

Apparent skin tone rendering in recommended undergraduate optometric textbooks in England: a descriptive content analysis

Daniela Oehring

daniela.oehring@plymouth.ac.uk

University of Plymouth

Andrew Taylor

University of Plymouth

Charlotte Barsdell

University of Plymouth

Haleema Ajaib

University of Plymouth

Kaashish Sidhana

University of Plymouth

Katey Hooper

University of Plymouth

Research Article

Keywords: skin tone representation, medical education, optometry, Fitzpatrick scale, content analysis, textbook imagery, health equity, visual pedagogy

Posted Date: July 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9855135/v1>

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

[Read Full License](#)

Additional Declarations: Competing interest reported. The study was conducted as part of an undergraduate research project at the University of Plymouth.

Abstract

Background

Lighter skin tones are overrepresented in medical educational imagery across several disciplines, yet no study has examined optometric teaching materials. This study describes the distribution of apparent skin tone rendering in recommended undergraduate optometric textbooks used in England.

Methods

A descriptive, cross-sectional content analysis was conducted on 28 textbooks identified through purposive sampling of recommended reading lists from eight of eleven English university optometry programmes. Two independent mixed-composition teams classified 1,623 gradable graphics (1,446 photographs, 177 schematics) on the six-point Fitzpatrick skin phototype (FP) scale. Inter-rater reliability was assessed using quadratically weighted Cohen's kappa. Proportions are reported with 95% Wilson confidence intervals. Textbook-level analyses were also performed, treating the textbook as the unit of analysis. Descriptive comparisons with the 2021 Census are presented as contextual, not normative, benchmarks; no inferential hypothesis tests were conducted against the census comparator.

Results

Quadratically weighted kappa was 0.69, indicating substantial agreement on the ordinal scale, while unweighted exact agreement (61.8%) reflected moderate category-by-category concordance. Across all graphics, 78.1% were classified as light (FP I-II; 95% CI 76.1–80.1%), 18.4% as intermediate (FP III-IV; 95% CI 16.6–20.4%), and 3.5% as dark (FP V-VI; 95% CI 2.7–4.5%). Fourteen of 28 textbooks (50%) contained no images classified as FP V-VI, and no schematic diagram across the corpus was classified as FP V-VI (0/177). At the textbook level, median FP V-VI prevalence was 0.9% (IQR 0.0–3.8%), substantially below the census contextual comparator of 18.95% non-White. Findings were robust to exclusion of the single largest contributing text and restriction to books with ten or more gradable images (photographs plus schematics).

Conclusions

In this sample of recommended optometric textbooks, graphics classified as depicting dark skin tones are rare. The finding is consistent across textbooks and robust to analytical choices. These data describe a pattern that warrants attention from publishers, educators, and curriculum developers, though the study does not directly assess educational outcomes or establish a normative standard of adequate representation.

Background

Visual imagery is a cornerstone of clinical training in medicine and its allied professions, including optometry. Educational graphics serve as tools through which clinicians learn to recognise and interpret disease presentations [1]. A well-documented tendency towards lighter skin tone depiction exists across medical educational resources, with potential implications for diagnostic training, clinical competence, and health equity [2, 3].

International research has established the existence of this representational pattern. In the United States, Louie and Wilkes [4] demonstrated that while major anatomy textbooks approximated the US racial demographic composition at an aggregate level, no individual textbook achieved proportional representation. Subsequent studies confirmed the pattern across specialties: dermatology textbooks showed marked predominance of lighter skin tones [5–7, 31], with similar observations documented in surgery [8, 9], gynaecology [10], and rheumatology [11]. Recent multi-modal analyses of anatomy atlases have extended these findings to schematic illustrations [12, 13]. Broader reviews have also identified curricular gaps in medical education on skin of colour [23].

This representational pattern may have clinical consequences, although the evidence base is indirect. A lack of educational exposure to disease presentations across the spectrum of skin tones has been associated with reduced diagnostic accuracy in dermatology assessments [14, 15]. Whether similar effects apply in optometry - where conditions such as conjunctival melanosis, anterior uveitis, and lid lesions can present differently across the pigmentation spectrum - remains unstudied.

Two recent studies have examined skin tone representation in ophthalmology textbooks. Cheng, Curley and Barmettler [16] analysed three leading US ophthalmology textbooks and found that 92.1% of 2,305 images depicted lighter skin tones (Fitzpatrick I-IV). Cooper, Li and Passby [17] conducted the first UK-based evaluation, analysing four core ophthalmology texts, and found that 86.5% of images were classified as Fitzpatrick types I-III. Within the UK, Gunowa et al. [18] found that undergraduate nursing teaching materials for pressure injuries lacked skin tone diversity.

No study has examined optometric educational materials in any jurisdiction. This is a notable gap given that optometrists serve as first-contact clinicians for ocular conditions whose clinical signs may manifest differently across the pigmentation spectrum.

This study describes the distribution of apparent skin tone rendering in recommended undergraduate optometric textbooks used in England. The objectives are: (1) to quantify the distribution of Fitzpatrick-coded skin tones across textbook graphics, (2) to compare observed proportions descriptively with national demographic data as a contextual reference point, and (3) to evaluate differences between photographs and schematic diagrams. The study does not seek to establish a normative standard for adequate representation, nor does it assess downstream educational outcomes.

Methods

Study design

This study was designed as a descriptive, cross-sectional content analysis of imagery in textbooks recommended for undergraduate optometry programmes in England. As the study analysed publicly accessible educational materials formal ethical approval was not required.

Textbook sampling and selection

A purposive sampling strategy was employed. All eleven universities in England offering an undergraduate optometry degree, as listed by the Universities and Colleges Admissions Service (UCAS) for the 2022-23 academic year, were invited to share their recommended reading lists. Eight institutions provided data (response rate 72.7%). Three non-responding institutions were not systematically compared with respondents, and the sample may therefore not capture all texts in current use across England.

A textbook was included if it appeared on the recommended reading list of at least two responding universities. This threshold was chosen to balance inclusivity against idiosyncratic local recommendations, though we acknowledge its arbitrariness: a text used intensively at one institution but absent from another's list would be excluded. The strategy yielded 28 textbooks spanning clinical ophthalmology, optometric practice, anatomy, physiology, and pharmacology. No weighting was applied for curricular centrality, frequency of student consultation, or assessment relevance; the corpus therefore represents nominated texts, not verified exposure.

Operational definition: graphic identification and classification

All graphics depicting identifiable human skin were extracted and classified as either photographic images or schematic diagrams (including anatomical illustrations, line drawings, and diagrammatic representations). A graphic was considered to depict 'identifiable human skin' if a visible area of skin was present sufficient for a coder to form a judgement about apparent tonal rendering. Graphics were excluded as ungradable if: the image was in greyscale, the skin area was too small or obscured for reliable classification, the skin was partially occluded by medical devices, or the image quality was insufficient. A total of 1,642 graphics were identified, of which 19 (1.2%) were excluded as ungradable, yielding 1,623 gradable graphics (1,446 photographs, 177 schematics).

Duplicate images appearing in multiple chapters of the same textbook were counted once. Where the same image appeared across different textbooks, it was counted once per textbook, because the unit of interest is textbook content rather than unique images. Edition standardisation was not performed: the most recent edition available through each university's library system was used.

Skin tone classification instrument

The six-point Fitzpatrick skin phototype (FP) scale [22] was used for classification. This scale was originally designed to categorise UV sensitivity and erythema response, not to measure visible skin colour or racialised appearance in educational imagery. Its use here as a proxy for apparent skin tone rendering requires explicit justification and qualification.

The primary justification for retaining the Fitzpatrick scale is comparability: it is the metric used consistently by Louie and Wilkes [4], Adekun et al. [7], Cheng et al. [16], and Cooper et al. [17], and its adoption here permits direct cross-study comparison. However, its construct validity for the present purpose is limited. When coders assign a Fitzpatrick grade to a textbook image, they are not assessing UV photobiology; they are making a perceptual judgement about apparent tonal rendering in a visual artefact. For schematic diagrams, this distinction is particularly salient: illustrations are editorial products, not biological phenotypes, and the assignment of Fitzpatrick grades to schematics should be understood as a coding convenience rather than a measurement of phototype.

The study therefore measures perceived tonal rendering in educational graphics using the Fitzpatrick scale as an ordinal coding framework, not as a biological measurement instrument. Misclassification is possible in both directions, particularly at the boundaries between adjacent grades. The known limitations of the Fitzpatrick scale for darker skin tones [26, 27] may further reduce precision at the upper end of the scale. Alternative frameworks, such as the Monk Skin Tone Scale [28] and the Taylor Hyperpigmentation Scale [30], offer finer gradations and may be more appropriate for future work, but were not available in validated form at the time of data collection.

Coding procedure and inter-rater reliability

Data extraction and classification were performed independently by two pre-assigned pairs of two researchers (four coders in total). Teams were balanced to include researchers self-identifying with at least two distinct ethnic backgrounds, a deliberate design choice intended to mitigate the risk of systematic bias in skin tone perception. We acknowledge that this compositional strategy, while methodologically motivated, does not guarantee perceptual equivalence and should not be interpreted as implying that ethnic background determines coding accuracy.

Prior to data collection, all four researchers undertook a calibration exercise using a set of 30 reference images spanning the full Fitzpatrick range. The reference set was assembled by convenience from exemplar images in published sources and was not independently validated; with five images per grade across a six-grade ordinal scale, the calibration set is small relative to the known heterogeneity of darker-tone exemplars and may be insufficient to span the full practical range at FP V-VI. Coders were not blinded to the study hypotheses - a limitation inherent to the study design - but were instructed to classify each image based on apparent skin tone alone, without reference to clinical content.

Coding was performed on digital images viewed on calibrated laboratory monitors at standardised brightness settings (luminance approximately 120 cd/m², sRGB colour profile). The use of physical

textbook copies was avoided to ensure consistent display conditions, though differences in image rendering between ebook and print formats were not assessed.

Disagreements between teams were resolved through a moderated consensus process. Moderation rules were prespecified: where teams disagreed by one grade, the higher (darker) grade was adopted as a conservative default; where teams disagreed by two or more grades, a joint review was conducted. This rule has a known directional consequence: 1-grade disagreements at the FP IV/V boundary are resolved in favour of FP V, so the reported FP V-VI prevalence is, if anything, biased upwards relative to a no-moderation alternative; the substantive finding that dark-tone graphics are uncommon is therefore conservative under this rule. All analyses used moderated (consensus) scores. Inter-rater reliability was assessed on the pre-moderation paired grades using Cohen's kappa (unweighted and quadratically weighted) on the full six-point scale.

Contextual demographic comparison

The 2021 ONS Census for England reported that 81.05% of the population identified as White and 18.95% identified with other ethnic groups [19]. This figure is used as a contextual comparator - not a normative benchmark or quasi-ground-truth denominator. The mapping between Fitzpatrick phototype (a phenotypic UV-sensitivity construct) and census-based ethnic self-identification (a social construct) is conceptually imprecise: these are different constructs measuring different dimensions, and the correspondence is approximate at best. The comparison is retained because it is the standard approach in the published literature [4, 7, 16, 17] and provides a useful, if imperfect, contextual reference point. Readers should interpret census comparisons as indicative rather than definitive.

Statistical analysis

Descriptive statistics report absolute frequencies and proportions with 95% Wilson score confidence intervals. The primary unit of reporting is the textbook ($n = 28$), with image-level data presented as supporting detail. Book-level summaries include medians, interquartile ranges, and the proportion of books containing zero FP V-VI images. Book-level proportions and the ≥ 10 gradable image inclusion threshold were computed using total gradable images per textbook (photographs plus schematics); the supplementary workbook accompanying this manuscript currently lists photograph-level aggregate counts per book, and per-book schematic counts are noted in the manuscript table only. No a priori power calculation was performed: this is a descriptive content analysis aimed at characterising the available corpus rather than testing a prespecified hypothesis at a target power.

To assess the potential impact of image clustering within textbooks, we computed the intraclass correlation coefficient (ICC) for FP V-VI classification and derived the design effect as $DEFF = 1 + (\text{mean cluster size} - 1) \times ICC$. The effective sample size was estimated as $n / DEFF$. Consistent with the descriptive framing of the study, book-level FP V-VI proportions are reported as descriptive statistics (mean, median, IQR, with 95% Wilson confidence intervals for proportions); we did not conduct inferential

hypothesis tests against the 2021 Census comparator, because that comparator is treated as a contextual reference point rather than a falsifiable population parameter (see Discussion).

Sensitivity analyses excluded: (1) the single largest contributing text, and (2) books with fewer than ten gradable images (photographs plus schematics). Image-level proportions are reported with 95% Wilson confidence intervals; no inferential goodness-of-fit tests against the census comparator are reported, consistent with the descriptive framing. Between-textbook heterogeneity in FP V-VI proportions is described by the range, median, and interquartile range across the subset of textbooks with ≥ 10 gradable images.

Results

Inter-rater reliability

The two teams showed exact agreement on 61.8% of 1,623 images when scored on the full six-point scale, rising to 90.4% within one grade. Quadratically weighted Cohen's kappa was 0.69 (95% CI 0.66–0.72), indicating substantial agreement on the ordinal scale; the unweighted exact agreement of 61.8% reflects moderate category-by-category concordance, a distinct property that reflects the inherent subjectivity of placing images into one of six adjacent ordinal categories. Linearly weighted kappa was 0.53. These values are consistent with published reliability estimates for Fitzpatrick classification in educational settings.

Overall distribution of apparent skin tones

Table 1 and Fig. 1 present the full Fitzpatrick grade distribution across graphic types.

Table 1

Distribution of Fitzpatrick grades across graphic types (moderated scores, n = 1,623).

Graphic type	FP I	FP II	FP III	FP IV	FP V	FP VI	Total
Photographs	832 (57.5%)	266 (18.4%)	217 (15.0%)	75 (5.2%)	33 (2.3%)	23 (1.6%)	1,446
Schematics	167 (94.4%)	3 (1.7%)	4 (2.3%)	3 (1.7%)	0 (0.0%)	0 (0.0%)	177
Total	999 (61.6%)	269 (16.6%)	221 (13.6%)	78 (4.8%)	33 (2.0%)	23 (1.4%)	1,623
Footnote: trichotomy groupings used elsewhere in the manuscript are defined as light = FP I-II, intermediate = FP III-IV, dark = FP V-VI.							

Using the primary trichotomy, 78.1% of all graphics (95% CI 76.1–80.1%) were classified as light (FP I-II), 18.4% as intermediate (FP III-IV; 95% CI 16.6–20.4%), and 3.5% as dark (FP V-VI; 95% CI 2.7–4.5%). Fitzpatrick Grade I alone accounted for 61.6% of all images.

At the textbook level, the median proportion of FP V-VI images was 0.9% (IQR 0.0-3.8%; mean 2.4%, SD 3.4%). Fourteen of 28 textbooks (50%) contained no images classified as FP V-VI. The book-level mean of 2.4% (95% CI for mean proportion: 1.2–3.7%) sits 16.5 percentage points below the 2021 Census contextual comparator of 18.95%; all 28 textbooks recorded an FP V-VI proportion below this contextual reference.

FP V-VI prevalence varied markedly between textbooks (range 0% to 14.3%; IQR 0.0-3.8%), and 14 of 28 textbooks contained zero FP V-VI images. The intraclass correlation was small (ICC = 0.006). Using $n = 1,623$ with a mean cluster size of $1,623 / 28 \approx 58.0$, the design effect $DEFF = 1 + (\text{mean cluster size} - 1) \times ICC \approx 1.34$, giving an effective sample size of approximately 1,210.

Photographs versus schematic diagrams

In photographs ($n = 1,446$), 75.9% were classified as light (FP I-II; 95% CI 73.7–78.1%), 20.2% as intermediate (FP III-IV), and 3.9% as dark (FP V-VI; 95% CI 3.0–5.0%). In schematics ($n = 177$), 96.0% were classified as light (95% CI 92.1–98.1%), 4.0% as intermediate, and 0.0% as dark (95% CI for zero proportion: 0.0-2.1%). No schematic diagram across the entire 28-textbook corpus was classified as FP V or FP VI.

Contextual comparison with national demographics

Using the standard clinical dichotomy (FP I-IV vs FP V-VI), image-level FP V-VI prevalence was 3.5% (95% CI 2.7–4.5%), 15.5 percentage points below the census comparator of 18.95% (95% CI for difference: –16.4 to –14.6 pp). Table 2 presents the contextual comparison under both classification thresholds. Figure 2 illustrates how the choice of classification threshold alters the direction of findings.

Table 2 Contextual comparison with 2021 Census under two classification thresholds.				
Classification	Graphic type	Observed (95% CI)	Census	Direction
FP I-IV vs V-VI	Photographs	3.9% (3.0–5.0%)	18.95%	Below
FP I-IV vs V-VI	Schematics	0.0% (0.0-2.1%)	18.95%	Below
FP I-II vs III-VI	Photographs	24.1% (21.9–26.4%)	18.95%	Above
FP I-II vs III-VI	Schematics	4.0% (1.9–7.9%)	18.95%	Below

When the alternative dichotomy (FP I-II vs FP III-VI) was applied, photographs showed a higher proportion of FP III-VI skin tones (24.1%; 95% CI 21.9–26.4%) than the census comparator, while schematics remained below (4.0%; 95% CI 1.9–7.9%). This divergence illustrates the sensitivity of conclusions to the chosen classification threshold and underscores the importance of transparent threshold reporting.

Sensitivity analyses

Kanski's Clinical Ophthalmology contributed 538 of 1,623 gradable graphics (33.1%). Excluding this text, FP V-VI prevalence was 3.2% (95% CI 2.3–4.4%) compared with 3.5% overall. Restricting to the 19 textbooks with at least ten gradable images (photographs plus schematics) yielded FP V-VI prevalence of 3.5% (95% CI 2.7–4.6%). Both sensitivity analyses confirmed the robustness of the primary findings. Figure 3 provides a composite visual summary, and Fig. 4 displays book-level FP V-VI prevalence.

Discussion

Principal findings

This study provides a descriptive quantification of apparent skin tone rendering in recommended undergraduate optometric textbooks used in England. The principal finding is a pronounced concentration of imagery at the lighter end of the Fitzpatrick scale: across 1,623 gradable graphics in 28 textbooks, 78.1% were classified as FP I-II, and only 3.5% as FP V-VI. Half of all textbooks contained no images classified at the darkest end of the scale. At the textbook level, the median FP V-VI proportion was 0.9%, substantially below the 18.95% non-White population share in the 2021 Census - though the limitations of this comparison are substantial and are discussed below.

The construct validity problem

The most important methodological caveat concerns the relationship between the measurement instrument and the construct of interest. This study uses the Fitzpatrick skin phototype scale - originally designed to assess UV sensitivity - as a coding framework for apparent tonal rendering in educational images. These are not the same construct. When a coder classifies a textbook photograph as FP III, they are not determining the depicted individual's erythema response; they are making a perceptual judgement about how skin colour appears in a particular image under particular display conditions. For schematics, the gap between instrument and construct is even wider: the skin tone of an illustration is an editorial artefact, not a biological attribute.

This limitation does not invalidate the study, but it does constrain the precision with which the findings should be interpreted. The data describe a distribution of coded apparent tonal rendering, not a distribution of phototypes. The Fitzpatrick scale was retained for comparability with the existing literature [4, 7, 16, 17], but readers should note that the known limitations of this scale for darker skin tones [26, 27] may compress the upper end of the distribution, potentially underestimating the true tonal range present in the imagery.

The sensitivity of findings to classification threshold

A notable finding is the demonstration that the choice of classification threshold substantially alters the direction of the census comparison. Using the standard clinical dichotomy (FP I-IV vs FP V-VI), the corpus appears to underrepresent darker tones relative to the census. Using the alternative dichotomy (FP I-II vs FP III-VI), photographs appear to overrepresent the FP III-VI group. These contrasting conclusions from the same dataset underscore the instability of threshold-dependent inferences and

argue for reporting the full six-grade distribution in future studies, enabling readers to apply their own analytical frameworks.

The census comparison: a contextual heuristic, not a normative standard

The comparison of Fitzpatrick-coded image proportions with census ethnic self-identification data is a standard approach in this literature [4, 7, 16, 17], but its conceptual limitations are more serious than is typically acknowledged. Phenotypic skin appearance, Fitzpatrick phototype, racialised visual perception, and self-reported ethnicity are distinct constructs that do not map onto one another in a straightforward way. A person who identifies as White in a census may have skin that a coder classifies as FP IV; a person who identifies as Asian may be coded as FP II. The census comparison therefore provides a rough contextual reference point, not a denominator from which representational adequacy can be precisely judged.

We have accordingly framed the census comparison as contextual throughout this manuscript and have avoided language implying that 18.95% FP V-VI representation would constitute a correct or sufficient target. The key finding - that half of textbooks contain zero FP V-VI images and the textbook-level median is 0.9% - does not depend on any particular benchmark to be informative.

Schematic diagrams: a pattern, not a proven editorial choice

The complete absence of FP V-VI classification in schematics is a striking pattern. Unlike clinical photographs, schematics are not constrained by patient availability or clinical presentation; the skin tone of an illustration is determined by the illustrator and publisher. However, the data demonstrate absence, not intent. We cannot determine from these data whether the pattern reflects deliberate editorial policy, unconscious default, production constraints, illustrator convention, or some combination thereof. What the data do establish is that the pattern exists across all 28 textbooks and that schematics are - by virtue of their editorial nature - a category of imagery where diversification does not require new clinical material.

Comparison with adjacent disciplines

The findings are broadly consistent with the published literature in ophthalmology [16, 17], anatomy [12, 13], and dermatology [7]. The observed 3.9% FP V-VI prevalence in photographs is comparable to patterns reported across these disciplines, suggesting a cross-specialty phenomenon rather than an optometry-specific problem. The cross-disciplinary consistency strengthens the case for a coordinated response, though the present study cannot determine whether the mechanisms are shared.

Implications for education

The study did not assess learners, diagnostic performance, or educational outcomes and therefore cannot directly establish that the observed representational pattern impairs optometric education. Evidence from dermatology suggests that curricula dominated by lighter skin imagery may reduce diagnostic accuracy for darker skin presentations [14, 15], but this finding has not been replicated in optometry. The present data describe a condition of the educational environment - the representational composition of recommended textbooks - without establishing its educational consequence. Future studies should directly assess whether representational diversity in teaching materials is associated with diagnostic competence across the skin tone spectrum in optometric practice. More broadly, frameworks for evaluating equitable imagery in preclinical and clinical curricula have been proposed [24] but have not yet been applied systematically to optometric education.

Policy relevance

The General Optical Council's updated Standards of Practice (effective 1 January 2025) require registrants to promote equality and value diversity [20]. The GOC's Equality, Diversity and Inclusion Annual Report [21] further emphasises that educational institutions bear responsibility for inclusive practices. The Equality Act 2010 imposes a public sector equality duty on universities in England (the duty applies to public bodies in England, Wales, and Scotland under section 149) [25]. These regulatory and legal frameworks establish a policy context in which the composition of educational materials is relevant, though the present data do not establish legal non-compliance, nor do they demonstrate that any specific institution or publisher has breached a statutory duty. We note the policy context without drawing legal conclusions.

Strengths and limitations

This study has several strengths: a corpus spanning the majority of English optometry programmes, systematic classification with dual independent teams, prespecified moderation rules, formal inter-rater reliability assessment, reporting of confidence intervals and textbook-level analyses, and multiple sensitivity analyses.

The limitations are substantial and should temper interpretation. First, the Fitzpatrick scale is an imperfect instrument for the construct of interest, as discussed above. The scale was designed for UV photobiology, not for assessing the visual diversity of educational imagery, and its six-point ordinal structure compresses the full range of human pigmentation. The Monk Skin Tone Scale [28] or automated approaches such as STAR-ED [29] may offer improvements in future work.

Second, images are clustered within textbooks. The intraclass correlation was small ($ICC = 0.006$, design effect approximately 1.34), suggesting modest clustering, but the textbook-level analyses are nevertheless the more appropriate basis for inference and are reported as primary.

Third, the sampling strategy captures recommended texts from eight of eleven universities, but 'recommended' is not equivalent to 'used', 'consulted', or 'assessed'. The corpus may not reflect actual

student exposure. Three non-responding institutions are unrepresented.

Fourth, the census comparison is conceptually imprecise, as discussed above. It serves as a contextual reference point, not a normative standard.

Fifth, coding was performed on digital ebook versions under controlled display conditions, but rendering differences between ebook and print formats were not assessed. Coders were not blinded to study hypotheses.

Sixth, the moderation process, while prespecified, was not independently validated, and the reliability statistic ($\kappa = 0.69$) reflects pre-moderation agreement, which is moderate to substantial rather than near-perfect on this scale. The moderated scores may reduce apparent disagreement without necessarily reflecting ground truth.

Seventh, the study did not assess educational outcomes, learner competence, or patient care effects. The step from representational description to educational inference requires evidence that this study does not provide.

Eighth, the moderation rule resolves 1-grade inter-team disagreements in favour of the darker grade as a prespecified conservative default. This means the reported FP V-VI prevalence is biased upwards relative to a no-moderation alternative, and the substantive finding that dark-tone graphics are uncommon is therefore conservative under this rule. A related limitation is the calibration set: 30 reference images (≈ 5 per Fitzpatrick grade) assembled by convenience and not independently validated may be insufficient to span the practical range of darker-tone exemplars; future replications should use a validated calibration set or a finer-grained instrument such as the Monk Skin Tone Scale [28].

Finally, the cross-sectional design captures a single academic year; reading lists and textbook editions change over time.

Conclusions

In this sample of 28 recommended optometric textbooks used in English university programmes, graphics classified as depicting dark skin tones (FP V-VI) were rare, comprising 3.5% of all gradable images (95% CI 2.7–4.5%). Half of all textbooks contained no such images. The pattern was consistent across textbooks, robust to multiple sensitivity analyses, and broadly concordant with findings in adjacent disciplines. The complete absence of darker-toned classification in schematics identifies a specific category of imagery where diversification does not require new clinical material.

These findings describe a representational pattern in recommended textbook imagery. Whether this pattern has consequences for optometric education, clinical competence, or patient outcomes remains an empirical question for future research. The data provide a baseline against which change can be measured and a descriptive foundation for discussions about visual diversity in optometric education.

Abbreviations

CI

confidence interval

DEFF

design effect

FP

Fitzpatrick skin phototype

GOC

General Optical Council

ICC

intraclass correlation coefficient

IQR

interquartile range

ONS

Office for National Statistics

SD

standard deviation

UCAS

Universities and Colleges Admissions Service

UV

ultraviolet.

Declarations

Ethics approval and consent to participate

Not applicable. This study analysed publicly accessible educational materials and did not involve human participants, human data, or human tissue.

Consent for publication

Not applicable.

Availability of data and materials

Data are available from the corresponding author on reasonable request.

Competing interests

No conflict of interest to declare.

Funding

This is an unfunded project. The project was part of an Undergraduate dissertation project.

Authors' contributions

[Authors' contributions are provided in the separate, unblinded title page document to preserve blinding.]

Acknowledgements

[Acknowledgements that may identify authors or institutions are provided in the separate, unblinded title page document.]

Authors' information

[Authors' information is provided in the separate, unblinded title page document to preserve blinding.]

References

1. Dhouha B, Mouna B, Nadia B, Karim K. Image in medical education: why, when and how to use it? A critical review of the literature. *Biomed J Sci Tech Res*. 2020;26(3):19912–5.
2. Turbes S, Krebs E, Axtell S. The hidden curriculum in multicultural medical education: the role of case examples. *Acad Med*. 2002;77(3):209–16.
3. Joseph OR, Flint SW, Raymond-Williams R, Awadzi R, Johnson J. Understanding healthcare students' experiences of racial bias: a narrative review. *Int J Environ Res Public Health*. 2021;18(23):12771.
4. Louie P, Wilkes R. Representations of race and skin tone in medical textbook imagery. *Soc Sci Med*. 2018;202:38–42.
5. Kaufman BP, Guttman-Yassky E, Alexis AF. Atopic dermatitis in diverse racial and ethnic groups. *Exp Dermatol*. 2018;27(4):340–57.
6. Lester JC, Taylor SC, Chren M-M. Under-representation of skin of colour in dermatology images. *Br J Dermatol*. 2019;180(6):1521–2.
7. Adelekun A, Onyekaba G, Lipoff JB. Skin color in dermatology textbooks: an updated evaluation and analysis. *J Am Acad Dermatol*. 2021;84(1):194–6.
8. Cho DY, Kneib CJ, Shakir A, et al. Underrepresentation of racial minorities in breast surgery literature. *Ann Surg*. 2021;273(2):202–7.

9. Smith RM, Andersen ES, Powell LE, et al. It's not all White: implicit racial bias in imagery used in plastic surgery resident education. *J Surg Educ.* 2022;79(4):943–9.
10. Gwan A, Tanner-Sanders LN, Nair N, Chapple AG, Jernigan A. Equity in visual representation of vulvar conditions in major gynecology textbooks. *J Natl Med Assoc.* 2022;114(3):314–23.
11. Rana A, Witt A, Jones H, et al. Representation of skin colors in images of patients with lupus. *Arthritis Care Res.* 2022;74(11):1835–41.
12. Yang H, Xie SY, Yuan SG, Hu GY, Zhu K, Zhou L, et al. Systemic representational biases in anatomical education: a multi-modal content analysis. *Ann Anat.* 2026;265:152801.
13. Beresheim AC, Luo H, Tubbs RS. Anatomy's missing faces: an assessment of representation gaps in atlas and textbook imagery. *Anat Sci Educ.* 2024;17(6):e2432.
14. Fenton A, Elliott E, Shahbandi A, et al. Medical students' ability to diagnose common dermatologic conditions in skin of color. *J Am Acad Dermatol.* 2020;83(3):957–8.
15. Akhtar S, Chen V, Benjamin L, Mayoral F. Photodocumentation in skin of color. *Pediatr Dermatol.* 2021;38(S2):86–9.
16. Cheng T, Curley M, Barmettler A. Skin color representation in ophthalmology textbooks. *Med Sci Educ.* 2022;32(5):1143–7.
17. Cooper S, Li D, Passby L. Not seen, not heard? Representation of skin of colour in ophthalmology educational texts. *Clin Teach.* 2026;23(1):e70293.
18. Gunowa NO, Hutchinson M, Brooke J, Aveyard H, Jackson D. Pressure injuries and skin tone diversity in undergraduate nurse education. *J Adv Nurs.* 2021;77(11):4511–24.
19. Office for National Statistics. Ethnic group, England and Wales: Census 2021. London: ONS. 2022. <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity>. Accessed 15 Sep 2024.
20. General Optical Council. Standards of Practice. London: GOC. 2024. <https://optical.org/en/standards-and-guidance/standards-of-practice/>. Accessed 10 Jan 2025.
21. General Optical Council. Equality, Diversity and Inclusion Annual Report 2024-25. London: GOC. 2025. <https://optical.org/en/about-us/equality-diversity-and-inclusion/>. Accessed [Daniela to insert access date before submission].
22. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol.* 1988;124(6):869–71.
23. Syder NC, Omar D, McKenzie S, et al. Gaps in medical education curricula on skin of color. *J Am Acad Dermatol.* 2022;89(5):885–92.
24. Martin GC, Kirgis J, Sid E, Sabin JA. Equitable imagery in the preclinical medical school curriculum. *Acad Med.* 2016;91(7):1002–6.
25. Equality Act. 2010. c.15. London: The Stationery Office; 2010. <https://www.legislation.gov.uk/ukpga/2010/15/contents>. Accessed 15 Sep 2024.

26. Ware OR, Dawson JE, Shinohara MM, Taylor SC. Racial limitations of Fitzpatrick skin type. *Cutis*. 2020;105(2):77–80.
27. Everett JS, Budescu M, Sommers MS. Making sense of skin color in clinical care. *Clin Nurs Res*. 2012;21(4):495–516.
28. Monk EP. Monk Skin Tone Scale. Open Science Framework; 2023. <https://osf.io/pdf5c/>. Accessed 20 Oct 2024.
29. Tadesse GA, Cintas C, Varshney KR, et al. STAR-ED: skin tone analysis for representation in educational materials using machine learning. *NPJ Digit Med*. 2023;6(1):151.
30. Taylor SC, Arsonnaud S, Czernielewski J. The Taylor Hyperpigmentation Scale. *Cutis*. 2005;76(4):270–4.
31. Reilley-Luther J, Shin TM, Gumila B, Lipoff JB. Representation of Fitzpatrick skin type in dermatology textbooks. *Dermatol Online J*. 2020;26(12):4.

Figures

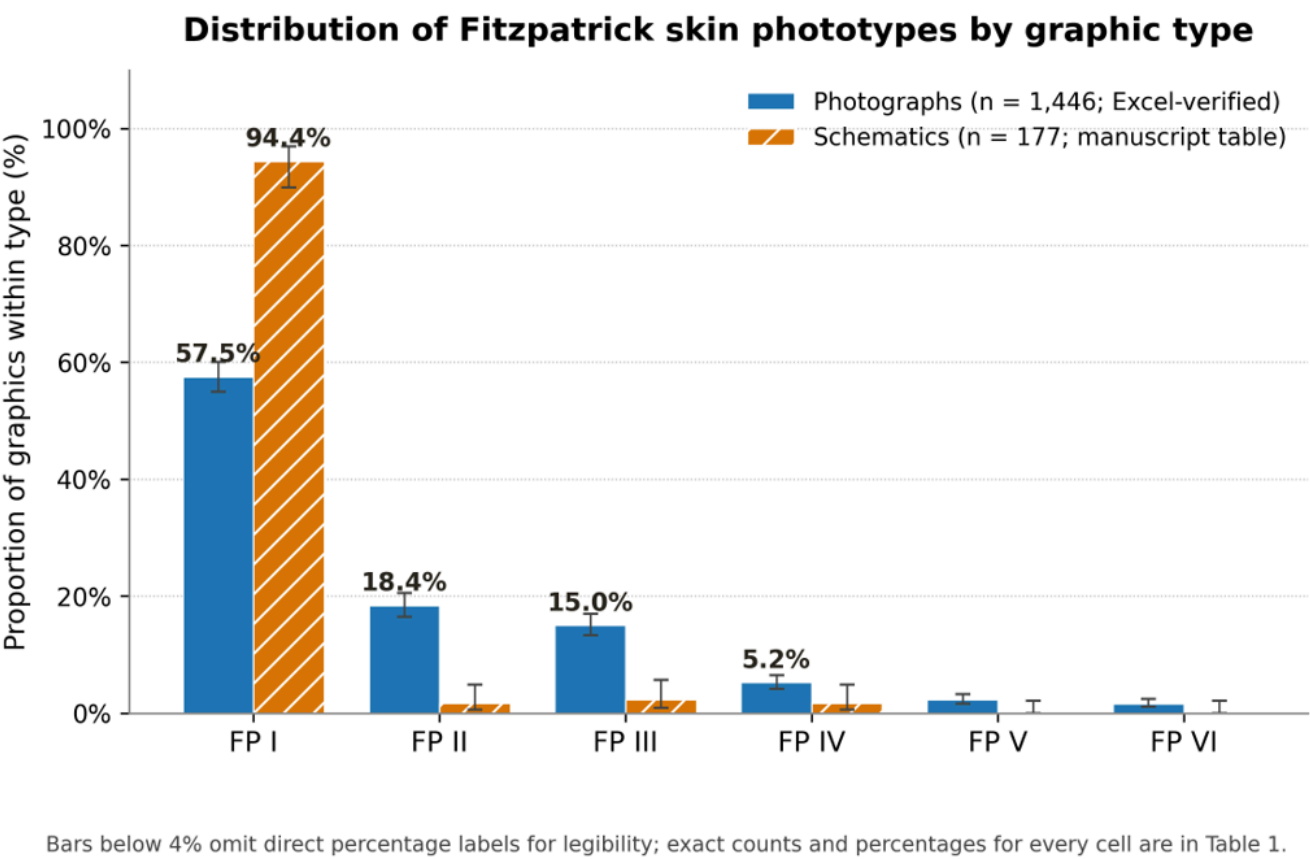


Figure 1

Distribution of Fitzpatrick skin phototypes by graphic type. Grouped bar chart comparing the proportion of photographs (n = 1,446) and schematic diagrams (n = 177) classified at each Fitzpatrick grade (I-VI) with

95% Wilson confidence intervals.

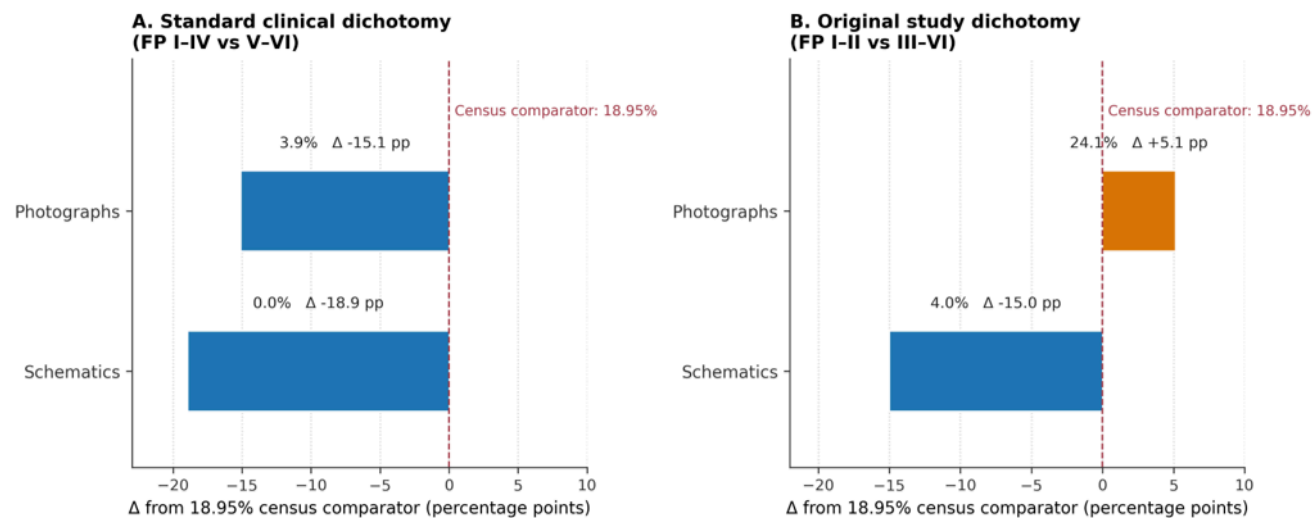


Figure 2

How classification threshold changes the narrative. Diverging bar charts showing deviation of observed proportions from the 2021 Census contextual comparator (18.95% non-White). Panel A: standard clinical dichotomy (FP I-IV vs V-VI). Panel B: alternative dichotomy (FP I-II vs III-VI). Bars indicate proportions below the comparator (left) and above (right).

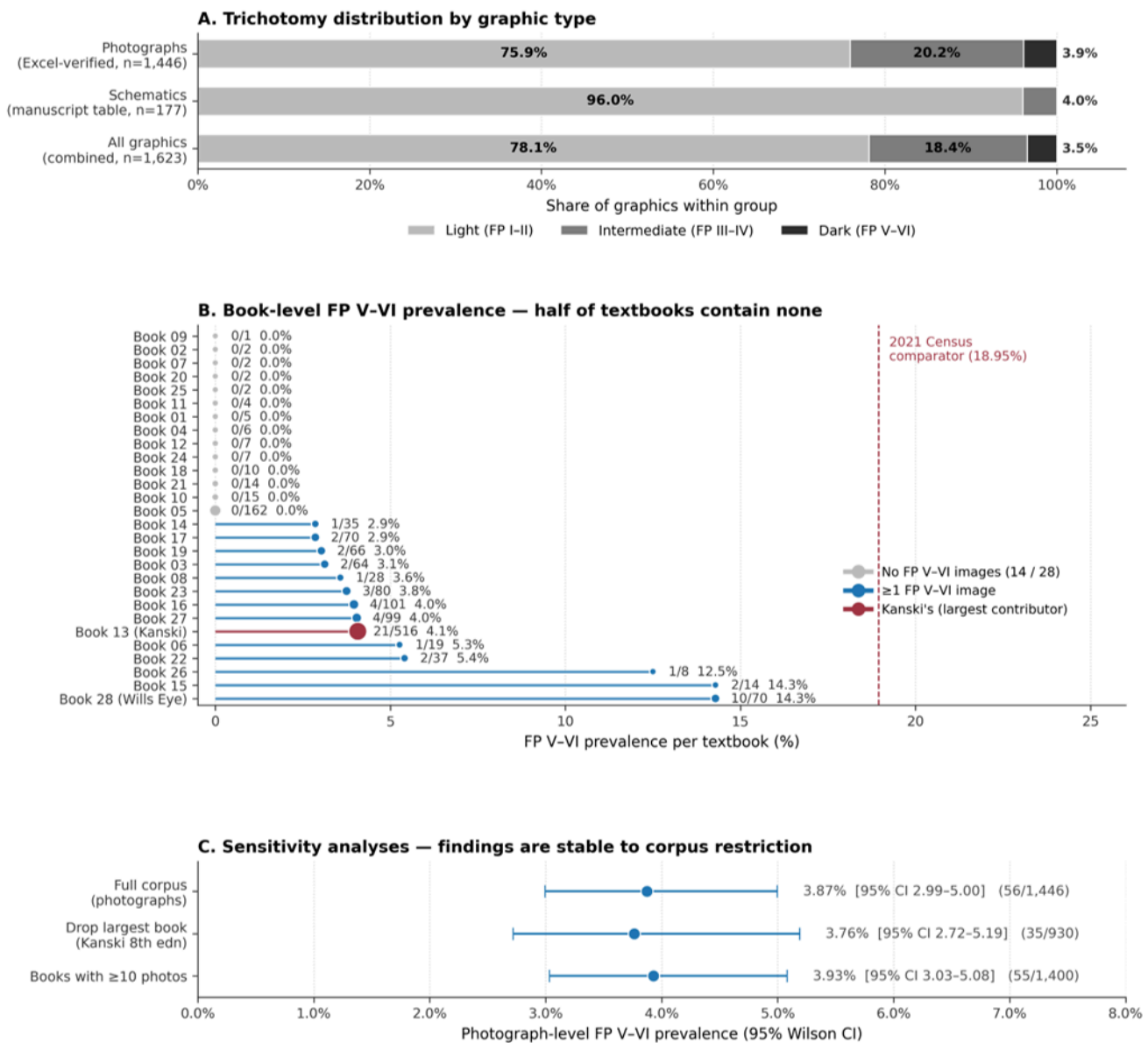


Figure 3

Composite visual summary. Panel A: 100% stacked bar chart of the trichotomy distribution (light FP I-II / intermediate FP III-IV / dark FP V-VI). Panel B: book-level lollipop chart of FP V-VI prevalence with the 18.95% census contextual comparator marked. Panel C: sensitivity analysis with and without the single largest contributing text and restricted to books with ≥10 gradable images.

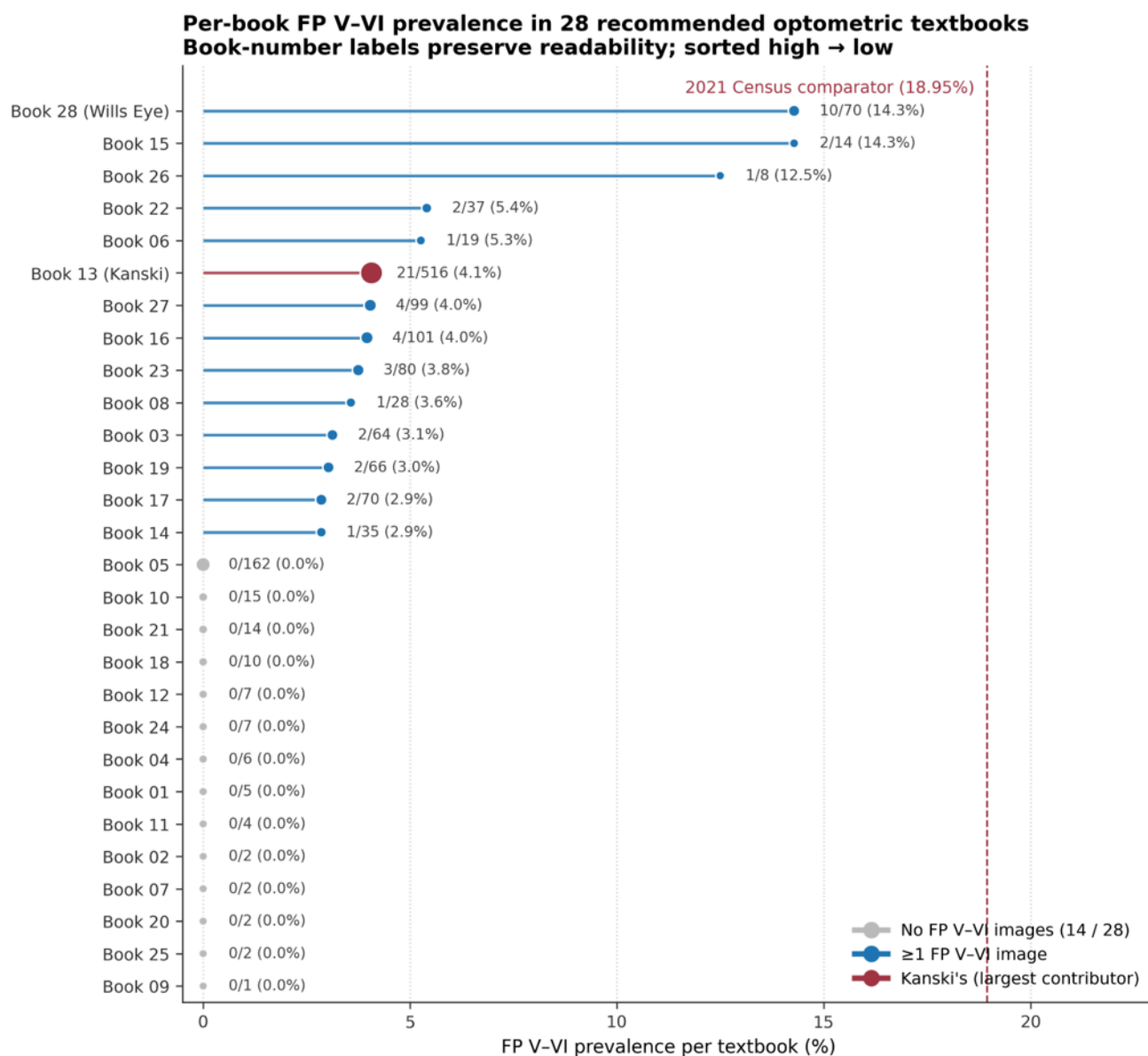


Figure 4

Book-level FP V-VI prevalence (lollipop). Per-textbook FP V-VI proportion against the 18.95% census contextual comparator, ordered to highlight between-textbook heterogeneity. Fourteen of 28 textbooks contain zero FP V-VI images.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [04STROBECrossSectionalChecklistBMC.docx](#)