt success was 5.94 Mbps (95% CI 4.0 to 15.0 Mbps). Descriptive operational tiers showed 0% success below 3 Mbps (two attempts), 71.8% success at 3–5 Mbps (117 attempts) and 89.0% success above 5 Mbps (118 attempts). Approximately 6 Mbps downlink may serve as a pragmatic deployment threshold for this cloud-dependent workflow in this setting. This threshold should not be treated as a universal technical requirement: uplink throughput, latency, jitter, packet loss and application-level failure states were not logged and could shift the true operating threshold.

User perceptions

Survey responses provide context for deployment constraints and opportunities. Patients most often flagged connectivity and accuracy as concerns (online supplemental figure S8), despite favourable comfort and experience ratings. The staff cadre was predominantly nursing (73.3%), reflecting the target user group for routine vitals collection (online

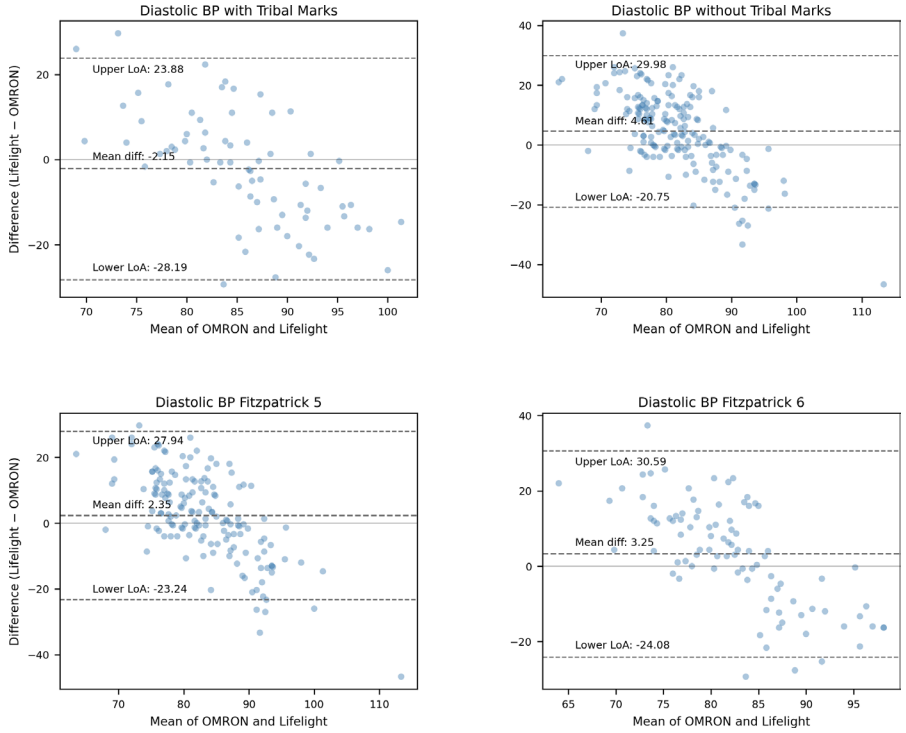

Figure 3 Bland–Altman for DBP by subgroup. Positive bias and wide LoA overall. BP, blood pressure; DBP, diastolic blood pressure; LoA, limits of agreement.

supplemental table S13). Perceptions varied by seniority, with more positive views among junior staff (online supplemental figure S9). Overall, 75.89% of patients rated comfort ≥4/5, 70.03% rated perceived accuracy ≥4/5 and 93.33% of staff expressed willingness to integrate the tool into workflows, indicating a substantial perception-performance gap.

Summary of key findings

Across sites, Lifelight operated for most urban participants but failed at the rural site due to bandwidth. Agreement with cuff measurements was weak (SBP MAE 15.37; DBP MAE 10.91; wide LoA), and diagnostic sensitivity was very low, particularly for Fitzpatrick type VI. Specificity was high for SBP but lower for DBP. Connectivity predominantly affected feasibility rather than error magnitude. Key quantitative indicators are summarised in online supplemental table S14.

DISCUSSION

Principal findings

In this prospective, multisite field evaluation among adults with Fitzpatrick V–VI skin types in Kebbi State, Nigeria, Lifelight produced valid readings for 81.4% of participants but showed poor agreement with the reference cuff (SBP MAE 15.37 mm Hg; DBP MAE 10.91 mm Hg; LoA approximately ±40 mm Hg) and very low sensitivity for hypertension detection (SBP 0.04; DBP 0.10). High SBP specificity (0.99) does not offset the clinical risk posed by missed hypertensives, which was most pronounced in Fitzpatrick type VI (SBP sensitivity 0.00). Connectivity primarily affected feasibility, with strong negative correlations between bandwidth and failure rates and complete failure at the rural site, while having negligible association with absolute error magnitude. Facial tribal markings did not reduce the probability of obtaining a reading; in an exploratory interaction analysis, type VI remained strongly associated with failure but no clear interaction with tribal marking status was identified.

Strengths and limitations

Strengths of this study include prospective field testing conducted in routine outpatient settings across three hospitals encompassing urban and rural environments, explicit STARD-compliant diagnostic accuracy reporting, focused enrolment of adults with Fitzpatrick V and VI

Table 2 Overall diagnostic accuracy (successful reads only; n=249)

	Sensitivity	Specificity	PPV	NPV
Systolic BP (≥140 mm Hg)	0.04 (0.00–0.13)	0.99 (0.96 to 1.00)	0.50 (0.07–0.93)	0.80 (0.74–0.84)
Diastolic BP (≥90 mm Hg)	0.10 (0.04–0.21)	0.89 (0.84 to 0.93)	0.23 (0.09–0.44)	0.76 (0.70–0.81)

BP, blood pressure; NPV, negative predictive value; PPV, positive predictive value.

skin types previously under-represented in rPPG validation, and integration of quantitative and qualitative data including patient and staff acceptability assessment.

Fitzpatrick type was assessed visually by trained staff using printed reference charts. No objective colorimetry, spectrophotometry, melanin index or lux measurement was available. Visual Fitzpatrick assessment may introduce misclassification because it combines observer judgement, ambient lighting and phenotypic appearance rather than direct measurement of melanin or skin reflectance.[31–33] Misclassification could attenuate or distort apparent subgroup differences, particularly between types V and VI. The absence of objective lux measurement also limits interpretation of whether failures or errors were attributable to skin optics, illumination, site conditions or their interaction.

Because the study enrolled only adults with Fitzpatrick V and VI skin types and did not include a Fitzpatrick I–IV comparator cohort, we cannot determine whether the observed performance reflects a disparity specific to darker skin, a general limitation of the tested app build or a combined effect of phenotype, lighting, connectivity and site workflow. The findings should be interpreted as evidence of inadequate performance in this darker-skinned field cohort, not as definitive proof that poorer performance was caused by skin tone alone.

Subgroup sensitivity estimates (particularly Fitzpatrick VI SBP sensitivity) are underpowered for definitive inference. The finding of zero detected SBP hypertensive cases in type VI is clinically concerning but the CI is wide because hypertensive case counts were limited. This result should be viewed as a safety signal requiring adequately powered validation rather than a precise estimate.

Facial tribal markings were recorded only as present or absent. This binary categorisation may obscure variation in pattern, location, depth, laterality and surface texture, all of which could plausibly affect local reflectance and rPPG signal quality. Future studies should include culturally sensitive, consented subtyping of marking patterns where appropriate.

Comparison with previous studies

These findings contrast with earlier evaluations of Lifelight (VISION-D/V), in which types V/VI were minimally represented, limiting equity conclusions. Our results align with broader concerns that under-representation of darker skin tones in rPPG training and validation data sets can degrade performance and with evidence that common benchmark data sets contain very few dark-skinned participants: MMSE-HR, AFRL and UBFC-rPPG contain approximately 10%, 0% and 5% dark-skinned subjects respectively. Emergency-department evaluations have shown promising correlations for some rPPG implementations but were not conducted in predominantly darker-skinned African populations under real-world connectivity constraints. The present evaluation is, to the best of our knowledge, the first independent field validation of a commercial rPPG BP app in a predominantly Fitzpatrick V–VI African outpatient cohort.

Clinical and policy implications

Given the very low sensitivity observed, the app must not be used as a rule-out test for hypertension in this or similar populations: a normal app reading must not replace or defer standard cuff-based BP measurement. In a Nigerian outpatient workflow, a possible restricted use would be adjunctive rule-in triage only: a high app reading would trigger immediate cuff confirmation and management according to the local hypertension pathway but a normal app reading would not substitute for cuff measurement. Failed app measurements should default immediately to cuff-based measurement. This workflow preserves the clinical role of the cuff as the definitive screening tool while preventing false reassurance from rPPG-negative results.

The observed connectivity requirements argue strongly for reliable on-device inference with graceful degradation as a prerequisite for rural African deployment. Procurement and evaluation guidance should require spectrum-balanced validation data and documented performance across Fitzpatrick types V and VI before approving rPPG tools for use in darker-skinned populations.

Unanswered questions and future research

Whether the observed performance gap is attributable to skin tone specifically or to the combined effect of phenotype, ambient lighting, image quality, connectivity and app build cannot be determined without a spectrum-balanced validation including Fitzpatrick I–VI participants, objective colorimetric skin-tone measurement, controlled and logged illumination and adequate power for subgroup sensitivity and specificity estimation. The exploratory bandwidth threshold analysis identified approximately 6 Mbps downlink as a pragmatic deployment marker but uplink, latency, jitter, packet loss and application-level failure states were not logged so this threshold should not be treated as a universal technical requirement.

Future priorities include retraining the algorithm with balanced V/VI representation and diverse facial phenotypes including scarification; assessing multi-band and near-infrared strategies for pulse-signal extraction in darker skin; delivering robust on-device inference with quality gating and asynchronous synchronisation for low-bandwidth settings; prospectively subtyping facial marking patterns as a study variable and using reference standards beyond clinic cuffs such as ambulatory or home BP monitoring where feasible. Prospective postmarket fairness monitoring is also recommended if the device is deployed.

CONCLUSION

In this independent field evaluation, Lifelight did not meet the sensitivity or agreement thresholds required for hypertension screening in adults with Fitzpatrick V–VI skin types in Kebbi State, Nigeria. Feasibility was bandwidth-dependent with complete failure at the rural site and diagnostic sensitivity was critically low particularly in Fitzpatrick type VI (SBP sensitivity 0.00). Individual-level agreement with the reference cuff was poor with wide LoA implying clinically meaningful error even when mean bias is small. Facial tribal markings did not reduce the chance of obtaining a reading and no clear interaction between skin type and marking status was identified though binary marking categorisation limits inference. These findings pertain to this specific Lifelight research build, population and deployment context; other rPPG products require independent, spectrum-balanced validation before generalisation. Algorithmic re-engineering for darker skin, equity-centred validation and robust offline operation are prerequisites before any screening use in similar settings.

Acknowledgements The authors thank the management and outpatient department staff of Sir Yahaya Memorial Hospital Birnin Kebbi, Martha Bamaiyi General Hospital Zuru and Bamaiyi Sundu Memorial Medical Centre Senchi for their assistance in organising participant recruitment and supporting data capture.

Contributors DD and PD conceived and designed the study. DD coordinated study conduct, data collection and project administration across all three study sites. DD performed the statistical analysis and drafted the initial manuscript. DD and PD contributed to interpretation of the data, critically revised the manuscript for important intellectual content, approved the final version to be submitted and agreed to be accountable for all aspects of the work. DD is the guarantor and accepts full responsibility for the overall content, had access to the data and controlled the decision to publish.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Ethics approval This study involves human participants and was approved by Kebbi State Ministry of Health Research Ethics Committee (107:017/2024). Bournemouth University Research Ethics Committee (55993). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data sets generated and analysed during the current study contain sensitive participant information and are not publicly available in accordance with participant consent agreements. Deidentified summary data supporting the findings of this study are available from the corresponding author upon reasonable request. Derived remote photoplethysmography time series were deleted after analysis and are therefore not available.

ORCID iD

David Dasa https://orcid.org/0009-0007-1535-5962

REFERENCES

1 Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. *Nat Rev Nephrol* 2020;16:223–37.
2 WHO. First who report details devastating impact of hypertension and ways to stop it. 2023. Available: https://www.who.int/news/item/19-09-2023-first-who-report-details-devastating-impact-of-hypertension-and-ways-to-stop-it
3 Zhou B, Perel P, Mensah GA, *et al*. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nat Rev Cardiol* 2021;18:785–802.
4 National Institute for Health and Care and C. National guideline, hypertension in adults: diagnosis and management. London National Institute for Health and Care Excellence; 2019.
5 Adeloye D, Owolabi EO, Ojji DB, *et al*. Prevalence, awareness, treatment, and control of hypertension in Nigeria in 1995 and 2020: A systematic analysis of current evidence. *J of Clinical Hypertension* 2021;23:963–77.
6 Kolo PM, Jibrin YB, Sanya EO, *et al*. Hypertension-related admissions and outcome in a tertiary hospital in northeast Nigeria. *Int J Hypertens* 2012;2012:960546.
7 Schmidt B-M, Durao S, Toews I, *et al*. Screening strategies for hypertension. *Cochrane Database Syst Rev* 2020;5.
8 Guirguis-Blake JM, Evans CV, Webber EM, *et al*. Screening for Hypertension in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2021;325:1657–69.
9 Boniol M, Kunjumen T, Nair TS, *et al*. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and "universal" health coverage? *BMJ Glob Health* 2022;7:e009316.
10 John O, Campbell NRC, Brady TM, *et al*. The 2020 "WHO Technical Specifications for Automated Non-Invasive Blood Pressure Measuring Devices With Cuff". *Hypertension* 2021;77:806–12.
11 Wang W, den Brinker AC, Stuijk S, *et al*. Algorithmic Principles of Remote PPG. *IEEE Trans Biomed Eng* 2017;64:1479–91.
12 Rouast PV, Adam MTP, Chiong R, *et al*. Remote heart rate measurement using low-cost RGB face video: a technical literature review. *Front Comput Sci* 2018;12:858–72.
13 Ba Y, Wang Z, Karinca KD, *et al*. Overcoming difficulty in obtaining dark-skinned subjects for remote-ppg by synthetic augmentation. *arXiv* [Preprint] 2021.
14 Nowara EM, McDuff D, Veeraraghavan A. A meta-analysis of the impact of skin type and gender on non-contact photoplethysmography measurements. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; Seattle, WA, USA, 2020 Available: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9142289
15 Heiden E, Jones T, Brogaard Maczka A, *et al*. Measurement of Vital Signs Using Lifelight Remote Photoplethysmography: Results of the VISION-D and VISION-V Observational Studies. *JMIR Form Res* 2022;6:e36340.
16 Ezeudu TS, Fadeyi TJ. Examining the Influence of Infrastructure Deficit on Economic Activities, Education, and Healthcare in Rural Areas of Nigeria. *Nnamdi Azikiwe Journal of Political Science* 2024;9:155–76.
17 Xim. How lifelight works, science. 2021. Available: https://lifelight.ai/science
18 NICE. Lifelight first for monitoring vital signs. London, 2020.
19 Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. *Opt Express* 2008;16:21434–45.
20 Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Stard 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology* 2015;277:826–32.
21 Association for the Advancement of Medical Instrumentation. ANSI/aami/iso 81060-2:2019 non-invasive sphygmomanometers part 2:

clinical investigation of intermittent automated measurement type. Arlington VA: AAMI; 2019.
22 Niel F, Sahasrabudhe M, Vives V, *et al*. Prospective clinical validation of a noncontact vital signs measurement smartphone application in emergency department. *Eur Heart J* 2023;44:ehad655.
23 Adisa TA, Mordi C, Timming AR. Employment Discrimination against Indigenous People with Tribal Marks in Nigeria: The Painful Face of Stigma. *Work, Employment and Society* 2024;38:787–808.
24 Wang W, Shan C. Impact of makeup on remote-PPG monitoring. *Biomed Phys Eng Express* 2020;6:035004.
25 O'Brien E, Atkins N, Stergiou G, *et al*. European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. *Blood Press Monit* 2010;15:23–38.
26 Sala C, Santin E, Rescaldani M, *et al*. How long shall the patient rest before clinic blood pressure measurement? *Am J Hypertens* 2006;19:713–7.
27 Sachdeva S. Fitzpatrick skin typing: applications in dermatology. *Indian J Dermatol Venereol Leprol* 2009;75:93–6.
28 Coleman W, Mariwalla K, Grimes P. Updating the Fitzpatrick Classification: The Skin Color and Ethnicity Scale. *Dermatol Surg* 2023;49:725–31.
29 Federal Government of NIgeria. Nigeria data protection act. Nigeria Federal Republic of Nigeria Official Gazette; 2023.
30 Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;327:307–10.
31 Eilers S, Bach DQ, Gaber R, *et al*. Accuracy of self-report in assessing Fitzpatrick skin phototypes I through VI. *JAMA Dermatol* 2013;149:1289–94.
32 Weir VR, Dempsey K, Gichoya JW, *et al*. A survey of skin tone assessment in prospective research. *NPJ Digit Med* 2024;7:191.
33 Ly BCK, Dyer EB, Feig JL, *et al*. Research Techniques Made Simple: Cutaneous Colorimetry: A Reliable Technique for Objective Skin Color Measurement. *J Invest Dermatol* 2020;140:3–12.