J. Cosmet. Sci., 71, 399–410 (November/December 2020)

Objective Methods for Quantifying Color Retention and Bleeding Potential of a Lipstick Through Large-Scale Panel Perception and Controlled Usability Studies

DI QU, ALYSSA JOHNCOCK, and JIANWEI LIU, *Amway Innovation and Science, Ada, MI (D.Q.), (A.J.), Amway Innovation and Science – East Hub, Shanghai, 201203 China (J.L.)*

Accepted for publication June 17, 2020.

Synopsis

Color retention and bleeding are two properties of lipstick performance which impact consumer acceptability and preference. Before now, the claims of long-wear and non-bleeding had yet to be meaningfully substantiated quantitatively. Lack of objectively defined parameters and measurement methods inhibits one's ability of obtaining adequate claims support. In this study, we designed a 30-member, controlled-usability study to develop methods that quantify color retention and bleeding potential. Image analysis methods were used to measure lip color properties at the time points of baseline, immediate after application, and 8 hours of wear. Large-scale panel perception studies were conducted to obtain consumer acceptability criteria and establish consumer relevance of the objectively measured parameters. Our results show that the methods developed can quantitatively describe the color properties. Eight hours after application, the mean levels of color retention and bleeding potential for this specific lip color were 82.67% and 2.43, respectively. The results of panel perception studies helped establish threshold values for lip color performance acceptability, that is, if a color retention level was lower than 82.7%, or a bleeding potential level was higher than 2.37, the performance would not be deemed acceptable. The methods and results reported here serve as a useful tool for the evaluation of lip color performance in future development of quality lip color formulations.

INTRODUCTION

The global lipstick market was worth $12 billion in 2018 and was expected to reach $17 billion by 2024 according to a 2019 report (1). Lipstick has been used for nearly five thousand years as a way to enhance the texture of the lips, provide protection from the environment, and create an overall radiant facial appearance (2). The cultural, psychological, and sociological implications of lipstick use have been well documented (3–7). Phenomena of consumer lipstick purchasing behavior have also been studied. In a recent report, Zhi-xuani et al. (8) researched factors influencing lipstick purchases among Chinese female college students. It was concluded that the types of beauty products purchased were significantly affected by multiple psychological factors including their beauty- and

Address all correspondence to Di Qu at di.qu@amway.com.

fame-seeking motivations, as well as their emotional and diverse personality traits. Hill et al. (9) reported a lipstick effect in the United States, which concluded that women would purchase more beauty products to improve their appearance during economic downturns. Consumer satisfaction over product performance, especially long-lasting or long wear on the lips, is another important, perhaps more influential factor. A 2012 study in France had shown that wearing lipstick could have a significant influence on restaurant patrons' tipping behavior (10), and scientists in Japan had detected that wearing red lipstick enhanced the lightness of facial skin complexion (11). Instrumental and image analyses of properties of skin and lips have been active, which provide important quantitative descriptions of multiple color and morphological properties (12–17). However, there is little information in the literature regarding objective measurements of lipstick color retention and bleeding potentials because, up to now, lipstick property and performance have been primarily evaluated by small panel studies or laboratory testing using non-skin surfaces and various instruments (18–20). Although those tests are quick and convenient for prototype testing, the small sample size may lead to inadequate representation of product performance when it comes to final claims substantiation. Relating to measurement of color retention and bleeding, there was a Korean research in 2005, which examined lip color tone and spread (bleeding) while studying the lip wrinkles of 20 participants (21). The emphasis was made, however, to the local area around specific lip lines, and the pixels showing color spread (bleeding) were manually picked and measured with assistance of an image analysis software. The changes in overall color retention and comprehensive measurement of lip color bleeding were not studied. It is our belief that the objectively measured parameters from a relatively large test population ($N \geq 30$) would help achieve better measurement consistency, less variability in assessing product performance, and more relevance to consumer experience. Therefore, the aim of this study was to develop methods for the objective measurement of lipstick performance with regard to color retention and bleeding properties in a controlled usability study in conjunction with a large-scale panel perception study. This would allow us to evaluate a lipstick formulation and derive meaningful substantiated claims to drive consumer satisfaction in the marketplace.

MATERIALS AND METHODS

STUDY DESIGN

A liquid lip color formulation with bright red shade and silky matte finish was tested in our clinical testing laboratory in Ada, MI, where the temperature and relative humidity were controlled at 70 ± 1°F and 39% ± 2%, respectively. Thirty female volunteers participated in the study, with ages ranging from 30 to 63 years and of various ethnicities. They came to the laboratory in the morning and washed their face. After acclimating to the room conditions for 15 min, their frontal facial images were taken using VISIA-CR 4.3 (Canfield Scientific, Parsippany, NJ) together with a plate of color standards. These images were defined as the baseline (BSL) time point. They then applied the test product on their lips under the guidance of a laboratory scientist. After allowing the product to dry for 15 min, their images were taken again and were defined as the immediate after lipstick application (IMM) time point. The volunteers were then dismissed from the laboratory with instructions to drink beverages through a straw, cut food into small pieces

to eat, and refrain from excessively licking/touching their lips during the day. Eight hours after application of the lip color, the participants returned to the laboratory to have their images taken for the final 8-h (H08) time point.

QUANTIFICATION OF LIP PROPERTIES BY IMAGE ANALYSIS

Image analysis was performed on the clinical images using ImageJ, a freeware created and distributed online by the National Institutes of Health. Algorithms and analysis programs were developed in our laboratory using JavaScript. Steps involved in the analysis included (i) color correcting the images, (ii) outlining the lip area, (iii) measuring color intensity, and (iv) calculating lip color properties, namely, color retention and bleeding potential. To carry out the aforementioned analysis, each clinical image was first color-corrected against the embedded standard color chip using the least-square method described elsewhere (22). Lip area was detected using our in-house developed, automated algorithms to minimize inconsistent measurement results introduced by manual processing. Color intensity of the lips was measured from the detected lip area using L^*, a^*, and b^* in CIE-LAB color space. Color retention was derived from the $L^*a^*b^*$ values based on the definition shown by equation (1):

$$\text{Color retention, } \% = \frac{\Delta E_{\text{IMM}-\text{BSL}} - \Delta E_{T-\text{IMM}}}{\Delta E_{\text{IMM}-\text{BSL}}} \times 100\%, \tag{1}$$

where $\Delta E = \sqrt{\left(L_1^* - L_2^*\right)^2 + \left(a_1^* - a_2^*\right)^2 + \left(b_1^* - b_2^*\right)^2}$, in which the subscripts 1 and 2 are any time points of interest. Subscript BSL = baseline, IMM = immediate after lipstick application, and T = any time point at which a lipstick is evaluated for color performance. In this study, T = H08, 8 h after wearing the liquid lip color.

Some sample images with lip areas outlined for color retention analysis are shown in Figure 1A and C.

Color bleeding describes a phenomenon that occurs when the applied lip color spreads outside the boundaries of the lips, typically through wrinkles and lip lines. Examining the phenomenon, we defined a parameter of lip color bleeding as the ratio of line roughness of the detected lip border at time point T (which can be any time point after IMM, e.g., in this study, T = H08) over that at IMM, together with the influence of its corresponding color retention, as shown by equation (2).

$$\text{Bleeding}_{T-\text{IMM}} = \frac{\text{Line Roughness}_T}{\text{Line Roughness}_{\text{IMM}}} \times \text{Color Retention}_{T-\text{IMM}}. \tag{2}$$

Sample images of detected lip border before and after color bleeding are displayed in Figure 2A and C.

PANEL PERCEPTION

Consumer relevance of the color retention and bleeding potential was obtained from two separate perception studies in Shanghai, China, as well as in the United States. Color

Figure 1. Illustration of outlined lip areas for color retention analysis. (A) Lips of a volunteer at the BSL time point, (B) lips at the IMM time point, and (C) lips at the H08 time point. Color parameters L^*, a^*, b^*, and ΔE were measured, and color retention were calculated from those outlined regions of interest.

retention acceptability was determined using seven pairs of lip images from the IMM and H08 time points with varying levels of retention. Eighty-four panelists were shown the images and instructed to select the pairs of images exhibiting "acceptable" color retention

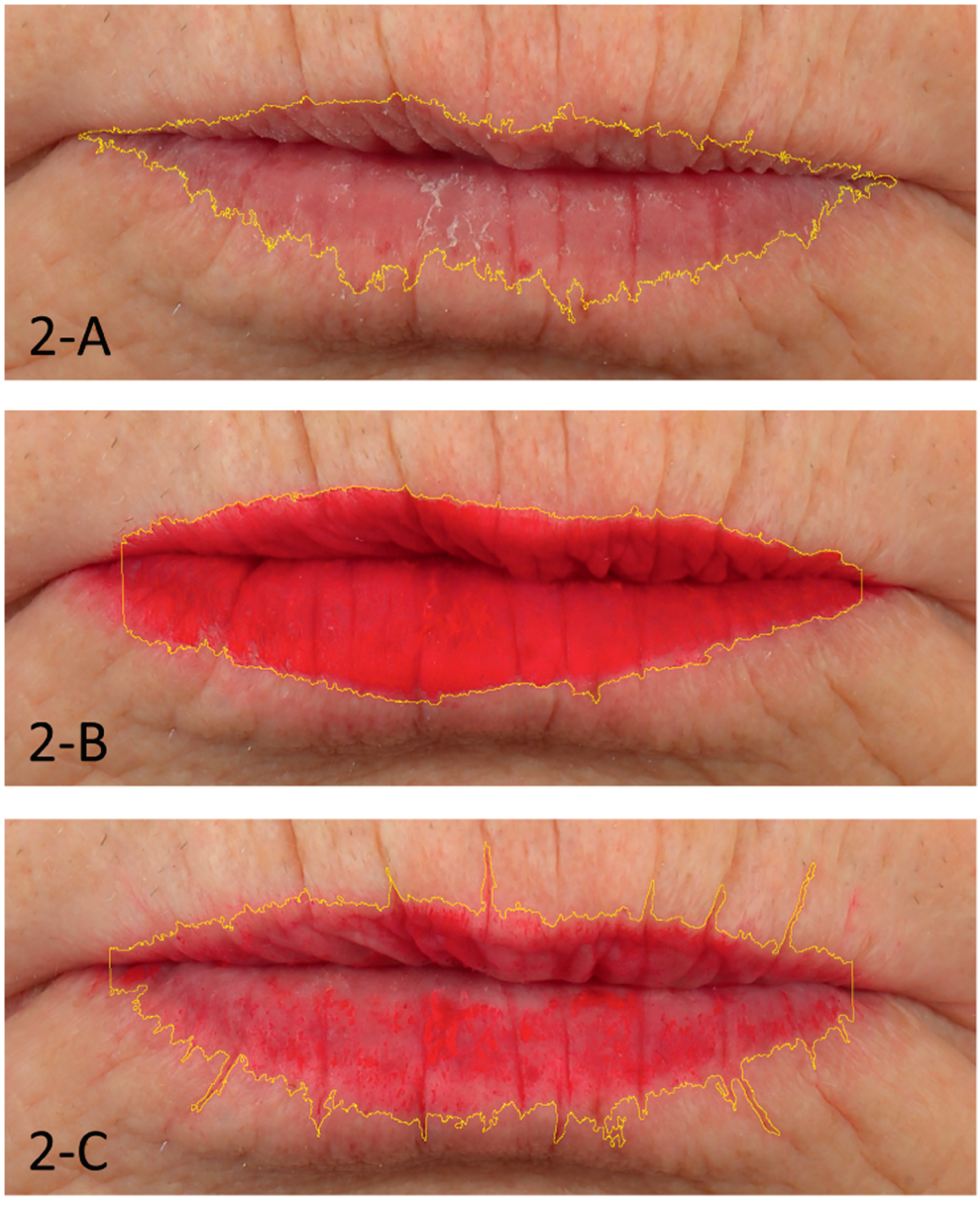


Figure 2. Illustration of detected lip areas for color bleeding calculation. (A) Lips at the BSL time point, (B) lips at the IMM time point, and (C) Lips at the H08 time point. The zig-zag lip border in (C) reflects the result of color bleeding. Line roughness and color bleeding potential can be calculated from the outlined regions of interest using equation (2).

after 8 h of wear. A similar method was used for color bleeding, in which 90 panelists were instructed to review six pairs of IMM and H08 images, and to select the pairs exhibiting "unacceptable" levels of bleeding. Those panel perceptual results were then correlated to their corresponding values of objectively measured parameters to determine a threshold of acceptability.

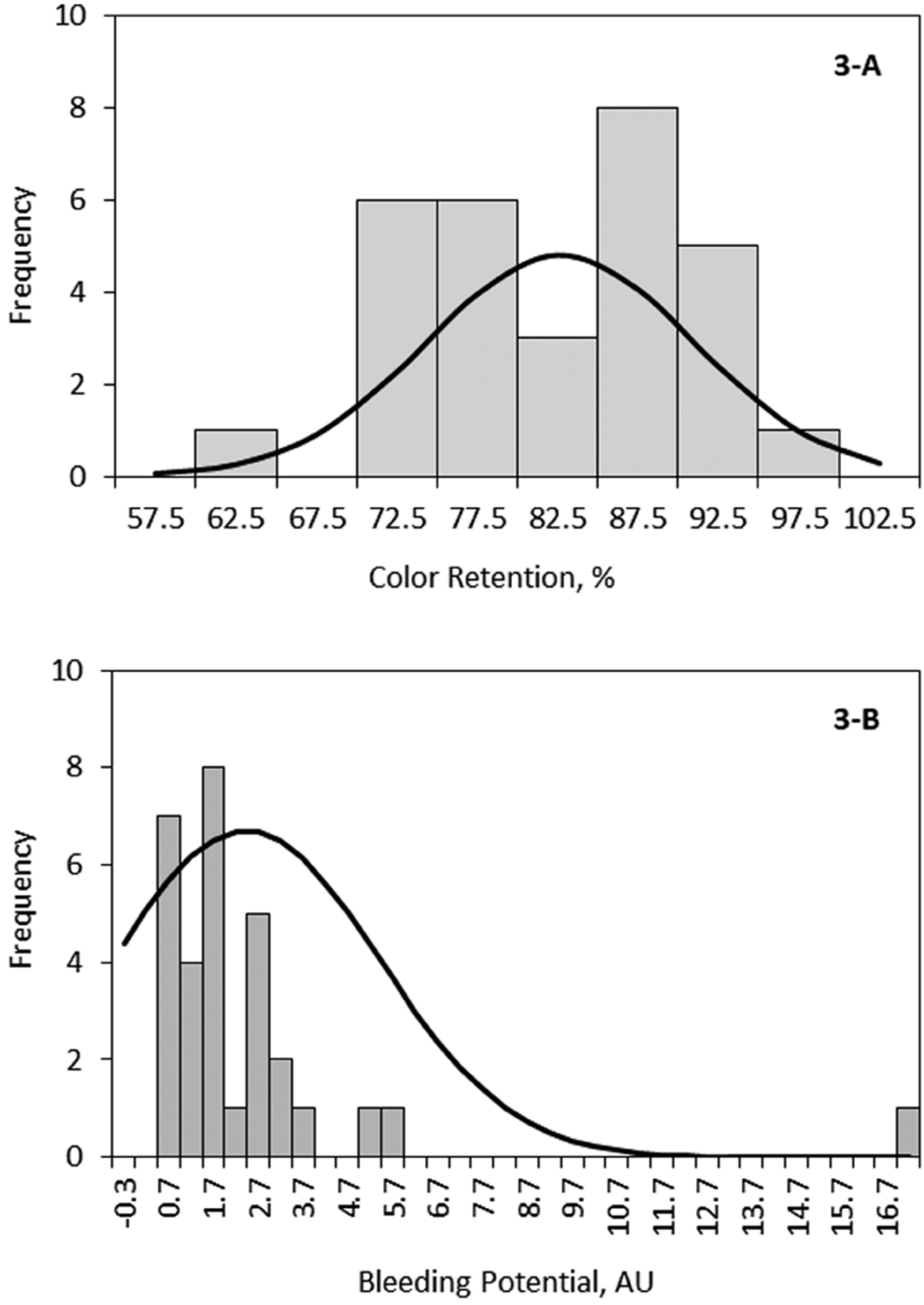


Figure 3. Distributions of the objectively measured lip color properties 8 h after application. (A) Color retention and (B) bleeding potential. Bars: actual distribution; curves: normal distribution.

STATISTICAL ANALYSIS

Binomial probability method was used to test the statistical significance of the panel perception results. Linear interpolation for threshold value between two measured datapoints was performed by determining the lowest possible percentage of panelists who perceived a color parameter acceptable while maintaining the statistical significance ($p < 0.05$ at the 95% confidence level).

RESULTS AND DISCUSSION

QUANTIFICATION OF COLOR RETENTION AND BLEEDING

Our automated lip detection algorithms adequately detected the lip area from the clinical images, such as those illustrated in Figures 1 and 2, which made it possible to measure color retention and determine bleeding potential 8 h after application. A mean color retention of 82.67% with a standard deviation of 8.27% was obtained from the test population in this study, as shown in Figure 3A. The observed variability was believed to come from the life habits of participants who paid variable attention to keeping the color on their lips in their daily activities.

For color bleeding potential (Figure 3B), the results fell in a linear scale ranging from zero to about 17, with a mean of 2.43 and a standard deviation of 2.97. It was noticed that most of the volunteers did not show much color bleeding around the lips, although the phenomenon was significantly visible on individuals with excessive wrinkles in the mouth area, skewing the data toward the higher end of the bleeding potential range.

PANEL PERCEPTION

The results of panel-perceived acceptability on lip color retention are shown in Table I, where the percentages of the panel who perceived pairs of before and after images as "acceptable" are displayed together with the corresponding lip images. The *p* values of binomial probability tests were also shown to validate the perception results with statistical significance.

For color bleeding, although the perception survey asked for unacceptability, we converted the results to acceptability, that is, %Panel (acceptable) = 100% – %Panel (unacceptable), to make the format of analysis consistent. The results are shown in Table II in the fashion similar to those of color retention.

Table I
Peception Results on Acceptability of Lip Color Retention Levels

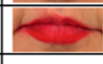
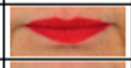
	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6	Pair 7
IMM							
H08							
Percentage of panel perceived color retention "acceptable"	92.9% (78/84)	85.7% (72/84)	79.8% (67/84)	13.1% (11/84)	4.8% (4/84)	3.6% (3/84)	3.6% (3/84)
Binomial *p* value	4.0E-15	5.9E-11	4.9E-08	1.3E-11	1.1E-16	1.7E-17	1.7E-17

Table II
Perception Results on Acceptability of Lip Color Bleeding Levels

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
IMM						
H08						
Percentage of panel perceived color bleeding "unacceptable"	93.3% (84/90)	72.2% (65/90)	13.3% (12/90)	61.1% (55/90)	5.6% (5/90)	11.1% (10/90)
Binomial *p* value	2.0E-16	2.5E-05	3.5E-12	3.5E-02	3.4E-17	1.6E-13

CONSUMER RELEVANCE OF THE QUANTIFIED LIP COLOR PROPERTIES

Combining panel perception results with their corresponding levels of color retention and bleeding potential measured objectively, we obtained charts to demonstrate consumer relevance. As shown in Figure 4A, the percentage of panelists who perceived the levels of color retention acceptable decreased with the levels of the objectively measured property. Between color retention levels of 85% and 77.1%, the percentage of acceptability dropped dramatically from 79.8% to 13.1%, indicating the existence of a threshold in the region.

For the bleeding potential, the percentage of panelists who perceived the levels of color bleeding acceptable decreased with increase in the bleeding potential measured objectively. Between the levels of 1.94 and 2.72, the percentage of acceptability dropped dramatically from 86.7% to 38.9%, suggesting a threshold in the region, as shown in Figure 4B.

Interpolating within the threshold region using a binomial probability test at the given population of each perception study and under the condition of $p < 0.05$, we first determined the minimal percentages of panel acceptability to be 60.12% and 60.56% for color retention and bleeding potential, respectively. The corresponding threshold values of color retention and bleeding properties were then derived from the linear equation, with the values being 82.67% and 2.37 for color retention and bleeding potential, respectively. Plotting the color parameters of all 30 participants in this study, we obtained Figure 5A and B to show the color properties relative to their threshold of acceptability. Sorting the color retention levels from high to low, we see in Figure 5A a gradual decrease in the property among different participants, which reaches the threshold at 82.67%, beyond which the color retention levels become unacceptable. Similarly, in Figure 5B, we see a gradual increase in bleeding potential reaching the threshold at 2.37 beyond which the bleeding level is too high to be acceptable. With these charts, we would be able to quantitatively evaluate the performance of a new lip color formulation during product development and final claims' substantiation. After using it for an expected time period such as 8 h in a usability study, should the measured mean value of color retention be higher than the threshold of 82.67%, or the mean value of bleeding potential be lower than 2.37, it would indicate with statistical confidence that a claim of "long-lasting" or "non-bleeding" be substantiated.

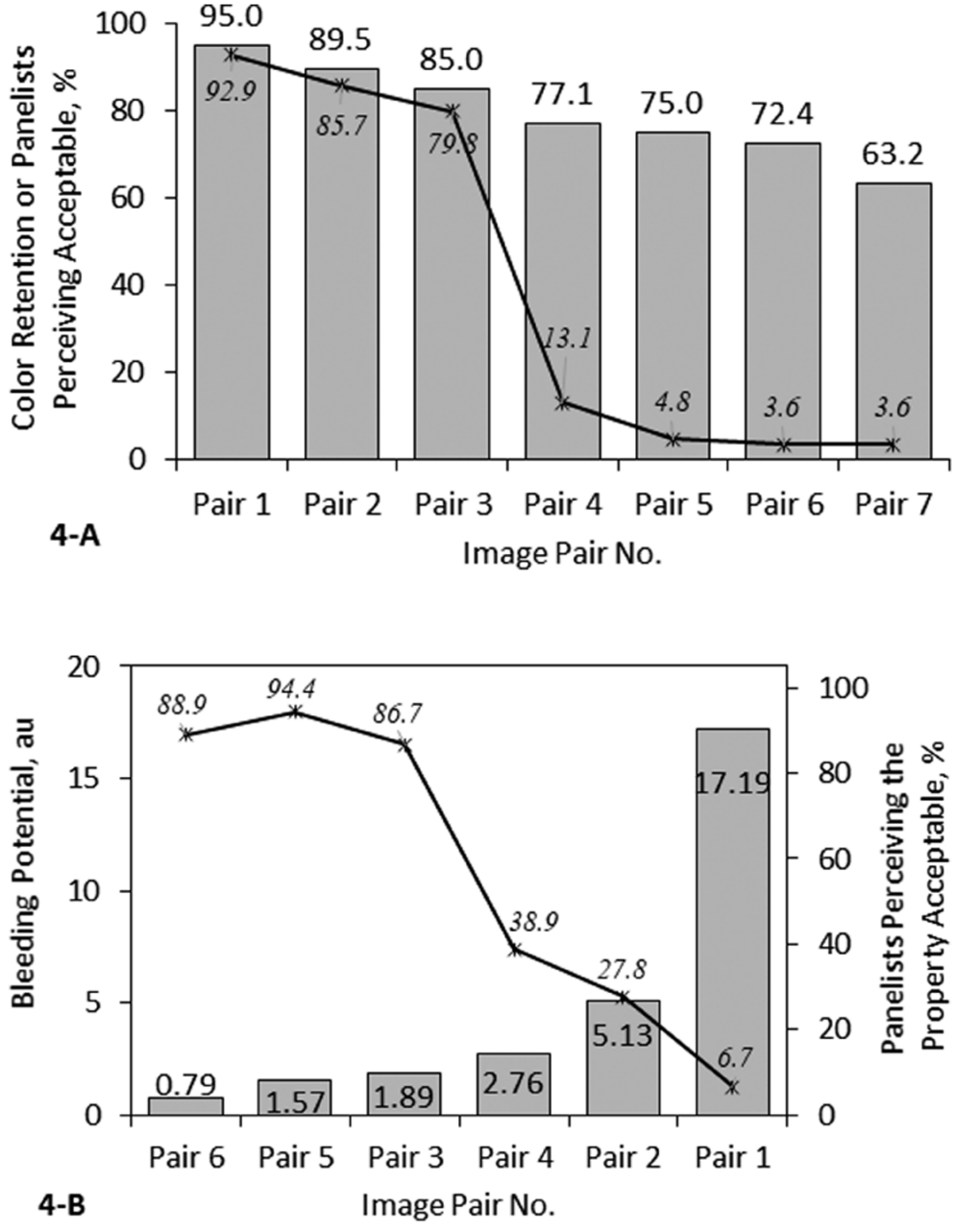
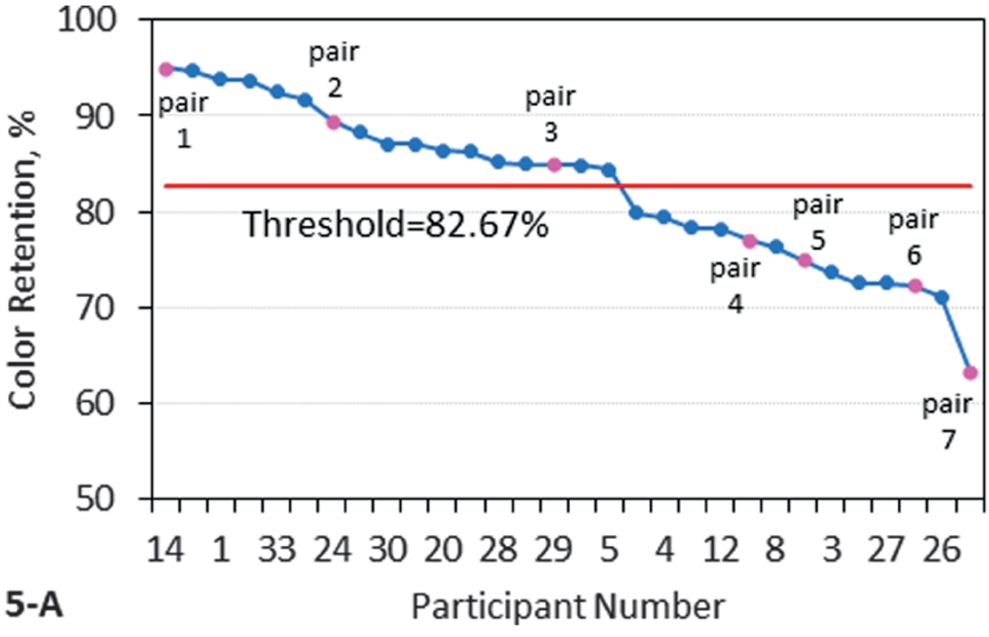


Figure 4. Correlations between results of panel perception and objectively measured lip color properties. (A) Color retention and (B) bleeding potential. Bars: objectively measured levels of color retention or bleeding. Solid lines with italic data label: percentage of panelists who perceived the property "acceptable."

CONCLUSION

This study quantifies changes in overall color properties of the lips after application of a liquid lip color in a controlled usability study. The image analysis methods were effective



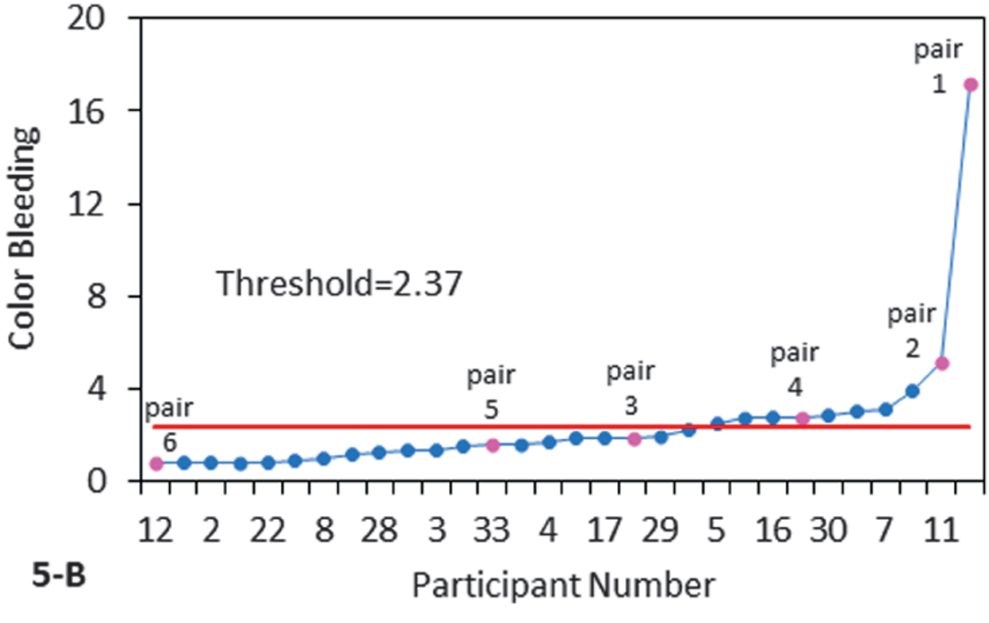


Figure 5. Quantified lip color properties and their corresponding threshold of acceptability. (A) Color retention and (B) bleeding potential. Colored dots with tracing line: objectively measured levels of color retention or bleeding potential. Pink dots: sample results used in panel perception studies. Red line: the threshold level within which consumers would consider the levels of a color property acceptable.

in outlining the lip area and detecting changes of color properties immediately after product application and after 8 h of wear. The color retention and bleeding potential parameters established in this study were adequate of quantitatively describing the phenomena. The panel perception study performed on a select group of the resulting images generated meaningful threshold levels when correlated with the objectively measured results. The controlled clinical setting of the study with a statistically acceptable study population helped generate consistent results for method development. Over the 8 h studied, the mean value of color retention decreased and that of color bleeding increased. The methods established in this study provide a quantitative description of those color properties with consumer relevance and could serve as an objective means to substantiate "long-lasting" and "non-bleeding" claims of lip color products.

ACKNOWLEDGMENTS

We would like to thank Robin Thacker, William Schaefer, and Elizabeth Gyurke for their assistance in carrying out the clinical study, and Sarah Whitehead for her review and comments on the article.

REFERENCES

(1) Lipstick Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2019–2024 IMARC (2019), https://www.imarcgroup.com/lipstick-market. Accessed on April 2, 2020.
(2) J. Pallingston, Lipstick: A Celebration of the World's Favorite Cosmetic (St. Martin's Press, New York, NY, 1999).
(3) W. McKeachie, Lipstick as a determiner of first impressions of personality: an experiment for the general psychology course, *J. Soc. Psychol.*, **36**, 241–244 (1952).
(4) L. Scott, "Rethinking necessities," in Book Fresh Lipstick: Redressing Fashion and Feminism (Palgrave MacMillan, London, UK 2005), pp. 191–222.
(5) L. Clarke and A. Bundon, From 'the thing to do' to 'defying the ravages of age': older women reflect on the use of lipstick, *J. Women Aging*, **21**, 198–212 (2009).
(6) M. Ogilvie and M. Ryan, Lipstick: more than a fashion trend, *Res. J. Soc. Sci. Manag.*, **1**, 117–128 (2011).
(7) N. Guéguen, Does red lipstick really attract men? An evaluation in a bar, *Int. J. Psychol. Stud.*, **4**, 206–209 (2012).
(8) L. Zhi-xuani, G. Hanchen, and L. Feifan, Analysis of influencing factors of Chinese female college students' lipstick purchasing decisions, Proceedings of 2019 International Conference on Emerging Researches in Management, Business, Finance and Economics (ERMBFE Oct. 15-16, 2019), Shanghai, China (2019), pp. 898–905.
(9) S. Hill, C. D. Rodeheffer, V. Griskevicius, K. Durante, A. E. White, Boosting beauty in an economic decline: mating, spending, and the lipstick effect, *J. Personal. Soc. Psychol.*, **103**, 275–291 (2012).
(10) N. Gueguen and C. Jacob, Lipstick and tipping behavior: when red lipstick enhance waitresses tips, *Int. J. Hospital. Manag.*, **31**, 1333–1335 (2012).
(11) Y. Kiritani, A. Okazaki, K. Motoyoshi, R. Takano, and N. Ookubo, Color illusion on complexion by lipsticks and its impression, *J. Psychon. Sci.*, **36**, 4–16 (2017).
(12) K. Kikuchi, Y. Masuda, T. Yamashita, E. Kawai, and T. Hirao, Image analysis of skin color heterogeneity focusing on skin chromophores and the age-related changes in facial skin, *Skin Res. Technol.*, **21**, 175–183 (2015).
(13) E. Tamura, J. Ishikawa, K. Sugata, K. Tsukahara, H. Yasumori, and T. Yamamoto, Age-related differences in the functional properties of lips compared with skin, *Skin Res. Technol.*, **24**, 472–478 (2018).
(14) H. Kim, Age-related changes in lip morphological and physiological characteristics in Korean women, *Skin Res. Technol.*, **25**, 277–282 (2019).
(15) M. Lee, J. Han, and E. Kim, An evaluation of the effects of makeup on perceived age based on skin color in Korean women, *Int. J. Dermatol.*, **18**, 1044–1051 (2019).

(16) F. Li, C. Zhao, Z. Xia, Y. Wang, X. Zhou, and G.-Z. Li, Computer-assisted lip diagnosis on traditional Chinese medicine using multi-class support vector machines, *BMC Compl. Altern. Med.*, **12**, 127–140 (2012).
(17) Y. Park and D. Qu, Lip appearance index derived from image analysis and consumer perception, Presented (Podium) at the Society of Cosmetic Chemists Annual Scientific Meeting, New York, NY (2013).
(18) F. Giron and F. Brie, Process for Measuring Transfer Resistance of a Cosmetic Product, US patent No. 7,258,023 (2007).
(19) C. Richard, B. Tillé-Salmon, and Y. Mofid, Contribution to interplay between a delamination test and a sensory analysis of mid-range lipsticks, *Int. J. Cosmet. Sci.*, **38**, 100–108 (2016).
(20) D. Rafferty, L Dupin, J. Zellia, A. Giovannitti-Jensen, Predicting lipstick sensory properties with laboratory tests, *Int. J. Cosmet. Sci.*, **40**, 451–460 (2018).
(21) J. Ryu, S.-G. Park, T.-J. Kwak, M.-Y. Chang, M.-E. Park, K.-H. Choi, K.-H. Sung, H.-J. Shin, C.-K. Lee, Y.-S. Kang, M.-S. Yoon, M.-J. Rang, and S.-J. Kim, Improving lip wrinkles: lipstick-related image analysis, *Skin Res. Technol.*, **11**, 157–164 (2005).
(22) D. Qu, Y.-B. Chen, L. Feng, Improving precision and accuracy of skin color measurement by using image analysis methods with individual image color correction algorithms, *Chin. J. Appl. Mech.*, **26**, 93–96 (2017).