

An Interrater and Intrarater Reliability Study of 3 Photographic Scales for the Classification of Perioral Aesthetic Features

JOEL L. COHEN, MD,* JANE THOMAS, AAS,[†] DEEPAI PARADKAR, PHD,[†] ADAM ROTUNDA, MD,^{‡§}
PATRICIA S. WALKER, MD, PhD,^{||} FREDERICK C. BEDDINGFIELD, MD, PhD,^{‡¶} ASHA PHILIP,[#]
PAULA G. DAVIS,[#] AND RAMANA YALAMANCHILI[#]

BACKGROUND Validated aesthetic rating scales for the perioral area provide objective evaluations for clinical trials and practice.

OBJECTIVE To confirm the reliability of 3 scales for evaluating dermal filler and neurotoxin treatments of the perioral area.

MATERIALS AND METHODS Three lip-specific photographic scales were developed from standardized 2-dimensional images to evaluate Perioral Lines at Rest (POL), Oral Commissures (OCS), and Perioral Lines at Maximum Contraction (POLM) severity scales. Each 4-grade scale (none to severe) had 3 representative images per grade. Physician validators rated volunteers on each scale (2 rounds of live review). Volunteers provided 2 series of self-assessments. Physician and subject intrarater reliability were based on the comparison of round 1 and round 2 scores (mean weighted kappa coefficient). Other measures were physician interrater agreement (intraclass correlation) and subject/physician interrater agreement (Pearson correlation).

RESULTS Physician intrarater agreement was almost perfect or substantial (POL, 0.725; OCS, 0.789; POLM, 0.826). Overall, physician interrater agreement was almost perfect for all 3 scales and ranged from moderate to substantial by grade. Subject intrarater agreement and subject/physician interrater agreement were substantial.

CONCLUSION All scales demonstrated a high degree of intrarater and interrater reliability during the validation process. Physician concordance was good; subject ratings were reliable and comparable to physician assessments.

