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.

Semi-Automated Erythema Quantification for Monitoring Propranolol Response in Infantile Haemangioma: A Pilot Feasibility Study

Bitesh Kumar
All India Institute of Medical Sciences
Vishesh Jain
All India Institute of Medical Sciences
Prabudh Goel
All India Institute of Medical Sciences
Devendra Kumar Yadav
All India Institute of Medical Sciences
Anjan Kumar Dhua
dhuaanjan@gmail.com
All India Institute of Medical Sciences

Research Article

Keywords: Infantile haemangioma, Propranolol, Erythema index, Image analysis, Photographic scoring, Intra-rater reliability, Feasibility, Vascular tumour

Posted Date: June 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9629074/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Pediatric Surgery International on September 22nd, 2026. See the published version at https://doi.org/10.1007/s00383-026-06590-5.

Abstract

Background

Infantile haemangioma (IH) is the most common vascular tumour of infancy. Propranolol is first-line treatment, but response monitoring relies on subjective clinical assessment. Semi-automated photographic erythema quantification may offer a reproducible, low-cost adjunct monitoring tool.

Objectives

To develop a semi-automated Python-based erythema quantification tool for serial clinical photographs of IH, evaluate its feasibility and intra-rater reliability, and characterise the relationship between colour-based erythema change and clinician-assessed composite response.

Methods

Retrospective analysis of a prospective photographic registry at a tertiary paediatric surgery centre. Forty-nine infants with IH treated with propranolol were included. Serial clinical photographs (n = 251) were analysed using a custom Python tool. The erythema index (EI = [R − G] / [R + G + B]), a formula validated for quantification of cutaneous erythema from digital photographs, was calculated from manually selected lesion regions of interest (ROI). Intra-rater reliability was assessed by re-rating 22 photographs from five patients one week apart. Absolute EI change was correlated with prospectively recorded, blinded clinician response scores (0–10).

Results

Of 251 photographs, 242 (96.4%) were analysable (mean 4.9 visits per patient, range 2–11). Intra-rater reliability was excellent: ICC(A,1) = 0.990 (95% CI 0.97–1.00); Bland-Altman mean difference − 0.005 EI units (95% limits of agreement − 0.023 to + 0.013). Mean baseline EI was 0.122 ± 0.058 and mean final EI was 0.115 ± 0.053. Absolute EI change did not correlate with clinician response scores (Pearson r = − 0.084, 95% CI − 0.357 to 0.202, p = 0.566; Spearman r = − 0.087, p = 0.554), consistent across three sensitivity analyses.

Conclusions

Semi-automated erythema quantification from clinical photographs of IH is feasible with excellent intra-rater reliability. Colour-based EI change alone was insufficient to track the clinician's composite response assessment under real-world uncalibrated photography conditions, consistent with the partial-biomarker

nature of colour-only analysis. Prospective studies with standardised photography incorporating colorimetric calibration and composite colour-plus-area scoring are warranted.

Introduction

Infantile haemangioma (IH) is the most common benign vascular tumour of infancy, with a prevalence of 4–5% in the general paediatric population [1]. Although most lesions involute spontaneously, a significant proportion require intervention due to functional impairment, ulceration, or cosmetic concerns [2]. Since its introduction in 2008, oral propranolol has become the standard of care for problematic IH, with response rates exceeding 85% in published series [3].

Clinical monitoring of propranolol response remains largely subjective. Most centres rely on sequential clinical photographs assessed by the treating clinician, with response graded using ordinal scales or physician global assessments. This approach is susceptible to inter-observer variability, recall bias, and the inherent difficulty of integrating multiple lesion characteristics — colour, size, and texture — into a single composite judgement [4]. Validated composite scoring tools including the Haemangioma Activity Score (HAS) and Physician Global Assessment (PGA) have been developed, but their adoption in routine practice remains inconsistent [5, 6].

Objective photographic erythema quantification offers a potential low-cost adjunct to clinical assessment. The erythema index (EI), derived from the red, green, and blue (RGB) channel intensities of digital photographs, has been validated for skin redness quantification in wound healing, psoriasis, and port wine stain (PWS) assessment [7, 8]. Abagge KT et al., previously demonstrated the feasibility of colorimetric photograph analysis in IH using calibrated reference patches with Adobe Photoshop, reporting a mean 45% colour clearing index over 16 weeks in a timolol-treated cohort [9]. Objective photographic methods have also been benchmarked against alternative imaging modalities including dynamic infrared thermography and colour Doppler ultrasound, though none has achieved widespread clinical adoption [10].

A critical conceptual point underpins the design of this study: EI captures only the colour dimension of IH response. It does not measure lesion volume, surface texture, or palpation characteristics. Testing EI against a composite clinical endpoint that encompasses colour, size, and texture is therefore not a like-for-like validation; it is a partial-biomarker assessment. The aim of this pilot study was to develop and evaluate a semi-automated open-source Python erythema quantification tool, establish its intra-rater reliability, characterise its feasibility in a real-world clinical cohort, and explicitly define the boundaries of colour-only erythema analysis as a monitoring biomarker.

Methods

Study design and setting

This was a retrospective analysis of a prospective photographic registry conducted at a tertiary paediatric surgery referral centre. Ethical approval for the registry was obtained from the Institutional Ethics Committee (Ref. No.: AIIMSA1384/07.06.2024, RP-40/2024). All patient data were anonymised prior to image analysis.

Patients

Inclusion criteria: clinical diagnosis of IH; propranolol treatment; at least one baseline and one follow-up clinical photograph. Exclusion criteria: (i) multiple simultaneous IH lesions where the photographed and assessed lesion could not be confirmed as identical; (ii) baseline photograph taken after propranolol initiation. Natural IH involution overlapping with propranolol response represents an acknowledged confounder inherent to retrospective design.

Sample size

This was a pilot feasibility study; no formal sample size calculation was performed for the primary correlation analysis. Post-hoc power analysis indicates n = 49 provides approximately 20% power to detect a moderate correlation of r = 0.30 at α = 0.05 (two-tailed), confirming the study is underpowered for definitive correlation testing. A definitive study would require approximately n = 84 patients for 80% power to detect r = 0.30.

Photographic data acquisition

Clinical photographs were obtained using a Nikon D750 digital SLR camera (Nikon Corporation, Tokyo, Japan) under ambient clinical lighting without colorimetric calibration targets or fixed camera-to-lesion distance. Photographs were stored as JPEG or NEF (Nikon raw) format. NEF files were converted to JPEG using rawpy (v0.18) with camera white balance preservation. Photographs with inadequate illumination (n = 7) or lesion obscured (n = 2) were excluded. The absence of photography standardisation is an acknowledged methodological limitation; future studies should incorporate calibration reference patches as described by Abagge KT et al., and Sommers MS et al., [9, 11].

Image analysis pipeline

A custom semi-automated scoring tool was developed in Python 3.13 using OpenCV (cv2) Dhua AK, NumPy, and Tkinter (Fig. 1) [12, 13]. The analyst manually selected a rectangular ROI encompassing the haemangioma and approximately 2–5 mm of surrounding normal skin, following pre-specified rules excluding scale markers, clothing, and background. The tool displayed the ROI with a dashed boundary overlay and computed erythema metrics in real time. The EI was calculated as:

$$EI = (R - G) / (R + G + B)$$

where R, G, and B represent mean pixel intensities in the red, green, and blue channels respectively (range 0–255), a formula validated for quantification of cutaneous erythema from digital photographs Takiwaki H [14]. Higher EI values indicate greater relative redness. The formula captures excess redness over green — the defining chromatic characteristic of haemangioma tissue — normalised for overall brightness, thereby amplifying the signal difference between lesion and surrounding normal skin. Additional metrics included red ratio (R / [R + G+B]), mean RGB values, Individual Typology Angle (ITA), and ROI pixel count. All scores were saved automatically to a comma-separated values file. Source code is freely available at https://github.com/dhuaanjancode/hemangioma-erythema-scorer (Python 3.8+; OpenCV, NumPy, Pillow) Dhua AK [12].

Intra-rater reliability

Intra-rater reliability was assessed prospectively by re-rating 22 photographs from five patients (representing the full observed range of baseline EI, lesion site, and treatment response) one week after initial scoring, without reference to previous results. The intraclass correlation coefficient [ICC(A,1), two-way mixed effects, absolute agreement, single measures] and 95% CI were calculated using pingouin (v0.5.3). Bland-Altman analysis characterised systematic bias and limits of agreement (LoA) [11].

Clinician response assessment

At each follow-up visit, the treating clinician prospectively recorded a response score (0–10), defined as the estimated percentage reduction in overall lesion burden (erythema, lesion size, and surface texture) divided by 10. Scores were recorded in routine clinical practice before any image analysis was performed; the scoring clinician was therefore fully blinded to EI values throughout. A single clinician applied this scale consistently to all 49 patients.

Statistical analysis

The primary endpoint was absolute EI change (baseline minus final follow-up EI). Pearson and Spearman correlation coefficients were computed against the most recent clinician response score, with 95% CI for Pearson r by Fisher z-transformation. Three pre-specified sensitivity analyses were performed: (i) visit-matched pairwise analysis; (ii) restriction to patients with baseline EI ≥ 0.08; (iii) Spearman rank correlation. All analyses used Python 3.13 with SciPy (v1.11) and pingouin (v0.5.3). Significance threshold: $p < 0.05$ (two-tailed).

Results

Patient cohort and photograph usability

Of 293 patients with IH in the registry, 49 met eligibility criteria (Fig. 2). Four patients were excluded for multiple concurrent lesions (n = 2) or uncertain baseline photograph timing (n = 2); 8 had single-visit photographs only. Of 251 photographs reviewed, 242 (96.4%) were analysable. Mean scored visits per patient was 4.9 (range 2–11; Table 1).

Intra-rater reliability

Re-rating of 22 photographs from five patients demonstrated excellent intra-rater reliability: ICC(A,1) = 0.990 (95% CI 0.97–1.00, $p < 0.001$). Bland-Altman analysis showed a mean difference of − 0.005 EI units (95% LoA − 0.023 to + 0.013), indicating negligible systematic bias. The maximum observed difference between sessions was 0.029 EI units, representing 10.9% of that patient's EI range. These results compare favourably with published ICC values for manual ROI-based digital image colour analysis of ≥ 0.91–0.99 [11].

Erythema index at baseline and follow-up

Mean baseline EI was 0.122 ± 0.058 (range 0.017–0.275) and mean final follow-up EI was 0.115 ± 0.053 (range 0.035–0.273). Mean absolute EI change was + 0.008 ± 0.063 (− 0.110 to + 0.171), with substantial inter-patient variability. Full study characteristics are in (Table 1).

Table 1
Study characteristics, image usability, intra-rater reliability, and correlation results (n = 49 patients)

Characteristic	Value
Number of patients	49
Total photographs reviewed	251
Photographs analysed	242 (96.4%)
Photographs excluded — poor illumination	7 (2.8%)
Photographs excluded — lesion obscured	2 (0.8%)
Visits per patient, mean (range)	4.9 (2–11)
Baseline erythema index (EI), mean ± SD (range)	0.122 ± 0.058 (0.017–0.275)
Final follow-up EI, mean ± SD (range)	0.115 ± 0.053 (0.035–0.273)
Absolute EI change, mean ± SD (range)	0.008 ± 0.063 (− 0.110 to + 0.171)
Intra-rater reliability, ICC(A,1) (95% CI)	0.990 (0.97–1.00)
Bland-Altman mean difference (95% LoA)	−0.005 (− 0.023 to + 0.013)
Pearson r, EI change vs clinician score (95% CI)	−0.084 (− 0.357 to 0.202)
Spearman r, EI change vs clinician score	−0.087
p-value (Pearson)	0.566
Visit-matched analysis (n = 122 pairs), Pearson r (95% CI)	−0.007 (− 0.185 to 0.171), p = 0.939
Subgroup baseline EI ≥ 0.08 (n = 37), Pearson r (95% CI)	−0.056 (− 0.373 to 0.273), p = 0.742

(EI = erythema index [(R − G)/(R + G+B)]; SD = standard deviation; ICC = intraclass correlation coefficient; LoA = limits of agreement; CI = confidence interval)

Longitudinal erythema index trajectories

The (Fig. 3) illustrates longitudinal EI trajectories for five representative patients over up to 23 months. Patient S demonstrated progressive EI decline from 0.239 to 0.103 over 17 months. Patient K.F. showed the largest absolute EI reduction in the cohort (0.215 to 0.044, 79.7%), corresponding to a clinician score of 9/10. Patient A.K. demonstrated a transient EI rise at one month before subsequent decline, consistent with the recognised natural proliferative phase of IH continuing briefly after propranolol initiation [15]. Per-patient data for selected patients are in (Table 2).

Table 2
Per-patient erythema index at baseline and final follow-up (10 selected patients; full dataset in supplementary material).

Patient	Visits	Baseline EI	Final EI	Δ EI (abs)	Δ EI %	Clinician score
Patient K.F.	4	0.215	0.044	+ 0.171	+ 79.7%	9
Patient B.S.	4	0.132	0.035	+ 0.098	+ 73.9%	6
Patient A.K.	6	0.124	0.041	+ 0.083	+ 67.2%	8
Patient L	7	0.275	0.122	+ 0.152	+ 55.5%	5
Patient S	5	0.239	0.120	+ 0.118	+ 49.5%	7
Patient H.K.	7	0.122	0.109	+ 0.013	+ 10.8%	9
Patient Ka.	8	0.062	0.122	−0.060	−96.8%	6
Patient R.	7	0.078	0.142	−0.065	−83.5%	8
Patient Su.	4	0.108	0.198	−0.090	−83.3%	7
Patient K.R.	5	0.033	0.143	−0.110	−338.0%	10

(EI = erythema index [(R − G)/(R + G+B)]. Green = EI reduction. Red = EI increase. Clinician score: estimated % reduction in overall lesion burden ÷ 10, recorded prospectively blinded to EI. Patient identifiers anonymised)

Correlation with clinician response scores

Absolute EI change did not correlate significantly with clinician-assessed response scores (Pearson r = − 0.084, 95% CI − 0.357 to 0.202, p = 0.566; Spearman r = − 0.087, p = 0.554; n = 49; Fig. 4). This was consistent across all three sensitivity analyses: visit-matched pairwise (n = 122 pairs, r = − 0.007, 95% CI − 0.185 to 0.171, p = 0.939); baseline EI ≥ 0.08 subgroup (n = 37, r = − 0.056, 95% CI − 0.373 to 0.273, p = 0.742).

Discussion

This pilot study demonstrates that semi-automated erythema quantification from clinical photographs of IH is technically feasible with excellent intra-rater reliability (ICC[A,1] = 0.990) in a real-world clinical cohort achieving a 96.4% photograph usability rate. To our knowledge, this is the first study to apply automated RGB-based erythema quantification to a clinical cohort of propranolol-treated IH with prospective, blinded clinician scoring. Abagge KT et al., previously demonstrated colorimetric photograph analysis feasibility in IH using a calibrated pixel-based approach [9]. Our study extends this by applying an open-source scripted pipeline to a larger propranolol-treated cohort with longitudinal follow-up of up to 23 months.

The central finding is that EI change did not correlate with clinician-assessed composite response (Pearson r = − 0.084, p = 0.566). This must be interpreted explicitly as a partial-biomarker finding. The clinician response score integrates erythema, lesion volume, and palpation characteristics; EI captures only the colour dimension. This mismatch between a colour-only metric and a composite clinical endpoint is expected and informative. It mirrors the experience in PWS literature, where significant heterogeneity in colorimetric outcome measures has been identified across 85 prospective studies [8]. Discordant cases such as Patient Ka. (EI increase, score 6) likely reflect real biological complexity — natural involution and treatment effect operating simultaneously — rather than measurement noise.

Three methodological factors likely attenuated correlation further. First, ambient photography without colorimetric calibration introduces illumination-dependent noise into RGB metrics. Sommers et al. demonstrated ICC ≥ 0.99 for ROI-based digital colour analysis under controlled calibrated conditions [11]. Our Bland-Altman LoA of − 0.023 to + 0.013 reflects within-session reproducibility under constant conditions, not the between-session variability introduced by variable clinical lighting. Abagge KT et al., addressed this directly by embedding standardised red and green reference patches beside each lesion — a practical low-cost approach we recommend for future prospective studies [9]. Second, manual rectangular ROI selection incorporates variable proportions of surrounding normal skin, diluting the erythema signal. Third, the study is underpowered for correlation testing (post-hoc power ∼20% for r = 0.30, n = 49).

Despite absent group-level correlation, individual EI trajectories showed clinically coherent patterns. Patient K.F. (79.7% EI reduction, score 9/10) and Patient A.K. (67.2%, score 8/10) illustrate cases where colour change dominated the clinical response. The transient EI rise in Patient A.K. at one month replicates a recognised natural history phenomenon: propranolol does not arrest the proliferative phase immediately, and EI capture of this biology supports construct validity [15]. Dynamic infrared thermography has demonstrated similar capacity to detect early physiological change in IH before visible structural changes occur [10]. Our EI approach offers a lower-cost, more widely deployable alternative for the colour component of this assessment.

The clinical utility of this tool extends beyond specialist monitoring. In telemedicine and hub-and-spoke care models, community practitioners who photograph lesions for remote specialist review could submit images for automated erythema quantification, enabling objective longitudinal tracking without tertiary attendance at every visit. Furthermore, objective erythema quantification may serve as a caregiver engagement instrument. Propranolol requires sustained daily administration for 6–12 months, and caregiver fatigue during the plateau phase of IH involution is a recognised clinical challenge. Presenting caregivers with a quantitative erythema trajectory — analogous to trending HbA1c in diabetes management — may reinforce treatment adherence by providing tangible evidence of biological response even when visual change appears subtle. Both hypotheses warrant prospective evaluation.

Future work should address four priorities: (i) prospective standardised photography with fixed illumination, controlled camera-to-lesion distance, and embedded colorimetric calibration targets as

used by Abagge KT et al., [9]; (ii) automated lesion segmentation using deep learning architectures to replace manual ROI selection and enable lesion area quantification; (iii) formal inter-rater reliability assessment across multiple operators; (iv) validation of a composite colour-plus-area score against standardised endpoints (HAS, PGA) in a prospective, adequately powered, multicentre cohort [5, 6].

Conclusions

Semi-automated erythema quantification from clinical photographs of infantile haemangioma is feasible using open-source Python tools, achieving 96.4% photograph usability and excellent intra-rater reliability (ICC[A,1] = 0.990) in a real-world cohort. Colour-based EI change alone was insufficient to track the clinician's composite response assessment under uncalibrated clinical photography conditions, consistent with the partial-biomarker nature of colour-only analysis. Prospective studies with standardised photography, colorimetric calibration, automated segmentation, and composite colour-plus-area scoring represent the necessary next steps. The tool is low-cost, open-source, and deployable without specialist infrastructure, with potential applications in telemedicine and caregiver engagement.

Declarations

Funding Support

No funding was received for conducting this study.

Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose.

Author Contribution

B.K. collected and analysed the data, performed image analysis, conducted the statistical analysis, and drafted the manuscript. V.J., P.G., and D.K.Y. contributed to patient management, clinical data acquisition, and manuscript review. A.K.D. conceptualised and supervised the study, developed the Python-based erythema scoring tool, interpreted the data, and critically revised the manuscript. All authors reviewed and approved the final manuscript.

References

1. Léauté-Labrèze C, Hoeger P, Mazereeuw-Hautier J et al (2015) A randomized, controlled trial of oral propranolol in infantile hemangioma. N Engl J Med 372(8):735–746. 10.1056/NEJMoa1404710
2. Haggstrom AN, Drolet BA, Baselga E et al (2006) Prospective study of infantile hemangiomas: clinical characteristics predicting complications and treatment. Pediatrics 118(3):882–887.

10.1542/peds.2006-0413
3. Léauté-Labrèze C, de la Dumas E, Hubiche T, Boralevi F, Thambo JB, Taïeb A (2008) Propranolol for severe hemangiomas of infancy. N Engl J Med 358(24):2649–2651. 10.1056/NEJMc0708819
4. Drolet BA, Frommelt PC, Chamlin SL et al (2013) Initiation and use of propranolol for infantile hemangioma: report of a consensus conference. Pediatrics. ;131(1):128–140. 10.1542/peds.2012-1691
5. Haggstrom AN, Beaumont JL, Lai JS et al (2012) Measuring the severity of infantile hemangiomas: instrument development and reliability. Arch Dermatol 148(2):197–202. 10.1001/archdermatol.2011.926
6. Janmohamed SR, Madern GC, de Laat PC, Oranje AP (2015) Educational paper: therapy of infantile haemangioma — history and current state (part II). Eur J Pediatr 174(2):259–266. 10.1007/s00431-014-2404-5
7. Weatherall IL, Coombs BD (1992) Skin color measurements in terms of CIELAB color space values. J Invest Dermatol 99(4):468–473. 10.1111/1523-1747.ep12616156
8. van Raath MI, Chohan S, Wolkerstorfer A et al (2020) Clinical outcome measures and scoring systems used in prospective studies of port wine stains: a systematic review. PLoS ONE 15(7):e0235657. 10.1371/journal.pone.0235657
9. Abagge KT, Sandri CDO, Sakai LSM, Mussato LP, Petterle RR, Oliveira de Carvalho VO (2020) Colorimetric analysis of images in the follow-up of infantile hemangiomas. Pediatr Dermatol 37(4):622–625. 10.1111/pde.14153
10. Burkes SA, Patel M, Adams DM et al (2016) Infantile hemangioma status by dynamic infrared thermography: a preliminary study. Int J Dermatol 55(10):e522–e532. 10.1111/ijd.13298
11. Sommers MS, Beacham B, Baker R, Fargo J (2013) Intra- and inter-rater reliability of digital image analysis for skin color measurement. Skin Res Technol 19(4):484–491. 10.1111/srt.12072
12. Dhua AK (2026) hemangioma-erythema-scorer [software]. GitHub; Available at: https://github.com/dhuaanjancode/hemangioma-erythema-scorer
13. Bradski G (2000) The OpenCV library. Dr Dobb's. J Softw Tools 25:120–125
14. Takiwaki H (1998) Measurement of skin color: practical application and theoretical considerations. J Med Invest 44(3–4):121–126
15. Hogeling M, Adams S, Wargon O (2011) A randomized controlled trial of propranolol for infantile hemangiomas. Pediatrics 128(2):e259–e266. 10.1542/peds.2010-0029

Figures

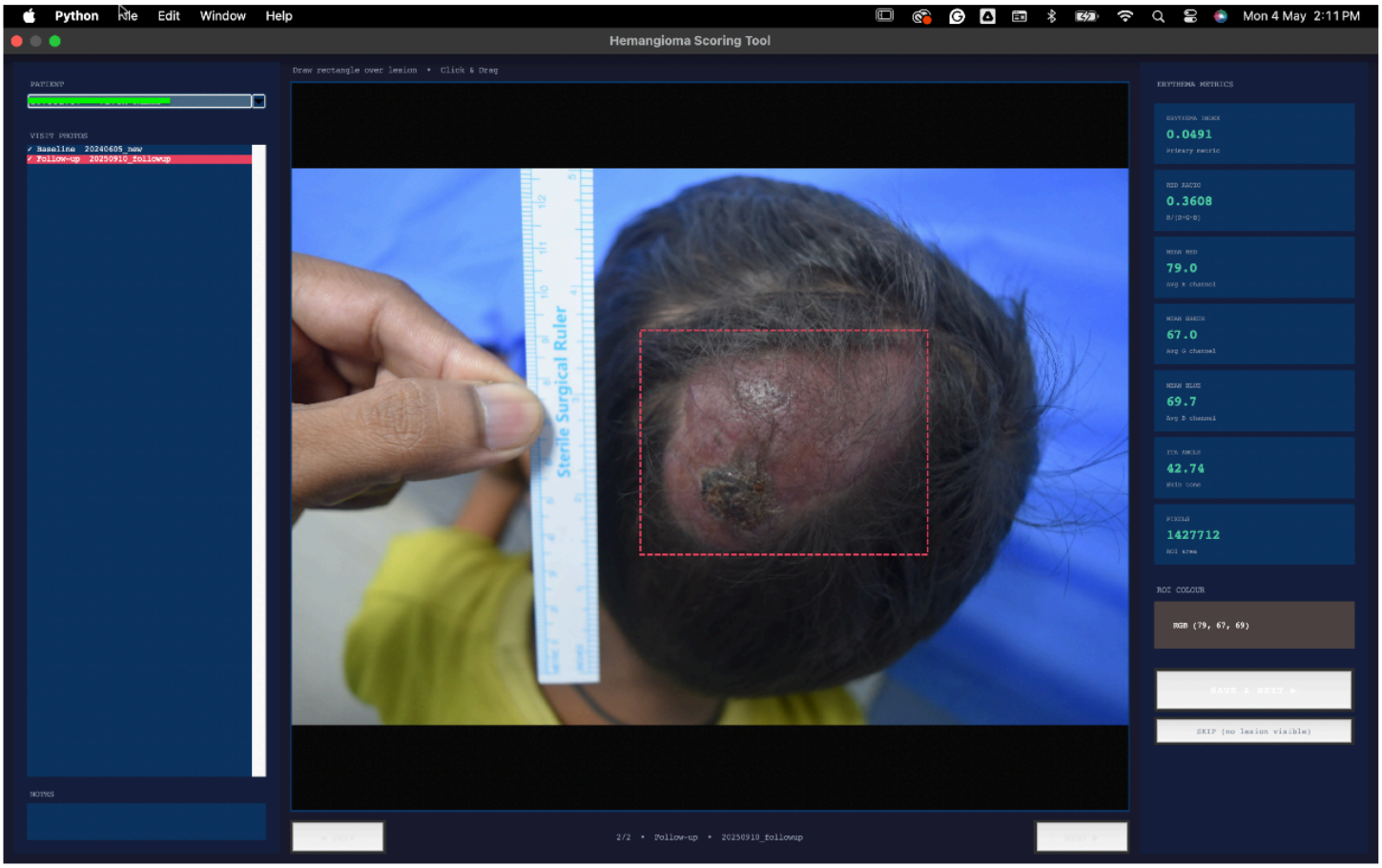


Figure 1

Screenshot of the semi-automated Python erythema scoring tool. The clinical photograph is displayed with the manually selected rectangular ROI (dashed red outline). The right panel shows real-time erythema metrics: EI (primary metric), red ratio, mean RGB values, ITA, pixel count, and mean ROI colour swatch.

(EI = erythema index; ITA = Individual Typology Angle; RGB = red, green, blue; ROI = region of interest)

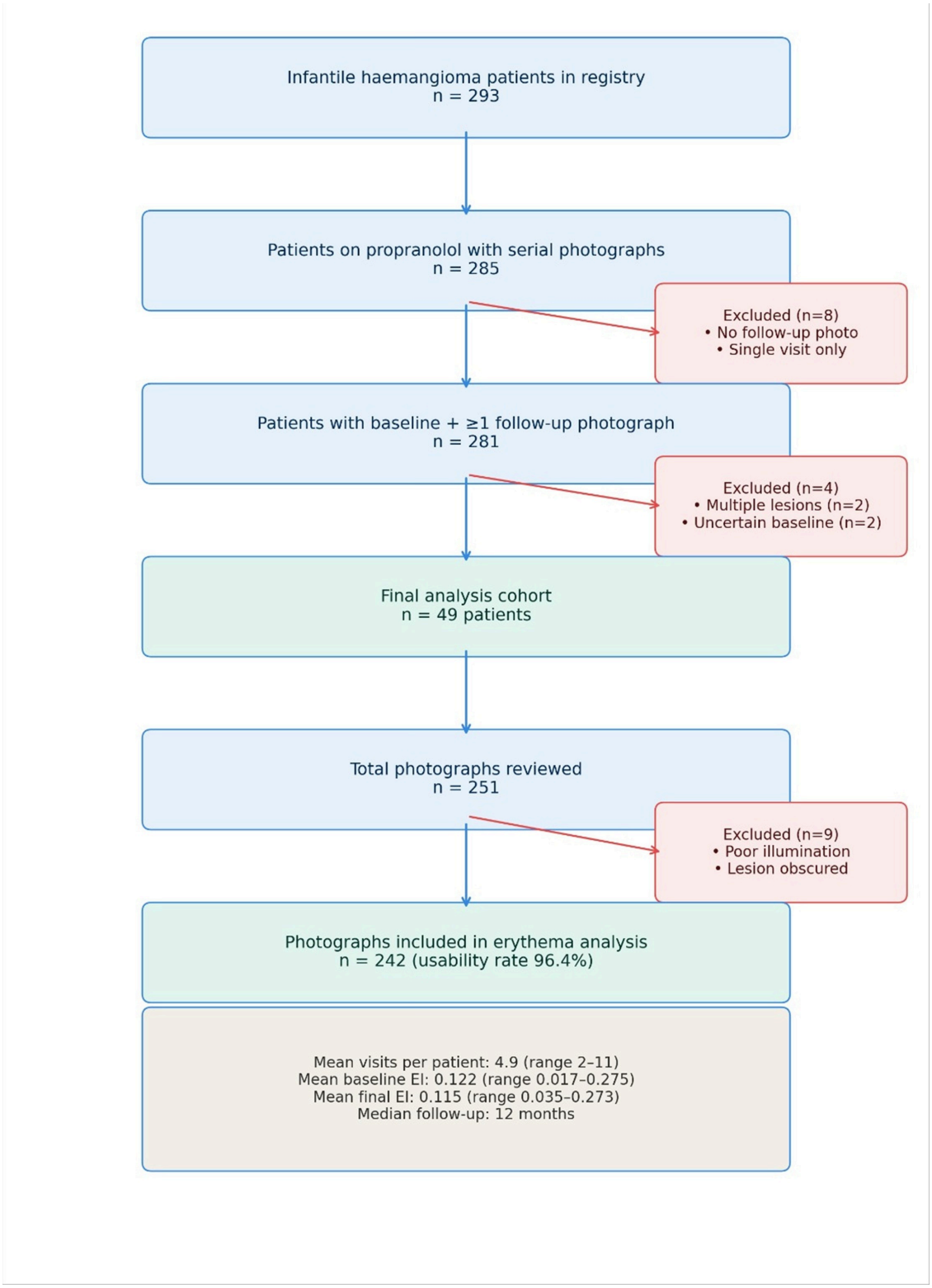


Figure 2

Longitudinal erythema index in five representative patients (of 49 total) with infantile haemangioma treated with propranolol.

Patient selection flowchart. Patients identified from a prospective infantile haemangioma photographic registry at a tertiary paediatric surgery centre. IH = infantile haemangioma; EI = erythema index.

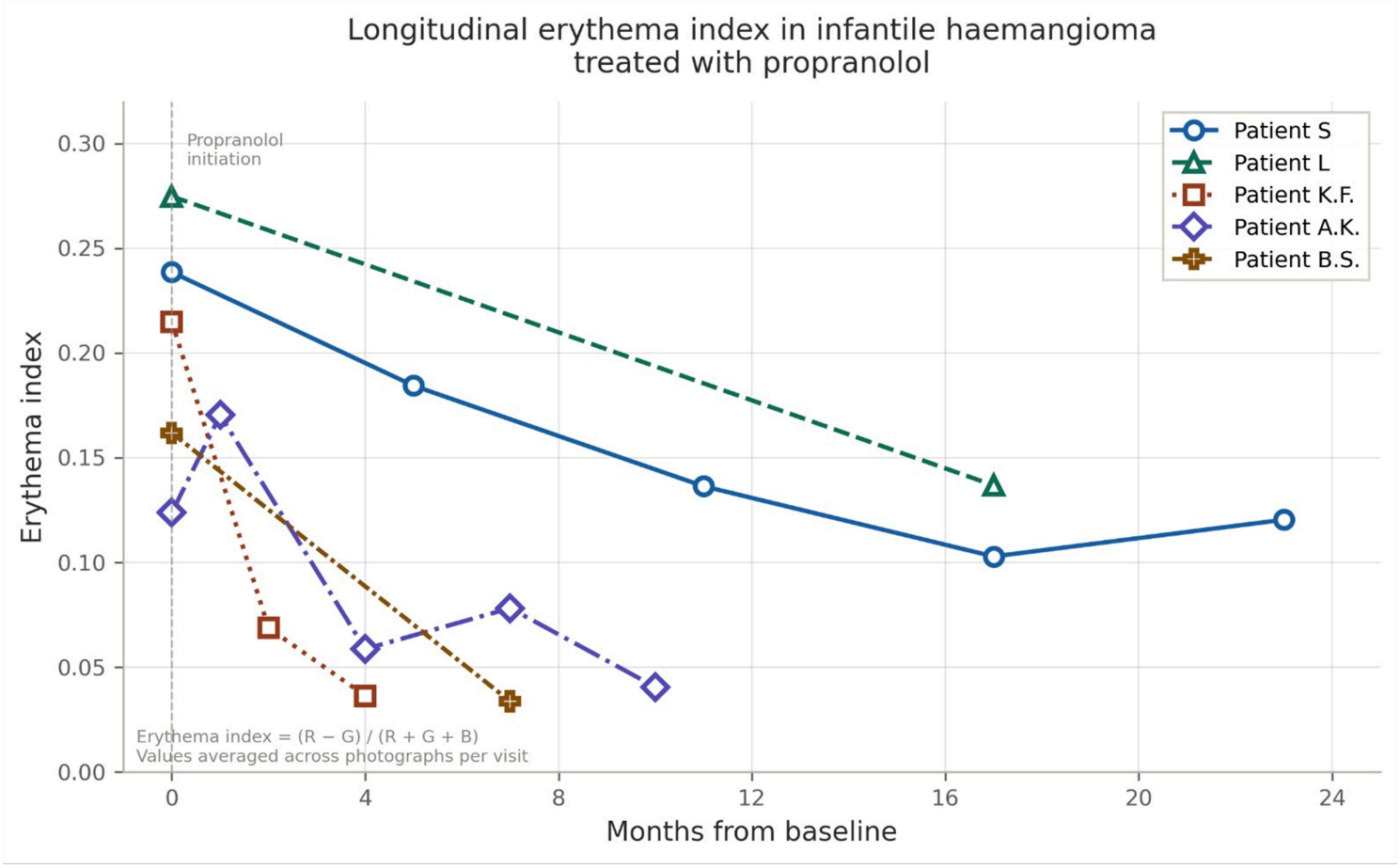


Figure 3

Longitudinal erythema index in five representative patients (of 49 total) with infantile haemangioma treated with propranolol. Each data point = mean EI from photographs at that visit. Dashed line = propranolol initiation (month 0). Patient A.K.: transient early EI rise at one month consistent with ongoing natural proliferative phase [15]. Patient K.F.: largest EI reduction in cohort (baseline 0.215, final 0.044). These are representative cases selected to illustrate biological plausibility; group-level correlation was not significant. EI = (R − G) / (R + G + B). Patient identifiers anonymised.

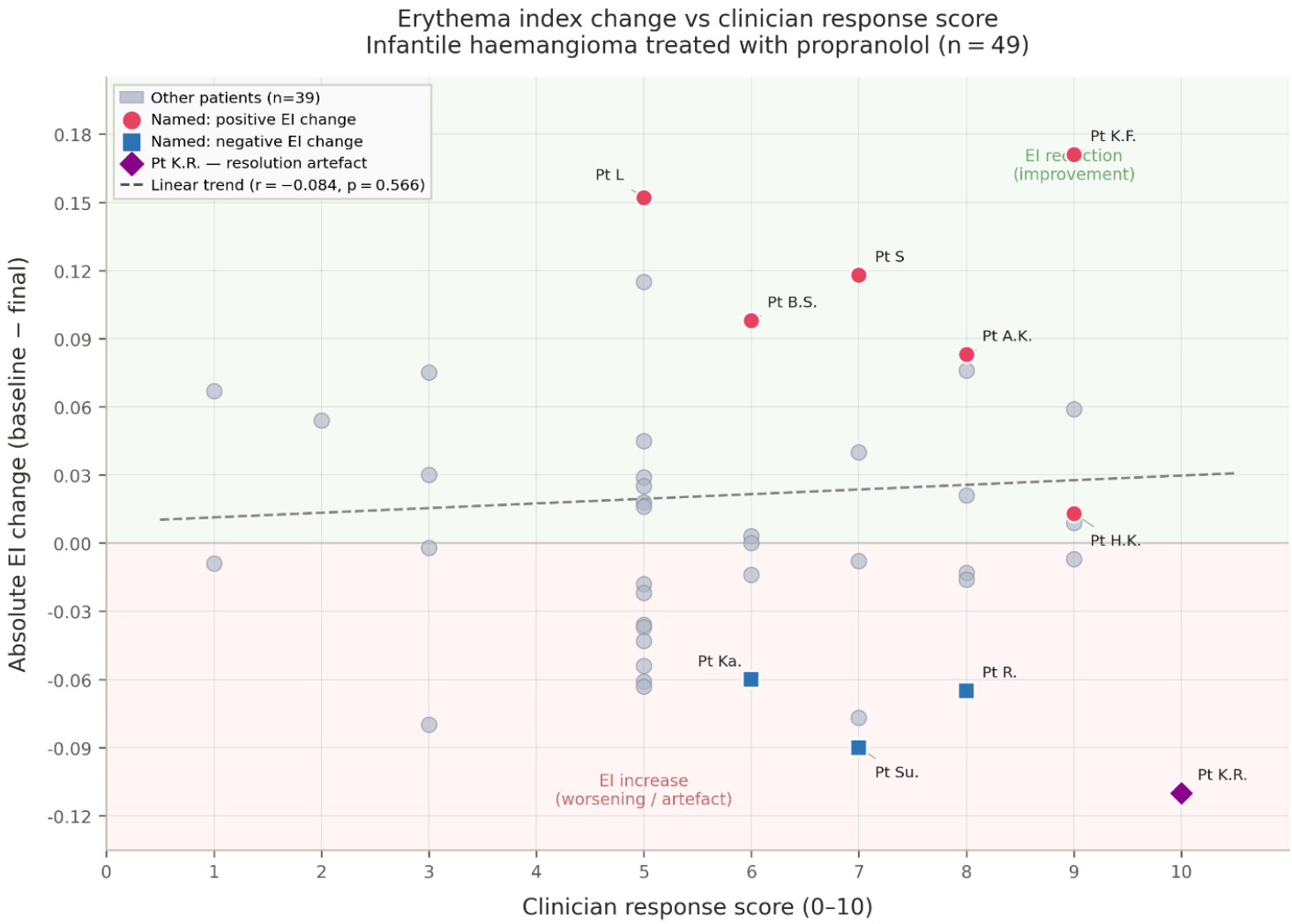


Figure 4

Scatter plot of absolute EI change against clinician response score (n = 49). Pearson r = −0.084 (95% CI −0.357 to 0.202), p = 0.566. Dashed line = linear trend. Each point = one patient.

Scatter plot of absolute erythema index (EI) change against clinician response score (n = 49 patients). Absolute EI change = baseline EI minus final follow-up EI; positive values indicate erythema reduction. Clinician response score = estimated percentage reduction in overall lesion burden ÷ 10 (0 = no change; 10 = complete resolution), recorded prospectively in clinic before any image analysis (blinded). Dashed line = linear trend (Pearson r = −0.084, 95% CI −0.357 to 0.202, p = 0.566). EI = erythema index; CI = confidence interval.