

Large Multiethnic Comparison Benchmark of the Fitzpatrick and Monk Skin Tone Scales

Ross O'Hagan MD, Divija Sharma BA, Benjamin Ungar MD

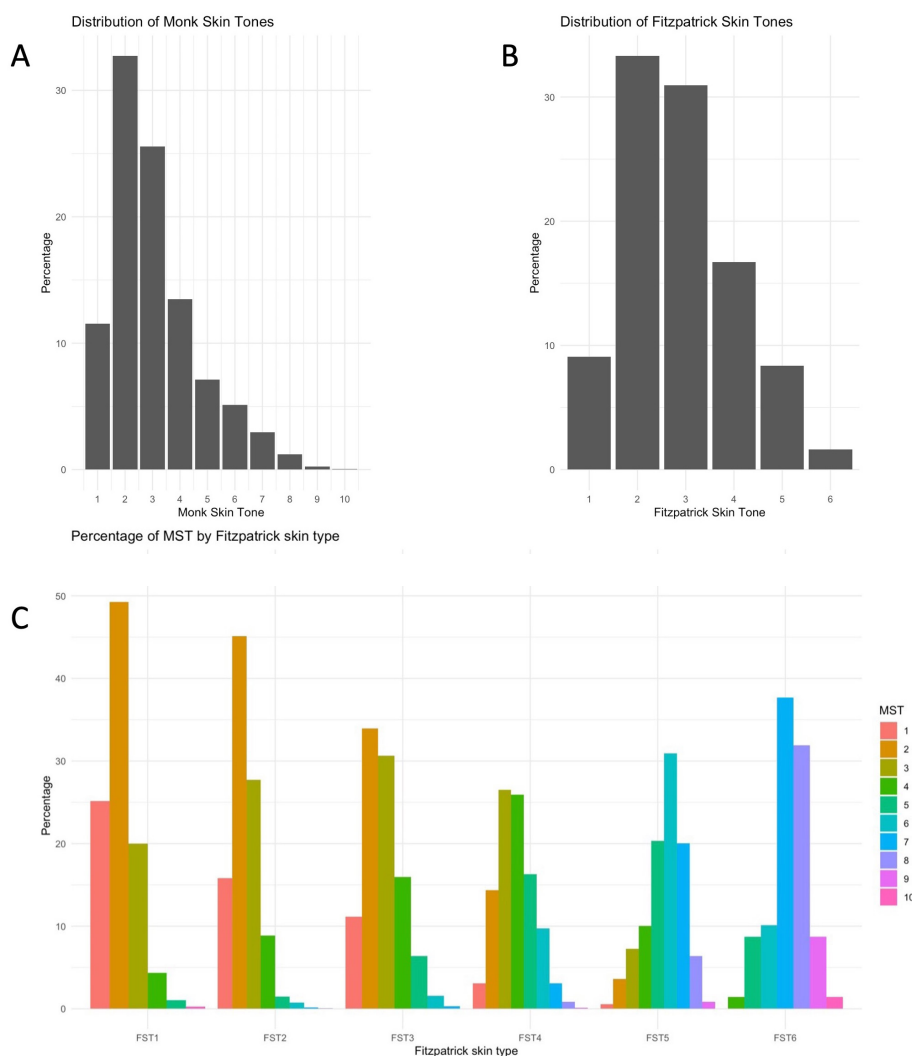
Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, NY

To the Editor:

The Fitzpatrick scale (FST) was developed to classify skin types based on responses to ultraviolet light, but is now commonly used in dermatology as a surrogate for skin tone. The FST does not adequately account for the diversity in patients with skin of color (SOC), however, limiting its clinical utility.^{1,2,3} In response to these shortcomings, the Monk Skin Tone (MST) Scale, a 10-shade grading scale, was developed as a more inclusive skin tone assessment.⁴ However, the relationship between FST

and MST is not well known. While clinical validation of MST is limited, there are efforts by Google to use this scale to help train their artificial intelligence (AI) models. Recently, a partnership between Google and Stanford researchers generated the Skin Condition Image Network (SCIN) dataset, which includes extensive dermatological images from patients recruited from Google search advertisements.⁵ Our study utilizes this dataset to benchmark the relationship between FST and MST across different evaluators, ethnic backgrounds, and skin types.

FIGURE 1. (A) MST score distribution, (B) FST score distribution, (C) Percentage of MST by dermatologist-rated FST.



SCIN includes images with dermatologist-rated FST, patient self-reported FST, and MST assigned by 2 trained non-clinical annotators, 1 from the United States (US) and the other from India. We analyzed the 4292 patient images with all FST and MST scores available. Statistical analyses were performed using R software (Version 4.2.2).

The Spearman correlation between dermatologist-rated FST and US evaluator-rated MST scores was strong and significant ($\rho=0.541$, $P<0.001$); correlations between self-reported FST and US MST scores were significant but weaker ($\rho=0.299$, $P<0.001$). There was a moderate correlation between self-reported and dermatologist-assigned FST ($\rho=0.318$, $P<0.001$). Despite the strong correlation between dermatologist-rated FST and MST scores, there was a range of different MST scores for each FST score (Figure 1). For FST 1, the most common MST score was 2 (out of 10). For an FST 6, the most common MST score was 7 (out of 10). Within racial groups, the correlations between self-reported FST and US MST were only significant among white and Hispanic patients (Table 1). The internal consistency of MST scores between the U.S. and Indian trained evaluators was robust ($\rho=0.640$, $P<0.001$).

The results of this study provide insights into relationships between FST, a widely used skin tone assessment method, and the newly developed MST. Skin tone self-assessments using FST were less correlated to trained MST scores than dermatologist-assessed FST, with a relatively weak correlation between self-assessed and dermatologist FST, suggesting that the FST scale requires training and may be less intuitive for the general public. Corresponding data to compare dermatologist and self-reported MST are lacking, requiring additional research. Our study found a strong correlation between MST and dermatologist-assigned FST and internal consistency of MST amongst different reader ethnic backgrounds.

TABLE 1.

Correlations Between Self-Reported Fitzpatrick scores and US Trained Monk Scores by Race			
Self Reported Fitzpatrick scores and US Trained Monk Scores by Race			
Racial Groups	N	Rho	P-value
White	1582	0.208	$P<0.001$
American Indian	57	0.251	$P=0.091$
Asian	84	0.194	$P=0.137$
African American	264	0.105	$P=0.216$
Hispanic	263	0.207	$P=0.003$
Middle Eastern	6	-0.207	$P=0.695$
Other	13	0.362	$P=0.247$
Two or More	69	0.236	$P=0.077$

Although dermatologist-assigned FST scores are correlated with Monk Skin Tone MST scores, notable divergences exist between the two systems. Our study was limited by insufficient power to detect significant associations within several ethnic groups and did not include comparisons between self-reported and dermatologist-assigned MST scores. Additionally, participant recruitment via online advertisements may limit the generalizability of our findings. As MST is increasingly adopted in machine learning models, it is essential for dermatologists to serve as key stakeholders in the evaluation of skin tone classification systems. Identifying gaps in knowledge surrounding MST and its relationship to FST is critical for ensuring accurate, equitable applications in clinical and technological contexts.

DISCLOSURES

BU is an employee of Mount Sinai and has received research funds (grants paid to the institution) from: Bristol Myers Squibb, Incyte, Rapt Therapeutics, Pfizer, and Sanofi. He is also a consultant for AbbVie, Arcutis Biotherapeutics, Bristol Myers Squibb, Botanix Pharmaceuticals, Castle Biosciences, Ebla Holdco, Fresenius Kabi, Galderma, J&J, Leo Pharma, Lilly, Pfizer, Primus Pharmaceuticals, Sanofi, and UCB. The other authors have no conflicts of interest to disclose.

REFERENCES

1. Goon P, Banfield C, et al. Skin cancers in skin types IV–VI: does the Fitzpatrick scale give a false sense of security? *Skin Health Dis.* 2021;1(3):e40. doi:10.1002/ski2.40

2. Pichon LC, Landrine H, et al. Measuring skin cancer risk in African Americans: is the Fitzpatrick Skin Type Classification Scale culturally sensitive? *Ethn Dis.* 2010;20(2):174-179.

3. Harvey VM, Alexis A, et al. Integrating skin color assessments into clinical practice and research: a review of current approaches. *J Am Acad Dermatol.* Published online February 9, 2024. doi:10.1016/j.jaad.2024.01.067

4. Monk E. The Monk Skin Tone Scale (MST). *SocArXiv.* Published March 23, 2023. doi:10.31235/osf.io/pdf4c

5. Ward JA, Wang J, et al. Crowdsourcing dermatology images with Google Search Ads: creating a real-world skin condition dataset. *arXiv.* Published February 28, 2024. doi:10.48550/arXiv.2402.18545

AUTHOR CORRESPONDENCE

Benjamin Ungar MD
E-mail:..... benjamin.ungar@mountsinai.org