

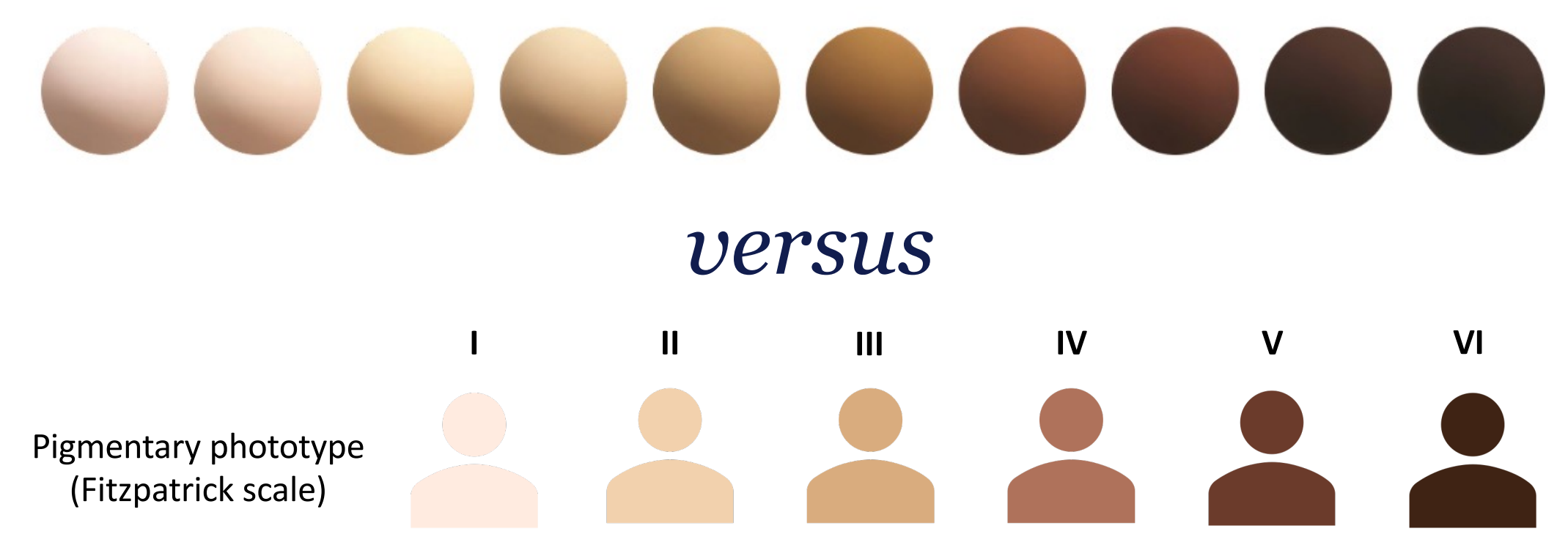
Raising the standard for skin tone labels: evaluating the Monk and Fitzpatrick scales in a multiethnic cohort

Christos Diamantidis (1), Åsa Ingvar (2), Kai Munte (3), Tamar Nijsten (1), Marlies Wakkee (1)

1. Department of Dermatology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, the Netherlands.
2. Department of Dermatology, Skåne University Hospital, Department of Clinical Sciences, Lund Skin Cancer Research Group, Lund University, Lund, Sweden.
3. Department of Dermatology, Maastad Hospital, Rotterdam, the Netherlands.

Introduction

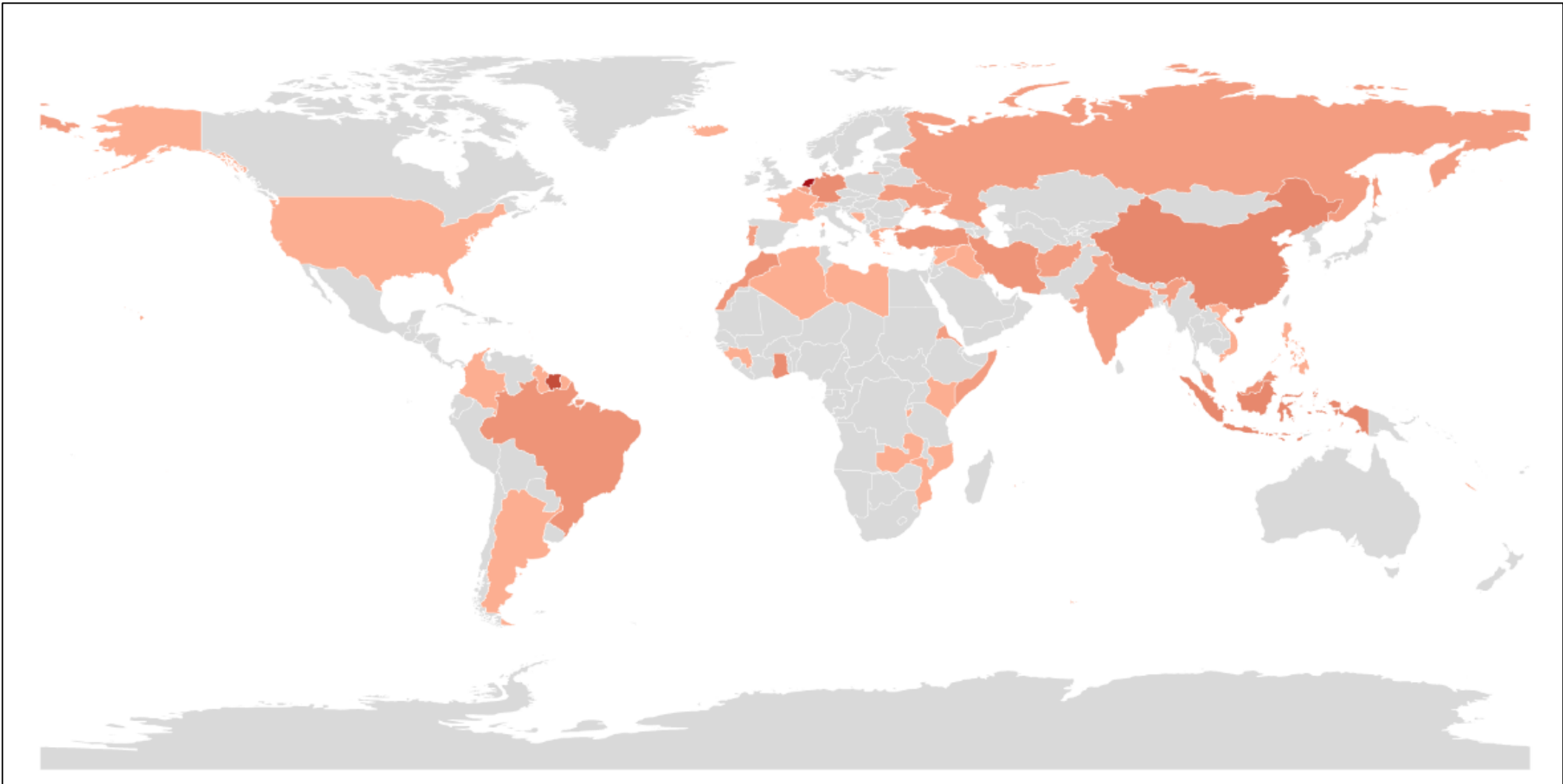
- Dermatologic AI often underperforms on darker skin, and must be shown to perform equivalently across the skin tones of its intended-use population.¹
- Skin tone labels are usually based on the Fitzpatrick scale (FST): applied inconsistently, poorly granular for dark skin, unreliable when self-reported.²
- We investigated whether the Monk skin tone scale (MST) better correlates with pigmentation, benchmarked against spectrophotometry.³



Methods & Patients

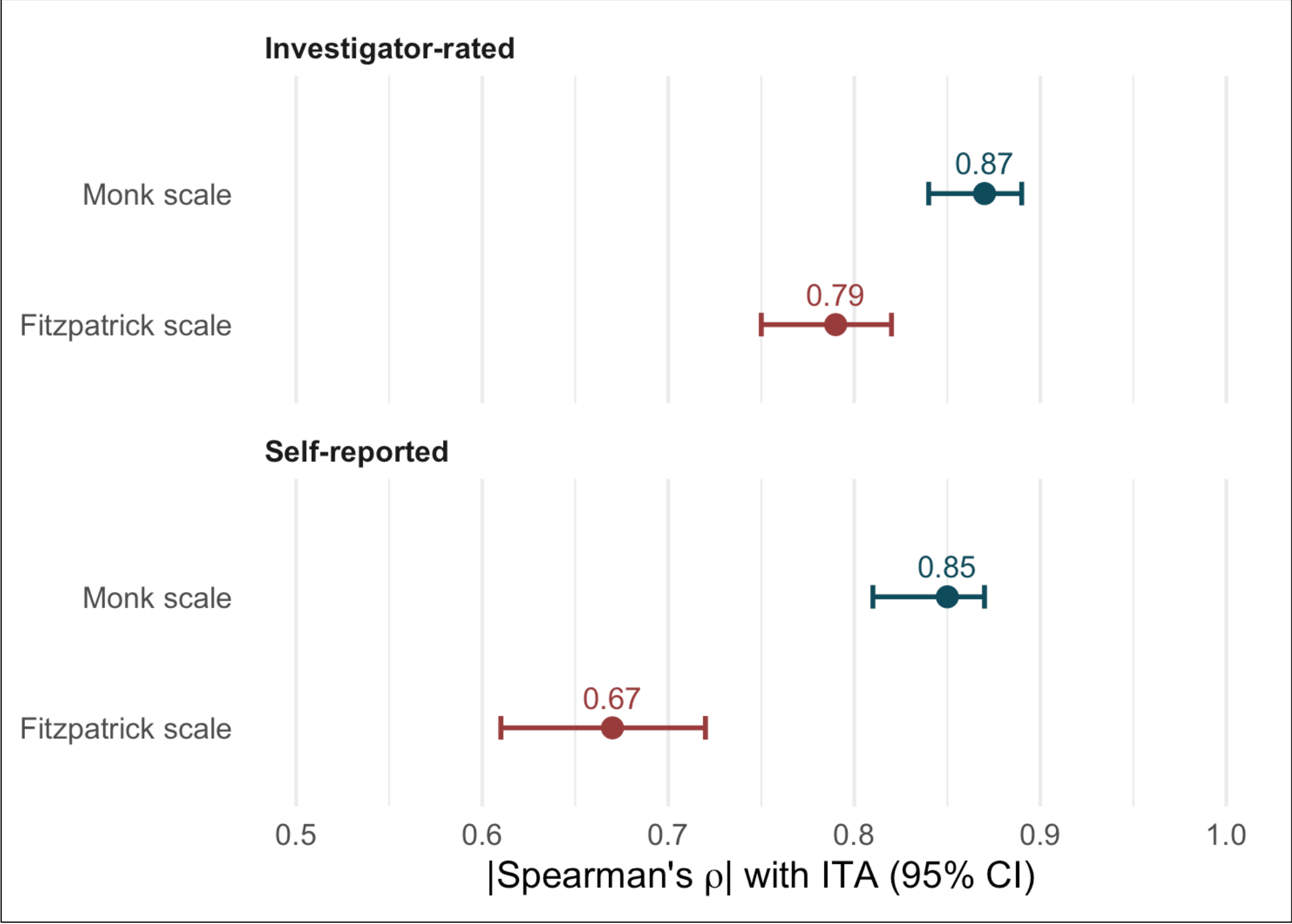
- n=517 consenting adults in dermatology outpatient clinics of Erasmus MC, Maastad Hospital
 - Targeted outreach to underrepresented communities
 - MST, FST determined by one of three trained investigators
 - Participants then self-rated with both scales (blinded)
 - Reference standard: spectrophotometry. Individual Typology Angle (ITA $\approx$ pigmentation) at the dorsal forearm.
-
- 44 countries of birth; 28% of sample born abroad
 - FST I-II 29%, III-IV 43%, V-VI 28%
 - MST 1-4 54%, 5-7 33%, 8-10 13%

Countries of birth in sample

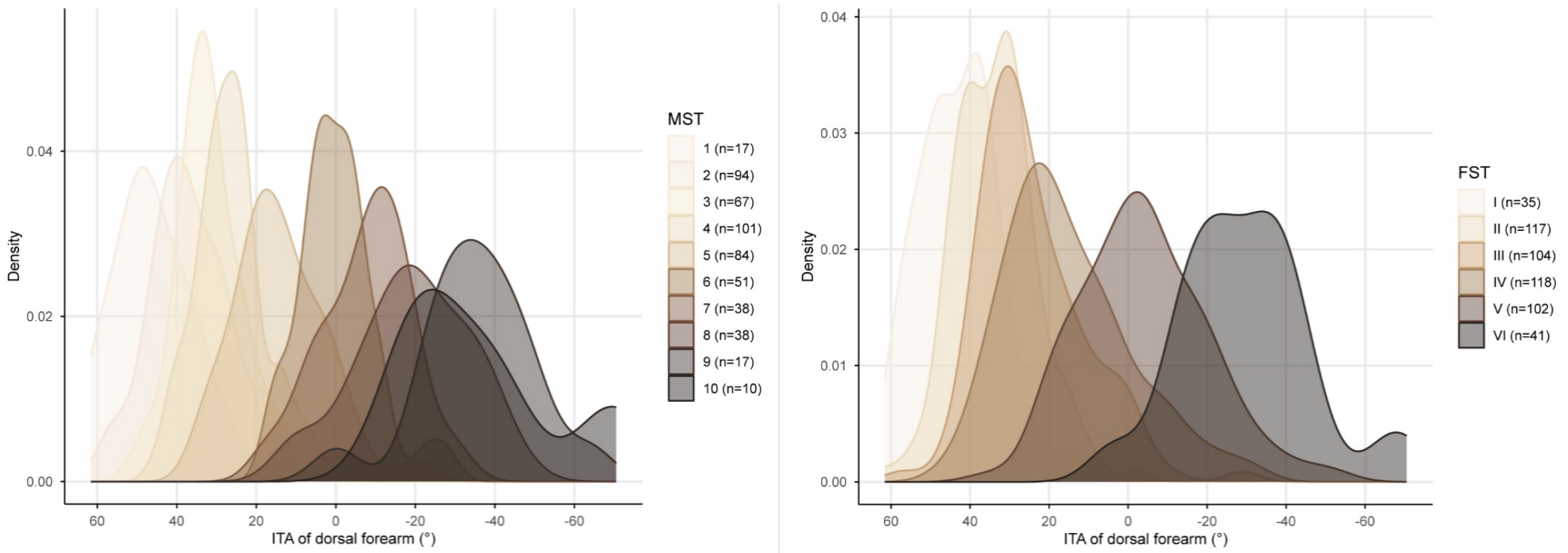


Results

The Monk scale correlates more strongly with pigmentation than the Fitzpatrick scale, particularly when self-reported

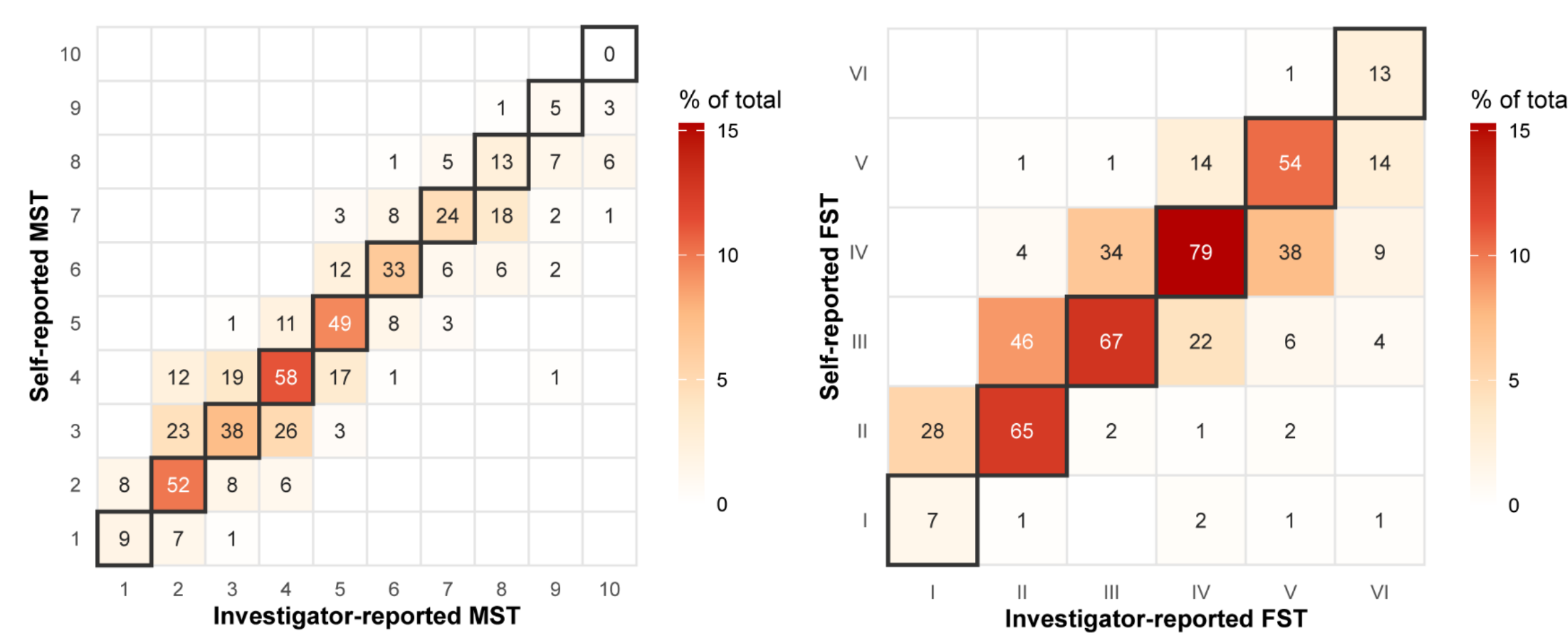


Adjacent categories of both scales overlap considerably in their objectively measured pigmentation ranges



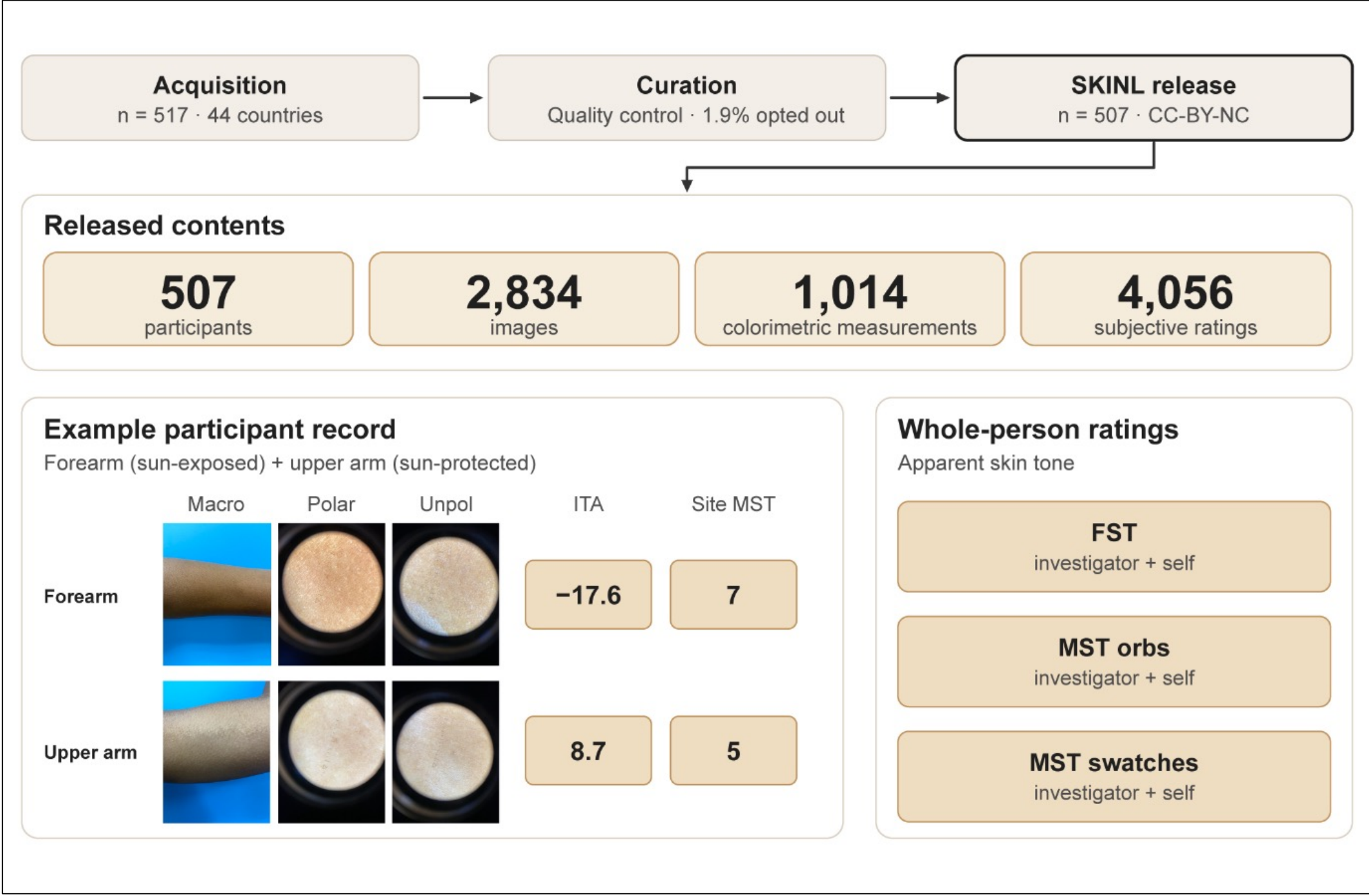
- An ideal scale would have minimally overlapping categories
- Within-category spread barely differed (mean SD_{MST} 11.8° versus mean SD_{FST} 13.7°)
- Under the FDA's Monk scale groupings⁴, 67.1% of MST_{5-7} overlapped with MST_{1-4} and 63.0% with MST_{8-10}

Untrained participants show excellent agreement with investigator-determined Monk labels



- $K_{w,quad}$ 0.91 for Monk versus 0.76 for Fitzpatrick
- Large disagreements (≥ 3 points) more common with Fitzpatrick
- Disagreement concentrated at the extremes of both scales

SKINL dataset for digital skin tone research



Conclusions

- The Monk scale aligned more closely with measured pigmentation than Fitzpatrick, supporting its use as a subjective skin tone label.
- Self-reported Monk ratings retain near investigator-level accuracy, offering a scalable source of skin tone labels for patient-facing AI.
- Where feasible, objective measurements should be preferred over either scale for benchmarking dermatologic AI across skin tones.