

New method to assess soft focus parameters on living skin: How does it compare to visual lay person rating of skin radiance?

Nathalie Lunau, Stephan Bielfeldt, Matthias Seise, Klaus-P. Wilhelm
proDERM Institute for Applied Dermatological Research, 22869 Schenefeld, Germany

Introduction

The purpose of this work was to develop a new method to measure soft focus properties of cosmetic products on living forearm skin as a screening method to select a candidate for a larger claim support study with the parameters skin radiance, attractiveness and skin evenness rated by lay persons and trained raters on standardized facial photographs. The results of the claim support study were expected to prove the success of the pre selection method.

Material & Methods

Screening Study on Volar Forearm

It was assumed that the soft focus on forearm skin (Figure 1 + 2) can be used to screen and select the best candidate for a larger study with facial application and assessment of facial radiance. We chose real in vivo skin on the arm, because it was assumed that it simulates facial skin better than any artificial material. However, the forearm partly lacks the inhomogeneity of the facial skin and contains almost no sebum. Six cosmetic products containing ingredients declared to increase facial radiance were applied to different test areas on the outer forearms of six female subjects. Then the baseline assessments of soft focus properties were performed. A defined amount of test product was applied on each marked test area. After product application, assessments of soft focus properties were repeated. Due to the fact that gloss intensity on forearms is low compared to facial skin, in a second trial artificial sebum was applied before product application. The amount of superficial sebum on the

skin was in the range from 50 to 200 $\mu\text{g}/\text{cm}^2$ as measured by Sebumeter. Then the products were applied and soft focus parameters were assessed (The surprising results of the soft focus screening are shown in Figure 3).

Rating of radiance, attractiveness and even skin tone on facial photographs

Since the selection of a best candidate based on the soft focus properties on forearm failed, a stepwise procedure with expert rating of pairs of facial photographs was performed. Only one stimulus person per test product was photographed before and after application of the respective test product. Three expert raters selected the product with the best result in enhancement of radiance and even skin tone. Based on this rating (data not shown) two products could not be discriminated. In a second run the experiment was repeated on 6 stimulus persons and only the two remaining products were tested. The three experts agreed in one of the two products to perform best (data not shown).

The pre-selected cosmetic test product was evaluated for facial radiance, attractiveness, and evenness of the skin on standardized facial photographs (Figure 4) by a panel of 27 female stimulus persons. It was predicted that the use of lipstick would have no influence on the rating results, because due to the short time between photographs untreated and treated they would appear identical in the two images. By 11 subjects no lipstick was used, while a group of 16 subjects was photographed with lipstick. On a separate analysis of the two groups was performed to proof this assumption.

One of four different nuances of the test product was chosen for each subject according to a good fit of the lightness of color to the skin type. After taking baseline photos, the test product was applied on the face of the stimulus persons by an experienced technician. Photographs were repeated after a two hours waiting time.

Judgment of Facial Photographs by Lay and Trained Raters

A group of lay raters (N=24) and a group of trained raters (N=6) assessed skin radiance, attractiveness, and evenness on facial photographs of the 27 female subjects presented in pairs (untreated versus treated) on color calibrated monitors. Both rater groups ranked the pictures with either „1“ (prefer) or „0“ (not prefer) for each image per pair separately in three runs for the parameters radiance, attractiveness and even skin tone.

Statistics

Mean values, standard deviations and 95% confidence limits were calculated from the raters' judgments concerning radiance, attractiveness, and evenness by use of the statistic software SAS (Figure 5).

Conclusion

The evaluation of the goniometric screening method for a selection of facial products concerning claims as improved radiance, attractiveness and even skin tone was not successful, though it was performed on living forearm skin and not artificial targets. The reason might be that the skin on the forearm is too even and homogeneous. Only one aspect, the gloss reducing efficacy (achievement of a mat skin surface) could be assessed. However, forearm skin is already extremely mat and we had to apply artificial sebum to increase glossiness. We expected that the test products would reduce glossiness on the sebum treated skin with different performance, so that the results could be used to discriminate between products and select candidates for a larger study on facial radiance. The results did not meet with our expectations. The finally selected well performing test product surprisingly increased glossiness on the sebum treated forearm (Figure 3). It seems that the reduction of gloss is not of main importance for the performance of the test products. The covering pattern of the reflected light, which are not covered in the forearm experiments, might be of higher importance. Our selection of product candidates based on facial images was successful, though only a few stimulus persons had to be photographed. This method is subjective and relies on the quality of the expert raters, but as can be seen from the very homogeneous ratings of trained and laypersons, even the human eye of a layperson is able to very robustly assess the investigated beauty parameters from highly standardized pairs of photographs. Further the result was not disturbed even by the use of other cosmetics as lipsticks. We believe that a low number of stimulus persons and expert raters are sufficient to successfully preselect candidate products for claim support studies on facial radiance and other beauty parameters.

Results



Figure 2: Soft focus assessment using the device of Figure 1

Representative pictures of forearm skin areas under 38 different angles to incident light. Shown is a test area of a subject treated with one of the 6 evaluated test products.

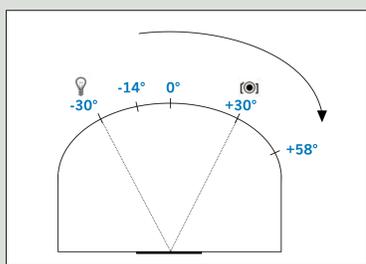


Figure 1: Scheme of the goniometric device for soft focus screening and set-up of the photography system: A stepwise camera movement from -14° to $+58^\circ$ (step size 2°) was utilized. From the series of images angle dependant reflection curves (Figure 3) were obtained by image analysis.

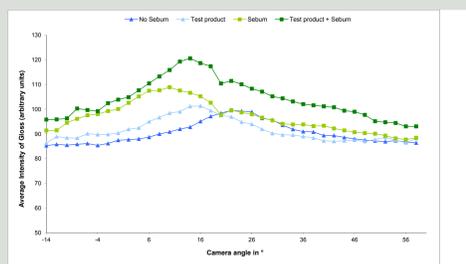


Figure 3: Results of soft focus screening

Representative example of the average gloss intensity on the volar forearm for the selected test product. Untreated forearm skin (no sebum, dark blue curve) showed a comparable reflectance curve as test product alone (light blue curve), meaning that no soft focus effect was found. Sebum alone (light green curve) showed a lower reflectance peak as the test product applied on sebum treated skin (dark green). In this case the product produced an inverse soft focus effect.



Figure 4:

Reproducible high quality photography before and after application of the test product. Photographs of two representative subjects without (upper image, left side) and with test product (upper image, right side) were presented to the raters on color calibrated monitors as pairs in a blinded and randomized manner. Raters subsequently judged which of the two images they preferred.

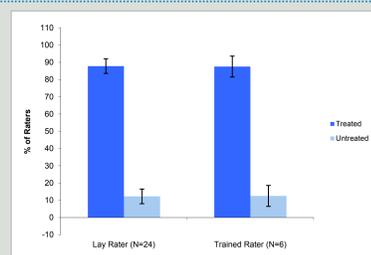
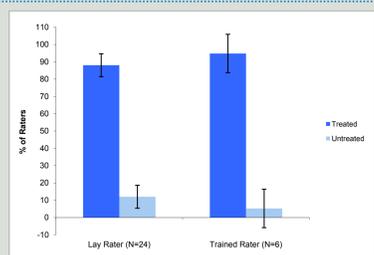
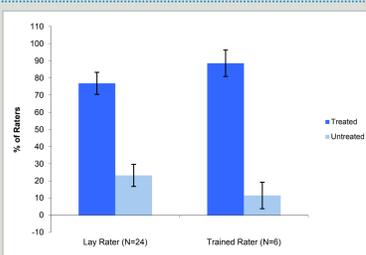
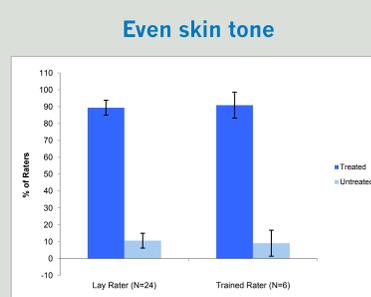
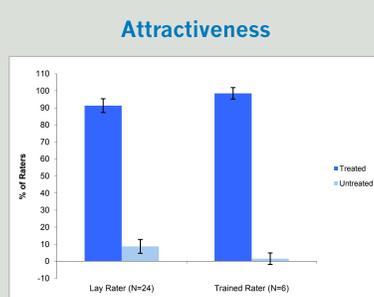
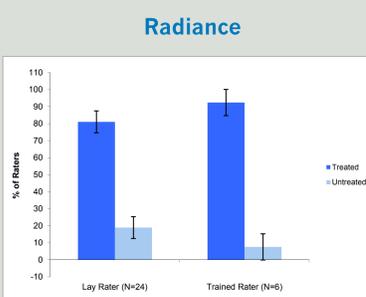


Figure 5: Percentage of raters seeing better radiance, attractiveness, even skin tone with the test product applied (dark blue) or untreated (light blue); (A) Stimulus persons without lipstick (N = 11); (B) Stimulus persons with lipstick (N = 16). For all three parameters lay raters and trained raters clearly assessed preference for the product treated faces. The lipstick had no relevant effect on the ratings.