OPEN

Development and Validation of the Clinician Reported and Patient Reported Photonumeric Scales to Assess Buttocks Cellulite Severity

Joel L. Cohen, MD,*† Neil S. Sadick, MD,‡ M. Todd Kirby, MD,§ Michael P. McLane, PhD,§ William R. Lenderking, PhD,‖ Randall H. Bender, PhD,‖¶ Jun Chen, PhD,‖ David Hurley, MD,§ Naomi B. Knoble, PhD,‖ Genzhou Liu, PhD,§ Qinfang Xiang, PhD,§ Jeffrey A. Davidson, PhD,§ and Mitchel P. Goldman, MD#

BACKGROUND The Clinician Reported Photonumeric Cellulite Severity Scale (CR-PCSS) and Patient Reported PCSS (PR-PCSS) are newly developed tools for assessing cellulite severity.

OBJECTIVE To report on the reliability, validity, and ability to detect a change in cellulite severity on the buttocks of adult women with the CR-PCSS and PR-PCSS.

MATERIALS AND METHODS Content validity of both scales was established through concept elicitation and cognitive interviews. Test–retest reliability was evaluated, and intra-rater (both scales) and inter-rater (CR-PCSS only) reliability were estimated using intraclass correlation coefficients (ICCs) for agreement and consistency. Ability to detect a change was determined using the Subject–Global Aesthetic Improvement Scale (GAIS) or Investigator-GAIS as anchors.

RESULTS For the CR-PCSS (n = 6) at baseline and Day 2, the mean interrater ICCs were ≥0.70 and mean intrarater ICCs (95% confidence interval [CI]) were ≥0.81 (0.72–0.90) for both buttocks. For the PR-PCSS (n = 99) at baseline and Day 14, the mean test–retest reliability ICCs (95% CI) were ≥0.86 (0.79–0.91) for both buttocks. A clinically meaningful change was 1.0 point on the PR-PCSS and 1.0 on the CR-PCSS.

CONCLUSION The CR-PCSS and PR-PCSS reliably assess cellulite severity of the buttocks and can detect a clinically meaningful change after treatment for cellulite.

Supported by Endo Pharmaceuticals Inc., Malvern, PA. J.L. Cohen N.S. Sadik, and M.P. Goldman have served as consultants for Endo Pharmaceuticals Inc.

Cellulite, a condition that affects 80% to 98% of postpubertal women, is an alteration in the skin topography that results in a dimpled or undulated appearance of the affected skin, primarily located on the buttocks and thighs.[1–3] Although several cellulite grading scales are available, none have integrated both patient's and clinician's perspectives into scale development consistent with the US Food and Drug Administration (FDA) recommendations for developing measures suitable for use in clinical research trials.

Scales commonly used to assess cellulite severity include the Nürnberger–Müller scale (1978) and the

AboutSkin Dermatology and DermSurgery, PC, Greenwood Village, Colorado; †Dermatology, University of California, Irvine, California; ‡Department of Dermatology, Weill Cornell Medical College, Cornell University, New York, New York; §Endo Pharmaceuticals, Inc., Malvern, Pennsylvania; ‖Evidera, Bethesda, Maryland; ¶Department of Anesthesiology, University of North Carolina, Chapel Hill, North Carolina; #Cosmetic Laser Dermatology: A West Dermatology Company and Volunteer Clinical Professor of Dermatology, University of California, San Diego, California

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the full text and PDF versions of this article on the journal's Web site (www.dermatologicsurgery.org).

ISSN: 1076-0512 • Dermatol Surg 2020;46:1628–1635 • DOI: 10.1097/DSS.0000000000002756

Hexsel Cellulite Severity Scale (CSS; 2009).[4,5] The photonumeric scale (Hexsel CSS) was the first standardized and objective method of grading cellulite severity[3,5]; it added additional morphologic features to the Nürnberger–Müller scale (i.e., number and depth of depressions; aspect of raised areas; and degree of laxity, flaccidity, or sagging skin).[5] The Hexsel CSS is reproducible but has been complex and cumbersome to administer[6] and does not have a patient self-assessment component. Additional scales, such as the Global Aesthetic Improvement Scale (GAIS),[7] although not designed specifically to assess cellulite,[7] have been adapted for use in assessing changes in cellulite severity in clinical trials,[8,9] by both clinicians (Investigator-GAIS [I-GAIS]) and patients (Subject-GAIS [S-GAIS]).[10]

Given the need for a simple, easily understood cellulite severity–specific grading scale for both clinicians and patients, the Clinician Reported Photonumeric Cellulite Severity Scale (CR-PCSS) and Patient Reported PCSS (PR-PCSS) were developed. These 5-point photonumeric severity scales were designed to assess cellulite severity on the buttocks or thighs. In a phase 2 collagenase clostridium histolyticum trial (ClinicalTrials.gov identifier: NCT02724644) that used the CR-PCSS to determine cellulite severity, the CR-PCSS score at Day 1 (screening) significantly correlated with the Hexsel CSS total score.[11] Aligned with guidance from the FDA, new CSS measurement properties to be assessed include reliability, validity, and ability to detect a change.[12]

The objective of this article is to report on a series of noninterventional studies that determined the reliability, validity, and ability to detect buttock cellulite severity changes using the CR-PCSS and PR-PCSS.

Methods

Clinician Reported Photonumeric Cellulite Severity Scale—Buttocks

Concept Elicitation and Cognitive Interviews

Photographs selected for use in the CR-PCSS and PR-PCSS were either from cellulite studies conducted by the sponsor or from separate photograph shoots supported by the sponsor. The 5-point severity scale labels were chosen based on review of previous aesthetic scales[13,14] and iterative feedback from the FDA. Descriptors were developed for each label to help ensure reliable distinctions between ratings levels. Twelve board-certified aesthetic clinicians ($n = 10$ dermatologists; $n = 2$ cosmetic surgeons) from geographically diverse US locations with substantial clinical practice experience in aesthetic medicine participated in qualitative interviews to provide guidance to scale developers in the terminology used to define cellulite and the assessment of cellulite severity. The clinicians also provided feedback on draft scales containing photographs with varying levels of cellulite severity and corresponding labels and text descriptors (CR-PCSS). Based on feedback from 12 interviews, a set of scale labels and descriptors was developed.

Clinician Reported Photonumeric Cellulite Severity Scale Intrarater and Interrater Reliability

Scale Reliability Using Photographic Images. Board-certified clinicians ($n = 5$) with ≥10 years clinical practice experience in aesthetic medicine were recruited for the study. Clinician training was achieved through video recording on use of the CR-PCSS, and a qualification assessment was performed to demonstrate proficiency in its use. CR-PCSS test–retesting involved assessments at Days 1 and 14 of digital images of women with cellulite who had participated in other clinical trials (ClinicalTrials.gov identifiers: NCT01518907 and NCT01987986). On Day 1, each clinician rated the cellulite severity of buttocks for 100 randomized digital photographs. On Day 14, the clinicians were retrained in use of the CR-PCSS and subsequently rated the same 100 digital photographs in a different random order. Intrarater and interrater reliability were determined using intraclass correlation coefficients (ICCs) and 95% confidence intervals (CIs) at Days 1 and 14. Two reliability estimates were used, both based on a 2-way analysis of variance (ANOVA) as follows: the ICC for consistency was based on consistent ordering of photographs, whereas the agreement estimate was based on exact agreement among clinicians.[15] Sample size determinations were based on unpublished data from a previous CR-PCSS interrater and test–retest reliability study of 3

clinicians. If the true reliability was 0.80, 5 clinicians rating 100 photographs at 2 time points would ensure that an observed ICC of 0.80 would exceed a 0.70 reliability criterion with 95% confidence (i.e., a lower bound of 95%
CI >0.70).

Scale Reliability Using Live Patient Assessments. The CR-PCSS was intended to be used clinically, thus it was necessary to validate the CR-PCSS using in-person (i.e., live) assessments. Women aged ≥18 years were recruited for participation and self-rated their overall cellulite severity as none, almost none, mild, moderate, or severe during a prescreen phone call. Women were screened by a clinician with >10 years of experience rating cellulite severity, and this individual was independent and separate from clinicians involved in reliability assessments. Women with inflammation; an active infection; tattoo(s) located within 2 cm of the areas to be evaluated; current treatment of cellulite; or recent use (within the previous 12 months) of injectables, laser treatment, liposuction, radiofrequency treatment, or surgery for cellulite in the areas to be evaluated were excluded.

Each of 2 areas (i.e., left and right buttock) was assessed individually, and 1 area was defined as the "screening area." This screening area was used to classify women to reach a target enrollment at each cellulite area of the 5 severity levels. A total of 6 board-certified clinicians with ≥4 years of clinical practice experience in core aesthetic medicine specialties and naive to the CR-PCSS were trained on the use of CR-PCSS by video and in-person instruction at baseline, using live models. Clinicians were blinded to the screening assessments. CR-PCSS test–retesting involved assessments at baseline and Day 2. Each clinician rated the left and right buttock of each woman. Various methods were used to minimize clinician recall bias, including rating a large number of women, changing the order of evaluations on Day 2, having the visual field limited only to areas under evaluation, and having no physical contact or verbal interactions with participants. Intrarater and interrater reliability were determined by calculating the ICC and 95% CI at baseline and Day 2 using the same statistical methodology[15] as in the photograph reliability study. If the true ICC was 0.80, 5 or 6 clinicians rating 80 patients at 2 time points would ensure that an observed ICC of 0.80 would exceed a 0.70 reliability criterion with 95% confidence (i.e., a lower bound of 95% CI >0.70), assuming all model assumptions were met (e.g., no rater or time main effects).

Ability to Detect Change

A clinically meaningful change was estimated using anchor-based methods (anchoring change to clinically understood and validated assessments) as recommended by the FDA guidance on patient-related outcomes.[12] For the CR-PCSS, the anchor was the I-GAIS. The ability of the CR-PCSS to detect a change was based on pooled data from 2 identically designed, phase 3, randomized, double-blind, placebo-controlled trials of collagenase clostridium histolyticum-aaes (CCH; QWO™, Endo Aesthetics LLC, Malvern, PA) for the treatment of cellulite of the buttocks (ClinicalTrials.gov identifiers: NCT03428750 and NCT03446781). Outcomes using the CR-PCSS (dependent variable) and the I-GAIS (independent variable) were analyzed with a Kruskal–Wallis 1-way ANOVA model to determine the clinically meaningful change thresholds. Statistical significance (e.g., H value, *p*-value) was determined by pairwise 2-sided multiple comparison analysis (Dwass-Steel-Critchlow-Fligner method). Spearman correlations were used to assess the association between changes in the CR-PCSS and I-GAIS.

Patient Reported Photonumeric Cellulite Severity Scale—Buttocks

Concept Elicitation and Cognitive Interviews

Twenty-four women (aged 23–55 years) with moderate to severe cellulite (thigh/buttocks) from 3 geographically diverse US locations participated in concept elicitation interviews to provide information on their experience with cellulite and how they defined cellulite severity. During the interviews, the women were also asked to rank-order by cellulite severity 5 photographs developed for the PR-PCSS

and provide spontaneous descriptors of the photographs. Participants then provided feedback on draft labels and descriptors. An additional 23 women (aged 26–55 years) with cellulite participated in cognitive interviews (qualitative step in a scale development process using verbal probing or thinking-out-loud methodology) and were asked, in the absence of photographic cues or labels, to rank-order descriptors that had been revised based on concept elicitation interviews. A second round of cognitive interviews was performed with 11 women (aged 21–60 years) with cellulite to confirm content validity of revised descriptors.

Patient Reported Photonumeric Cellulite Severity Scale Test–Retest Reliability

Scale Reliability Using Photographic Images. Adult women ($n = 99$) with varying degrees of cellulite from 6 US locations were recruited for the study. On Day 1, digital photographs were taken of each woman's left and right buttock. Women were instructed in the use of the PR-PCSS and also provided written instructions. Women self-rated the cellulite severity of their left and right buttock images displayed on a computer. On Day 14, women were retrained on use of the PR-PCSS and subsequently self-rated the Day 1 digital images in a different random order. Test–retest reliability was determined using ICCs and 95% CIs, using the same statistical methodology as studies described above. To increase the opportunity for recruitment of less common cellulite severity levels (e.g., none and severe) and to account for attrition between visits, it was estimated that 84 enrolled participants would ensure an observed ICC of 0.80.

Ability to Detect a Change

As noted above, a clinically meaningful change was estimated based on FDA guidance.[12] The S-GAIS was the anchor for the PR-PCSS. As with the CR-PCSS, the ability of the PR-PCSS to detect a change in cellulite severity was based on pooled data from the 2 identically designed CCH trials with outcomes using the PR-PCSS (dependent variable) and the S-GAIS (independent variable) analyzed using similar statistical methodology. Correlation between changes in the PR-PCSS and S-GAIS was determined by Spearman correlation.

Figure 1. Patient Reported Photonumeric Cellulite Severity Scale and Clinician Reported Photonumeric Cellulite Severity Scale for the buttocks.

Results

Clinician Reported Photonumeric Cellulite Severity Scale

Concept Elicitation and Cognitive Interviews
Cognitive interviews with 10 dermatologists and 2 plastic surgeons (75.0% men) with a mean ± SD of 23.0 ± 10.7 years in clinical practice were conducted to establish content validity for the CR-PCSS by adapting and refining the labels and text descriptors (Figure 1); this was conducted to make the descriptors broadly relevant to cellulite and to distinguish between categories without being overly complex.

Interrater and Intrarater Reliability

Scale Reliability Using Photographic Images
CR-PCSS intrarater reliability was calculated at Days 1 and 14 with 5 male clinicians (mean age ± SD, 57.6 ± 7.6 years) with an average of 26 ± 10.0 years of clinical practice. At Days 1 and 14, individual clinician ICC point estimates for intrarater reliability were >0.80. Point estimates for ICC (95% CI) for consistency ranged from 0.84 (0.77–0.89) to 0.90 (0.86–0.93). Point estimates for ICC (95% CI) for agreement ranged from 0.80 (0.63–0.89) to 0.89 (0.83–0.93). There was 1 clinician for whom the CI lower limit for the ICC for agreement was <0.70. Overall CR-PCSS interrater photographic reliability for the 5 clinicians was calculated on Days 1 and 14 for the 100 digital photographs of varying levels of cellulite severity of the buttocks. The lower limits for the 95% CI of the mean ICC point estimates were >0.70 for both days. At Day 1, the mean ICC (95% CI) point estimates were 0.86 (0.81–0.89) and 0.84 (0.78–0.88) for consistency and agreement, respectively. At Day 14, the mean ICC (95% CI) point estimates were 0.84 (0.80–0.88) and 0.83 (0.78–0.88) for consistency and agreement, respectively.

Scale Reliability Using Live Patient Assessments
Six clinicians (3 plastic surgeons; 3 dermatologists [5 men; 1 woman]) were included in the in-person assessments (18.0 median years in practice, post-residency). Ninety-three women were screened, 81 were enrolled, and 76 (mean age ± SD, 45.1 ± 15.5 years) participated in both CR-PCSS assessments. CR-PCSS intrarater reliability for the 6 clinicians was calculated between baseline and Day 2 (See **Supplemental Digital Content**, Table S1, http://links.lww.com/DSS/A489); overall mean ICC point estimates were within an acceptable range with a mean of ≥0.81 for both the left and right buttock. When assessed at baseline and Day 2, the CR-PCSS mean interrater reliability scores among the 6 clinicians were within the acceptable range at ≥0.70 for each buttock for both assessments (See **Supplemental Digital Content**, Table S1, http://links.lww.com/DSS/A489).

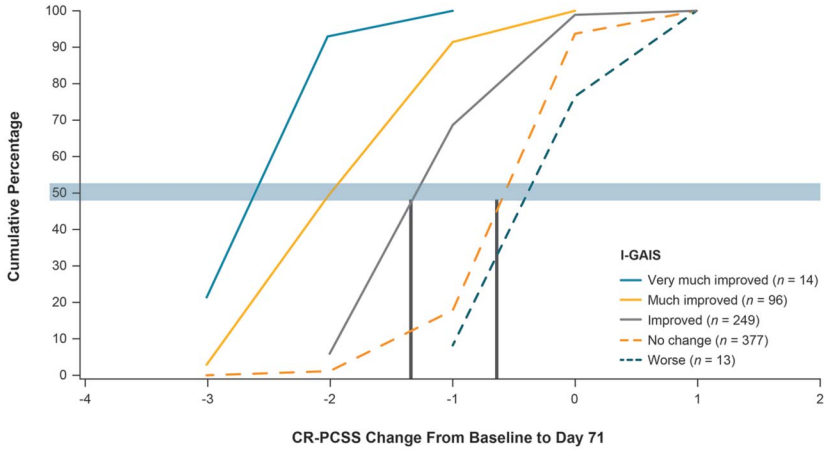

Figure 2. Cumulative distribution function for a CR-PCSS change from baseline to Day 71 in buttocks, by I-GAIS category. CR-PCSS, Clinician Reported Photonumeric Cellulite Severity Scale; I-GAIS, Investigator–Global Aesthetic Improvement Scale.

Ability to Detect a Change

Pooled data from 2 identically designed clinical trials of 793 women with cellulite (mean age ± SD, 46.7 ± 10.6 years; range 18–78 years) treated with CCH were analyzed to determine the threshold for a meaningful change in cellulite severity. Most of the 793 women were White (n = 620 [78.2%]) and predominantly Fitzpatrick skin Types I-III (n = 473 [59.6%]; Types IV-VI, n = 320 [40.4%]). Based on a 1-point and 2-point change in the I-GAIS for the target buttock, the mean ± SD threshold of significant change for the CR-PCSS was 0.75 points ±0.58 (p < .0001) and 1.45 points ±0.69 (p < .0001), respectively (Figure 2). Effect sizes were −1.55 and −3.11 for a 1-point and 2-point change, respectively. Correlation between the change from baseline to Day 71 in CR-PCSS and I-GAIS was −0.65 (p < .001).

Patient Reported Photonumeric Cellulite Severity Scale

Concept Elicitation and Cognitive Interviews

Concept elicitation and cognitive interviews established content validity of the PR-PCSS. For the PR-PCSS, 47 women (White [89.4%]; mean age ± SD, 40.0 ± 9.0 years) with cellulite were interviewed over several stages to adapt and refine the labels and descriptors from the CR-PCSS to create the PR-PCSS (Figure 1). Content validity of the PR-PCSS was established by correct ordering of photographs during the concept elicitation phase and correct ordering of descriptors during the cognitive interviews.

Test–Retest Reliability

Scale Reliability Using Photographic Images

For the PR-PCSS test–retest reliability assessment, data from 99 women (mean age ± SD, 40.6 ± 12.9 years; range 19–72 years) with cellulite were included. The PR-PCSS demonstrated good reliability with the same point estimates for test–retest reliability for agreement and consistency within each quadrant ranging from 0.86 (right buttock) to 0.87 (left buttock) and lower bounds of the CI all ≥0.79 (See **Supplemental Digital Content**, Table S2, http://links.lww.com/DSS/A490).

Ability to Detect a Change

Based on a 1-point and 2-point improvement in the S-GAIS for the target buttock, the mean threshold ± SD of significant change for the PR-PCSS was 0.94 points ±0.76 (p < .0001) and 1.45 points ±0.81 (p < .0001), respectively (Figure 3). Effect sizes were −1.89 and −2.89 for a 1-point and 2-point change, respectively. Correlation between the change from baseline to Day 71 in the PR-PCSS and S-GAIS was −0.58 (p < .001).

Discussion

The CR-PCSS and PR-PCSS were developed to provide clinicians and patients with simple, easily understood tools to rate cellulite severity on the buttocks during clinical research and routine clinical practice. Both the CR-PCSS and PR-PCSS demonstrated good test–retest reliability assessing cellulite severity of the buttocks using

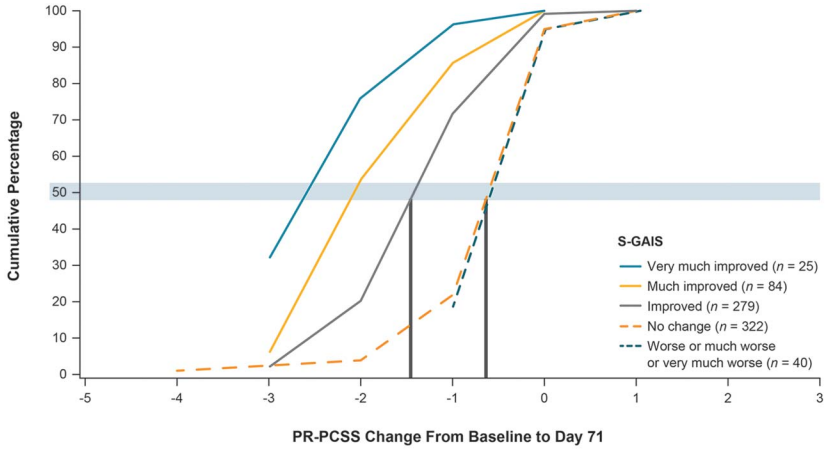

Figure 3. Cumulative distribution function for a PR-PCSS change from baseline to Day 71 in buttocks, by S-GAIS category.

digital photographs. The CR-PCSS also demonstrated good test–retest and interrater reliability using live patient assessments. Interrater reliability was not performed for the PR-PCSS because women did not evaluate cellulite severity of other women. Importantly, clinically meaningful change thresholds for cellulite severity were established for the CR-PCSS and PR-PCSS.

Tissue stabilized–guided subcision system[16] cellulite studies included a 4-point photonumeric scale in which clinicians scored photographs for the number and depth of depressions.[8] Intraclass correlation coefficient for intrarater reliability was 78%; interrater reliability was ≥0.81 at 3 and 12 months. Correlations between clinician reported and patient reported outcomes were not reported. Cellulite studies using a 1440-nm neodymium-doped yttrium aluminum garnet laser with a side-firing fiber and temperature-sensing cannula included a 5-point photonumeric scale with which clinicians scored the number of evident dimples and severity of linear undulations using photographs.[17] Interrater kappa values ranged from 0.69 to 0.90, and intrarater kappa values ranged from 0.75 to 0.92.[17] In this study, the CR-PCSS using live assessment and the PR-PCSS using photographs produced acceptably comparable ratings of cellulite severity. As well, good agreement between CR-PCSS and PR-PCSS ratings has been presented.[18] These data compare favorably versus data from other clinical trials that have assessed concordance between other validated patient-rated and clinician-rated aesthetic scales.[19] For example, the correlation coefficient between patient-rated and clinician-rated outcomes was 0.28 in a 2014 trial in patients with submental fat.[20]

No firm cutoff for an ICC has been universally established. Other health-related quality of life instruments have defined the threshold for acceptable test–retest reliability as an ICC of >0.60 or ≥0.70.[21–23] In contrast to defining a threshold for acceptable reliability, other publications have suggested grades of reliability based on the ICC. In a study by Cicchetti and colleagues, the authors recommended a scale range as follows: <0.40 (poor), 0.40 to 0.59 (fair), 0.60 to 0.74 (good), and ≥0.75 (excellent),[24] with slightly different cutoffs recommended by Koo and colleagues as follows: <0.5 (poor), 0.5 to 0.75 (moderate), 0.75 to 0.9 (good), and >0.9 (excellent).[25] In this study, the threshold for acceptable reliability was considered an ICC ≥0.70, with the 95% CI lower boundary ≥0.62, which is consistent with other published criteria.[23]

Unlike some other cellulite severity scales,[8,17,26] the CR-PCSS was validated using live patient assessments. The CR-PCSS and PR-PCSS were shown to be sensitive to clinically meaningful changes in cellulite severity. Effect sizes for 1-level and 2-level improvement in the PR-PCSS were large at −1.89 and −2.89, respectively. The data suggest a PR-PCSS rating change score of 1.0 from baseline to Day 71 is indicative of meaningful change for patients because this level of change was associated with ratings of meaningful improvement and satisfaction on external anchor variables (S-GAIS) and also was associated with an effect size >1, which is consistent with a large effect. Similar results were observed for the CR-PCSS (based on I-GAIS). Thus overall, a 1-level change in cellulite severity was found to be clinically meaningful by both patients (PR-PCSS) and clinicians (CR-PCSS).

Limitations of these analyses include the restricted range of skin tone (60% of women were Fitzpatrick skin Types I–III) among the women included in the studies, and the potential inherent risk of subjectivity and emotional response to self-evaluation with the PR-PCSS. In addition, clinician training could have been more extensive. In conclusion, the CR-PCSS and PR-PCSS are valid and reliable outcome measures for determining cellulite severity and should be considered for evaluating cellulite severity in clinical trials and real-world settings.

Acknowledgments Technical editorial and medical writing assistance was provided, under the direction of the authors, by Mary Beth Moncrief, PhD, and Julie B. Stimmel, PhD, Synchrony Medical Communications, LLC, West Chester, PA. Funding for this support was provided by Endo Pharmaceuticals Inc., Malvern, PA.

References

1. Hexsel D, Mazzuco R. Cellulite. In: Tosti A, Hexsel D, editors. *Update in Cosmetic Dermatology*. Berlin, Germany: Springer-Verlag; 2013; pp. 21–32.
2. Khan MH, Victor F, Rao B, Sadick NS. Treatment of cellulite: part I. Pathophysiology. J Am Acad Dermatol 2010;62:361–70; quiz 71–72.

3. Friedmann DP, Vick GL, Mishra V. Cellulite: a review with a focus on subcision. Clin Cosmet Investig Dermatol 2017;10:17–23.

4. Nürnberger F, Müller G. So-called cellulite: an invented disease. J Dermatol Surg Oncol 1978;4:221–9.

5. Hexsel DM, Dal'Forno T, Hexsel CL. A validated photonumeric cellulite severity scale. J Eur Acad Dermatol Venereol 2009;23:523–8.

6. Callaghan DJ, Robinson DM, Kaminer MS. Cellulite: a review of pathogenesis-directed therapy. Semin Cutan Med Surg 2017;36: 179–84.

7. Narins RS, Brandt F, Leyden J, Lorenc ZP, et al. A randomized, double-blind, multicenter comparison of the efficacy and tolerability of Restylane versus Zyplast for the correction of nasolabial folds. Dermatol Surg 2003;29:588–95.

8. Kaminer MS, Coleman WP III, Weiss RA, Robinson DM, et al. Multicenter pivotal study of vacuum-assisted precise tissue release for the treatment of cellulite. Dermatol Surg 2015;41:336–47.

9. Petti C, Stoneburner J, McLaughlin L. Laser cellulite treatment and laser-assisted lipoplasty of the thighs and buttocks: combined modalities for single stage contouring of the lower body. Lasers Surg Med 2016;48:14–22.

10. Sadick NS, Goldman MP, Liu G, Shusterman NH, et al. Collagenase clostridium histolyticum for the treatment of edematous fibrosclerotic panniculopathy (cellulite): a randomized trial. Dermatol Surg 2019;45: 1047–56.

11. Young VL, Sadick N, Liu G, Shusterman N, et al. Comparisons of clinician-reported and patient-reported cellulite severity scales with existing scales for measurement of cellulite severity. Presented at the Northeastern Society of Plastic Surgeons 34th Annual Meeting; September 8–10, 2017; Newport, RI.

12. US Department of Health and Human Services, Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available from: https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed April 23, 2019.

13. Pascoe VL, Enamandram M, Corey KC, Cheng CE, et al. Using the Physician Global Assessment in a clinical setting to measure and track patient outcomes. JAMA Dermatol 2015;151:375–81.

14. U.S. Department of Health and Human Services, Food and Drug Administration. *Acne Vulgaris: Establishing Effectiveness of Drugs Intended for Treatment Guidance for Industry*. 2018. Available from: https://www.fda.gov/media/71152/download. Accessed September 26, 2019.

15. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30–46.

16. Green JB, Cohen JL. Cellfina observations: pearls and pitfalls. Semin Cutan Med Surg 2015;34:144–6.

17. DiBernardo B, Sasaki G, Katz BE, Hunstad JP, et al. A multicenter study for a single, three-step laser treatment for cellulite using a 1440-nm Nd:YAG laser, a novel side-firing fiber, and a temperature-sensing cannula. Aesthet Surg J 2013;33:576–84.

18. Kirby MT, Lenderking WR, Bender RH, Chen J, et al. Assessing cellulite severity: test-retest reliability of and concordance between new clinician reported and patient reported photonumeric scales. Value Health 2018;21:S229.

19. Dayan SH, Schlessinger J, Beer K, Donofrio LM, et al. Efficacy and safety of ATX-101 by treatment session: pooled analysis of data from the phase 3 REFINE trials. Aesthet Surg J 2018;38: 998–1010.

20. McDiarmid J, Ruiz JB, Lee D, Lippert S, et al. Results from a pooled analysis of two European, randomized, placebo-controlled, phase 3 studies of ATX-101 for the pharmacologic reduction of excess submental fat. Aesthet Plast Surg 2014;38:849–60.

21. Wiklund I, Raluy-Callado M, Chen WH, Muenzer J, et al. The Hunter Syndrome-Functional Outcomes for Clinical Understanding Scale (HS-FOCUS) questionnaire: item reduction and further validation. Qual Life Res 2014;23:2457–62.

22. Leidy NK, Revicki DA, Geneste B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. Value Health 1999;2:113–27.

23. Hays R, Anderson R, Revicki DA. Assessing reliablity and validity of measurement in clinical trials. In: Staguet MJ, Hays RD, Fayers PM, editors. *Quality of Life Assessment in Clinical Trials*. New York, NY: Oxford University Press; 1998.

24. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic 1981;86:127–37.

25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15: 155–63.

26. Hexsel D, Fabi SG, Sattler G, Bartsch R, et al. Validated assessment scales for cellulite dimples on the buttocks and thighs in female patients. Dermatol Surg 2019;45(Suppl 1):S2–S11.

Address correspondence and reprint requests to: Joel L. Cohen, MD, AboutSkin Dermatology and DermSurgery, PC, 5340 South Quebec Street, Suite 300, Greenwood Village, CO 80111, or e-mail: jcohenderm@yahoo.com