Reliability and Validity of a Photographic Method for Measuring Facial Hair Density in Men

Many studies have investigated hair removal or growth prevention treatments, but they often measure hair density using noninvasive methods that are subjective and qualitative.[1] Although photographic and digital hair-counting methods have been used, their reliability and validity remain unknown.[2] We describe a simple, noninvasive method of hair counting used in a hair growth prevention treatment trial and assess its reliability and validity.

Methods. The data are from the first 14 healthy men consecutively enrolled in a randomized, double-blinded, placebo-controlled trial of a topical agent for hair growth prevention. Eligible subjects were required to shave at least once daily to avoid a beard with hair length visible above the skin line and to have a baseline physician global assessment (PGA) score for hair density of 4 or 5 in the beard area. The PGA was developed by us for the larger clinical trial as a visual analog scale for rating hair density by overall impression (**Figure**).

Subjects were randomized as to which side of their face would receive drug or placebo, which was then applied once daily after shaving to a treatment area within the beard region in a split-face design (Figure). The duration of active treatment was 6 or 8 weeks; subjects were assessed every 2 or 4 weeks for up to 8 to 16 weeks. Subjects did not shave for 48 hours prior to each visit so that they would have enough visible hair for assessment. At each visit, the PGA and digital photography of the treatment areas were performed (Figure). The study was approved by the University of Pennsylvania institutional review board.

Two of us (J.W. and J.M.S.) independently counted hairs in all photographs to assess interrater reliability (Figure). Five months after the initial measurement, hairs were recounted in all photographs to assess test-retest reliability. We used the intraclass correlation coefficient (ICC) and Spearman ρ correlation to assess reliability. Construct validity was evaluated by comparing hair counts with respect to corresponding PGA ratings using the *t* test. We conservatively estimated a sample size of 100 photographs with 85% power to detect an ICC of 0.6, assuming null ICC of 0.4 and $\alpha = 0.05$.

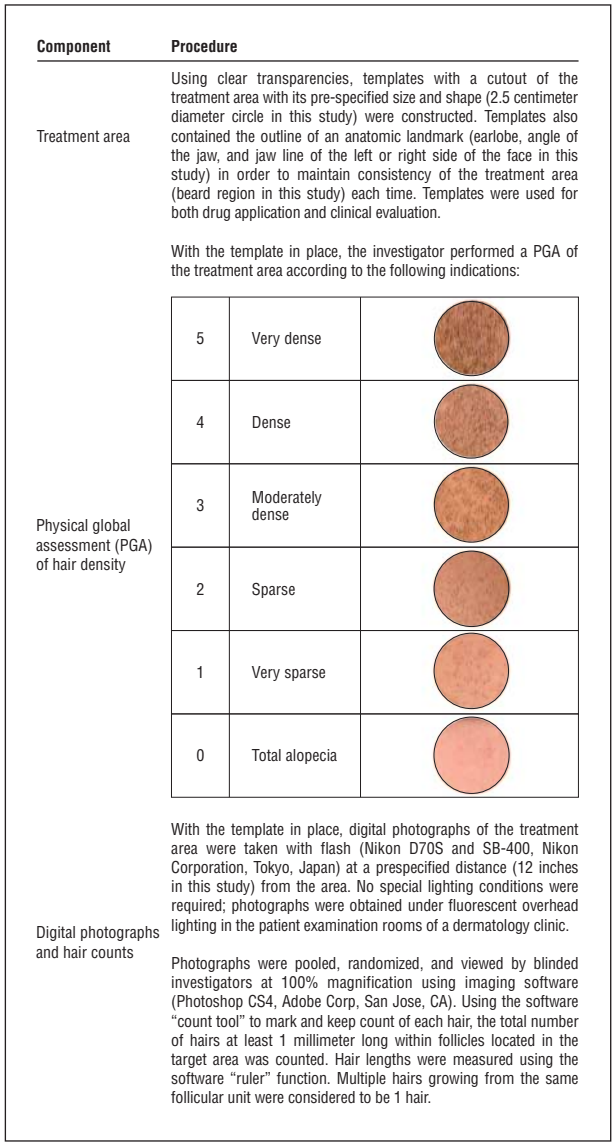

Figure. Protocol for hair density assessments. The image herein is copyrighted by the University of Pennsylvania, Philadelphia, and is reproduced with permission.

Results. The median age of the subjects was 28 years (interquartile range [IQR], 26-38 years). Eleven subjects were white (79%), and 3 were Asian (21%). All subjects had brown or black hair. A total of 130 photographs were obtained. Hair counts were approximately normally distributed, ranging from 2 to 391. The subject PGA scores were available for 114 photographs and ranged from 2 to 5 (median, 4; IQR, 4-4). Test-retest reliability demonstrated an ICC of 0.90 (95% confidence interval [CI], 0.86-0.93) and a Spearman ρ of 0.88 (95% CI, 0.84-0.92). Interrater reliability demonstrated an ICC of 0.81 (95% CI, 0.74-0.86) and a Spearman ρ of 0.81 (95% CI, 0.75-0.87). In the validity analysis, we included only PGA scores for which there were at least 10 corresponding photographs. Photographs with a PGA score of 3 had a lower mean hair count (mean [SD] count, n = 195.0 [16.5]) than those with PGA score of 4 (mean [SD] count, n = 237.2 [5.8]) ($P = .003$).

Comment. Our hair counting method demonstrates excellent interrater and intrarater reliability as well as construct validity based on its ability to discriminate categories of a PGA.[3] In contrast to other methods, our approach does not require expensive or specialized equipment. It provides better quantification of hair changes than global assessment scales, which may be too qualitative for clinical trials.[1] Moreover, it is less tedious and labor intensive than the manual collection, counting, and weighing of hair.[4] Although automated methods such as the TrichoScan (TRICHOLOG GmbH, Freiburg, Germany) have reported high reliability, fully automated approaches are hindered by imperfect algorithms, which can lead to inaccuracy.[1,5]

We recognize several limitations. First, hair diameter and length were not evaluated. Second, the camera was not mounted, and the skin in the treatment areas was not marked so as to guarantee the same exact evaluation distance and site every time. The generalizability of our results to areas with different hair density or to people with darker skin is unknown. Finally, additional studies are required to determine if this technique is responsive to true changes in hair density and to compare this method to other approaches such as digital photodermoscopy. Nevertheless, our simple, noninvasive method of hair counting demonstrates excellent reliability and discrimination validity and deserves further evaluation as an assessment tool for hair removal or growth prevention studies.

Joy Wan, BA
Katrina Abuabara, MD, MA
Shanu K. Kurd, MD, MSCE
Amy Musiek, MD
Jane M. Steinemann, BA
Carmela C. Vittorio, MD
Joel M. Gelfand, MD, MSCE

Accepted for Publication: July 10, 2011.
Author Affiliations: Departments of Dermatology (Mss Wan and Steinemann and Drs Abuabara, Kurd, Musiek, Vittorio, and Gelfand) and Internal Medicine (Drs Abuabara and Kurd) and Center for Clinical Epidemiology and Biostatistics (Dr Gelfand), University of Pennsylvania Perelman School of Medicine, Philadelphia; and Philadelphia VA Medical Center, Philadelphia (Dr Musiek). Dr Musiek is now with the Department of Dermatology, Washington University, St Louis, Missouri; Ms Steinemann is now with Drexel University College of Medicine, Philadelphia.
Correspondence: Dr Gelfand, 1471 Penn Tower, One Convention Avenue, Philadelphia, PA 19104 (Joel.Gelfand@uphs.upenn.edu).
Author Contributions: Ms Wan and Dr Gelfand had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Abuabara, Kurd, Vittorio, and Gelfand. *Acquisition of data:* Wan, Abuabara, Musiek, Steinemann, Gelfand. *Analysis and interpretation of data:* Wan and Gelfand. *Drafting of the manuscript:* Wan. *Critical revision of the manuscript for important intellectual content:* Wan, Abuabara, Kurd, Musiek, Steinemann, Vittorio, and Gelfand. *Statistical analysis:* Wan and Gelfand. *Obtained funding:* Vittorio and Gelfand. *Administrative, technical, and material support:* Wan, Abuabara, and Kurd. *Study supervision:* Musiek, Vittorio, and Gelfand.
Financial Disclosure: Dr Vittorio has filed a patent application for the use of DNA polymerase inhibitors in inducing alopecia. Dr Gelfand served as consultant and investigator with Abbott, Amgen, Centocor, Genentech, Novartis, and Pfizer; consultant with Celgene, Covance, Galderma, Shire Pharmaceuticals, and Wyeth; and investigator with Shionogi.
Funding/Support: This study was supported in part by grants from the Department of Dermatology at the University of Pennsylvania (Dr Vittorio), the Edwin and Fannie Gray Hall Center for Human Appearance at the University of Pennsylvania (Dr Vittorio), National Institutes of Health training grant T32-AR07465 (Ms Wan and Dr Musiek), and the Doris Duke Clinical Research Fellowship (Dr Abuabara).
Role of the Sponsors: The sponsors had no role in the design and conduct of the study; in the collection, analysis, and interpretation of data; or in the preparation, review, or approval of the manuscript.
Additional Contributions: Jennifer Goldfarb, RN, Albana Oktrova, and Debbie Leahy, LPN, did an outstanding job coordinating the clinical trial associated with this study.

1. Chamberlain AJ, Dawber RP. Methods of evaluating hair growth. *Australas J Dermatol*. 2003;44(1):10-18.
2. Hamzavi I, Tan E, Shapiro J, Lui H. A randomized bilateral vehicle-controlled study of eflornithine cream combined with laser treatment versus laser treatment alone for facial hirsutism in women. *J Am Acad Dermatol*. 2007; 57(1):54-59.
3. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.
4. Price VH, Menefee E, Strauss PC. Changes in hair weight and hair count in men with androgenetic alopecia, after application of 5% and 2% topical minoxidil, placebo, or no treatment. *J Am Acad Dermatol*. 1999;41(5, pt 1):717-721.
5. Van Neste D, Trüeb RM. Critical study of hair growth analysis with computer-assisted methods. *J Eur Acad Dermatol Venereol*. 2006;20(5):578-583.

VIGNETTES

Connubial Androgenetic Alopecia

Report of a Case. We report herein the case of a 52-year-old, postmenopausal Hispanic woman who developed severe androgenetic alopecia following involuntary exposure to topical testosterone gel used by her spouse for the treatment of hypogonadism. The patient presented to our clinic with complaints of a 1-year history of hair loss. Results of the pull test were negative; physical examination revealed severe hair thinning involving the crown and the frontotemporal regions. Dermoscopy of the scalp revealed more than 20% hair diameter variation. No other signs of hyperandrogenism were identified, such as hirsutism, acne, or obesity. Diagnosis of androgenetic alopecia with Hamilton pattern was established based on the clinical and dermoscopic findings.

Owing to the abrupt onset and Hamilton pattern, a laboratory workup was performed to exclude endocrine abnormalities. The results revealed high levels of testosterone (146 ng/dL; normal. 2-45 ng/dL; to convert testosterone to nanomoles per liter, multiply by 0.0347) and free testosterone (22.7 pg/mL; normal, 0.2-5.0 pg/mL). To exclude ovarian malignant neoplasm, abdominal and transvaginal ultrasonograms were obtained. The findings were normal.

Further investigation into the patient's history revealed that her spouse had started using a topical testosterone gel (containing 1% testosterone) 18 months previously for the treatment of hypogonadism. He had been applying the topical testosterone once daily (5-g packet) to his upper arm.