The skin sensitivity index: paving the way to objective evaluation and grading of skin sensitivity

Rengot, Juliette[1*]; Stuhlmann, Dominik[2]; Meyer, Imke[2]; Le Maire, Marielle[3]; Chamla, Julie[3]; Gierschendorf, Jordan[1]; Cherel, Marie[1], Prestat-Marquis, Elodie[1]

[1] Newtone Technologies, Lyon, France; [2] Symrise AG, Holzminden, Germany; [3] Symrise SAS, Clichy-la-Garenne, France.

* 13bis place Jules Ferry 69006 Lyon, +33 (0)4 72 69 83 20, jrengot@newtone.fr

Abstract

Background: Sensitive skin impacts life quality, causing skin dryness, itching sensations, appearance of pimples and redness… Its evaluation is still now mainly based on consumers' self-perception. In this study, we present a method to calculate objective skin sensitivity grades from multi-parametric hyperspectral image analysis.

Methods: *In vivo* studies were performed to collect data from a large and representative panel: sting test results, self-assessment questionnaires, expert assessments and SpectraFace® hyperspectral images.

First, we selected the most relevant information from self-perception questionnaires based on clinical experience. We designed a clinical sensitivity index with a weighting of only four answers.

To go further, we developed an objective instrumental sensitivity index based on hyperspectral images and machine learning. Face skin color, homogeneity and chromophore parameters were computed and used to train a neural network with regards to the clinical sensitivity index results.

Results: The sting test results were not well correlated with the sensitivity expert assessments.

The questionnaire answers presented a low repeatability, proving the complexity of repeatable self-assessment. We selected the most reliable subjects to construct trustworthy indices.
The introduced indices provided continuous and fine sensitivity scales. For sensitive *versus* non-sensitive classification, the clinical sensitivity index achieved a precision of 97%. The instrumental sensitivity index reached a correlation $R^2 = 0.9726$ with the clinical index.

Conclusion: The proposed innovative tool allows calculating skin sensitivity indices from self-perception questionnaires or instrumental measurements. It could support the assessment of skin sensitivity evolution under environmental changes, over time, or upon the use of a soothing product.

Keywords: Skin sensitivity; Artificial Intelligence; Hyperspectral imaging; Consumer sciences; Skin chromophores measurement;

Introduction. Even if skin sensitivity is one of the most common disturbing skin conditions, its definition, the recurrent causes and the possible soothing factors remain unclear. Sensitivity can be identified as the feeling of skin discomfort without any clinical evidence of skin lesions [1]. It can highly impact life quality. According to Symrise consumer and market insights [2], 56% of subjects declare having sensitive skin, a large and increasing number. A quarter of them considers it is a persistent condition. As first perceptions of their skin sensitivity, consumers identify a skin dryness, an itching sensation, as well as the appearance of pimples and redness. Adverse reactions occur in response to various types of agents [3, 4]: environmental factors (sun, wind, cold…), hygiene, skincare or cosmetic products (soap, cream, foundation…), stress, emotional burden…
Analyzing skin sensitivity is a challenge because sensations are not visible and objective attributes. Until now, sensitivity studies are mainly based on consumers' self-perception. These self-assessments can be collected with questionnaires [5]. Some tests, like the prick test [6] or the sting test [7, 8], have been implemented to better understand the sensitive skin reactions. However, they are still difficult to standardize, objectify and interpret.

No analytical tool and reliable approach to evaluate accurately skin sensitivity has already reached a consensus among experts. In this study, we present a complete method to calculate objective and robust skin sensitivity indices based on multi-parametric face skin hyperspectral image analysis. We developed two objective skin sensitivity indices based on either the self-perception or instrumental data. Both are complementary and serve the experts according to their study targets and constraints. This article details the development process of these indices.

Materials and Methods. We aimed at designing skin sensitivity indices that enables experts to objectively quantify how much the subject 's skin is sensitive, either based on subject 's self-perception or on instrumental measures. The evaluation methods should be simple, quick, reliable, repeatable and not expert dependent. Our endpoint was to develop a solution that improves the classical binary classification "sensitive / not sensitive". Our tools should assign continuous grades to propose a finer skin assessment that could shed light on several severity levels of skin sensitivity.

Sting test. To start with, we conducted a first clinical study including different instrumental measurements and sensitivity assessments. The objective was to study the correlation between sensitivity clinical assessment and classical sting test results. Data acquisitions lasted from December 2018 to March 2019 in a set of 90 Caucasian women, with sensitive skin or not, in France (Lyon). All participants provided written informed consent.

During the inclusion interview, the subjects answered a questionnaire in which they declared the cutaneous sensations they felt towards various factors. Based on these answers and additional visual skin evaluations (shiny aspect, redness, dryness…), a trained expert judged whether the subjects had a sensitive skin or not, and thus provided a binary classification of the panel..

The sting test was done on nasogenian fold area. The subjects were placed 10 minutes in front of a Vapozone®. 4mL of lactic acid solution at 10% was applied on right or left side (randomization). On the opposite side, 4 mL of water physiological serum was applied, as a control area. We collected the subject 's self-assessment of the intensity of the felt prickling, heating and burning sensations immediately after the application of the solutions. By using

visual scales, the subjects attributed a score between 0 and 4 to each kind of reaction. A global score is defined as the difference of the intensities of the felt sensations obtained on treated side by lactic acid and the one treated with physiological serum. The global score can vary from -12 to 12. The higher the score, the higher the skin sensitivity. A subject is considered as stinger positive if the global score is higher or equal to 4.

Clinical skin sensitivity index. A binary classification may not be sufficient to appraise soothing product efficacy, because it does not enable to highlight slight improvement evolutions of skin sensitivity. A more precise tool is needed to quantify the benefit of cosmetic products on skin sensitivity status.

To begin with, we conceived a clinical skin sensitivity index based on self-perception questionnaires. Only four questions were selected to make the evaluation quick and easy to implement:

- Do you feel abnormal and repeated reactions on the face to care products?
- Do you feel abnormal and repeated reactions on the face to hygiene products?
- Do you feel abnormal and repeated reactions on the face to the environment?
- Do you feel abnormal and repeated reactions on the face to other factors?

The respondents had the possibility to answer: “no reaction”, “slight reactions”, “moderate reactions” or “marked reactions”. This questionnaire covered a large skin sensitivity range with a few simple questions. We gave to each question and each answer a weight [Table 1], based on our understanding of skin biology, but as close as possible to the study approach, because they do not have the same impact on skin sensitivity severity. “No reaction” was given a score of zero and increasing declared severity were given increasing scores. A higher score was attributed to reactions to skin care cosmetics with regards to hygiene products or environmental factors. The sum provided the clinical skin sensitivity index. The scores could vary from 0 (not sensitive) to 15 (highly sensitive), and thus provide a fine sensitivity scale. A subject was considered as sensitive if the clinical index was higher than or equal to 2.

	No reaction	Slight reactions	Moderate reactions	Marked reactions
Care products	0	2	4	6
Hygiene products	0	1	2	3
Environment	0	1	2	3
Other factors	0	1	2	3

TABLE 1: Question-and-answer weightings for clinical skin sensitivity index computation.

Questions were always asked the same way, under a calm environment, taking care not to hurry the subjects.

Instrumental skin sensitivity index based on hyperspectral image analysis. An index based on self-perception scores has the advantages to be simple and remotely computable without any equipment. However, it has also some drawbacks. It contains a part of subjectivity and may vary according to the subject ’s mood and the way the questions are asked. We developed an additional instrumental skin sensitivity index to bring a more objective solution for sensitivity classification.

A new data collection campaign was carried out from December 2021 to March 2022. By phone calls, the questionnaire answers are collected, enabling us to compute the clinical index of each potential subject. It enables us to create a panel well distributed over the whole sensitivity severity range. At the end, 183 Caucasian women were included in the study. Written informed consent was obtained from all participants. The same questionnaire was asked two additional times, with a delay of two weeks. It allowed analyzing repeatability, by computing the maximal deviation, the mean deviation and the standard deviation of the clinical index. It also allowed us to identify the most reliable and repeatable subjects. In addition, two consecutive images of the subject ’s faces were acquired with SpectraFace® [9], a hyperspectral acquisition system. Images were collected for all the subjects on both front face and profile views. From the acquired images (in 31 wavelengths over the whole visible light spectrum), skin color images and skin chromophore maps (oxygenated and deoxygenated hemoglobin) were reconstructed [10, 11, 12].

Then, we defined four Regions Of Interest (ROIs): the nasogenian folds, the sides of the nose, the cheeks and the whole half-faces [Figure 1]. We manually drew the ROIs for all the subjects at the first timepoint. Spatial rigid registration based on image intensities was used to automatically put the ROI back on the second timepoint. A visual quality control guaranteed the algorithm was reliable and precise.

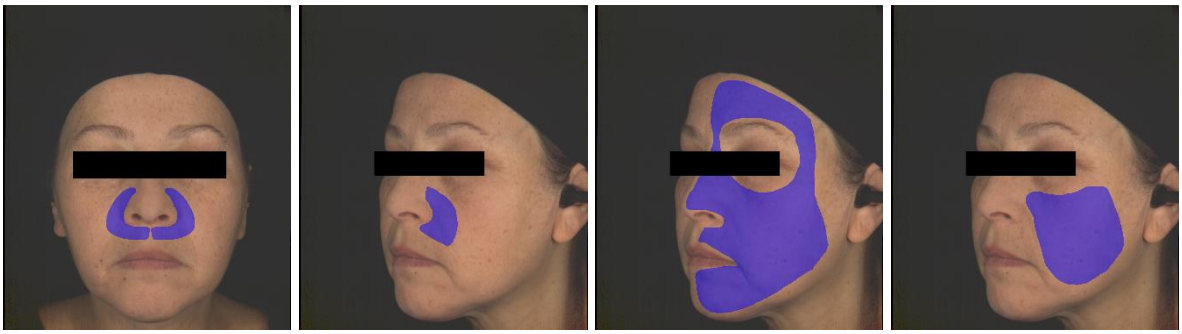

FIGURE 1: The regions of interest, illustrated on RGB reconstructions of one anonymized subject.

For each analyzed skin area, we computed six parameters of interest, characterizing skin color and homogeneity or chromophores features:

- L*, a* and b* for skin color information [13]
- H76 for homogeneity information

$$H76 = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(L_i^* - \mu L^*)^2 + (a_i^* - \mu a^*)^2 + (b_i^* - \mu b^*)^2}$$

with N the number of pixels inside the ROI

L_i^*, a_i^* and b_i^* the L*a*b* values of the pixel i

μL^*, μa^* and μb^* the L*a*b* average values inside the ROI

- Oxygenated hemoglobin and deoxygenated hemoglobin rates for chromophore information

These parameters were selected because they provide diverse and complementary measurements. Average values from the right and left sides of the face were used in the analysis to gain in robustness.

The proposed approach for creating an instrumental skin sensitivity index rests on machine learning. We designed a multi-layer perceptron (MLP) neural network [14]. The model input was the set of 24 parameters previously introduced. The objective was to predict a skin sensitivity score. The ground truth was the median of the clinical index repetitions. To train our regression model, we needed to separate our database into a training dataset and a testing dataset. The training dataset was used to fit the model parameters. The testing dataset was used to check the model performances and its ability to generalize to new samples. Of course, no data sample belonged to both sets at the same time. We rejected the subjects with too high variations in clinical index repetitions because they were judged not reliable enough.

Results.

Sting test. A majority of the subjects was found stinger negative (52 negatives against 38 positives) whereas most of the subjects was judged as having a sensitive skin by the clinical expert (70 subjects with a sensitive skin against 20 not having a sensitive skin). This discrepancy was confirmed by the low results of accuracy (0.51) when using the sting test as a predictive tool to determine skin sensitivity status [Figure 2]. The accuracy is the number of correct predictions divided by the total number of predictions. These results encouraged us to define another method to predict the sensitivity status with a better accuracy.

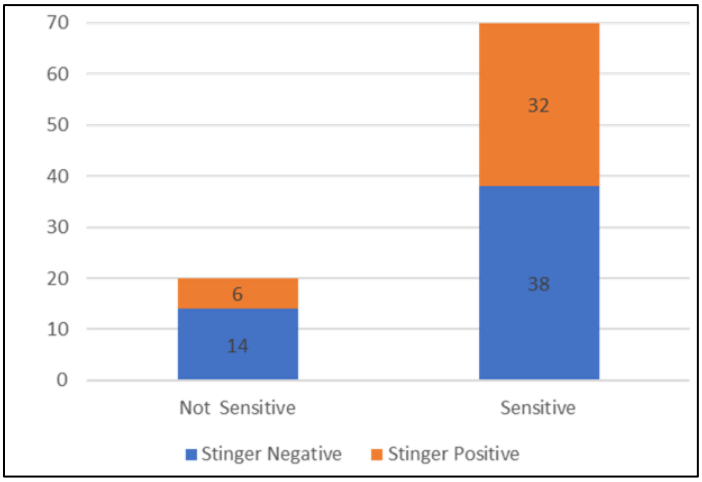

FIGURE 2: The sensitivity distribution across stinger status.

Clinical skin sensitivity index. Our index based on self-perception was defined according to expert evaluation methods. Its coherence was validated by an accuracy of 0.97 in the sensitive / not sensitive predicted classification by an expert [Table 2].

Confusion matrix		**Status prediction**		***Total***
		Not sensitive	**Sensitive**	
Expert classification	**Not sensitive**	17	3	*20*
	Sensitive	0	66	*66*
Total		*17*	*69*	*86*

TABLE 2: The confusion matrix between the predicted sensitivity status and the expert classification.

Questionnaire repeatability. In the second study, the four questions were asked three times to the subjects. Maximal, mean and standard deviations of the resulting scores were computed [Table 3]. High variations in the answers were observed. Only 13% of subjects gave the exact same answers for all repetitions [Figure 3]. Some subjects went from one extreme to the other. It proved the high difficulty to make trustworthy evaluations of skin sensibility only based on self-perception. It also encouraged us to develop another index based on objective measurements.

	Maximal deviation	Mean deviation	Standard deviation
Mean	3.34	1.51	2.09
Median	3.00	1.11	1.53
Maximum	12.00	6.00	8.49

TABLE 3: Deviation computations for questionnaire repeatability.

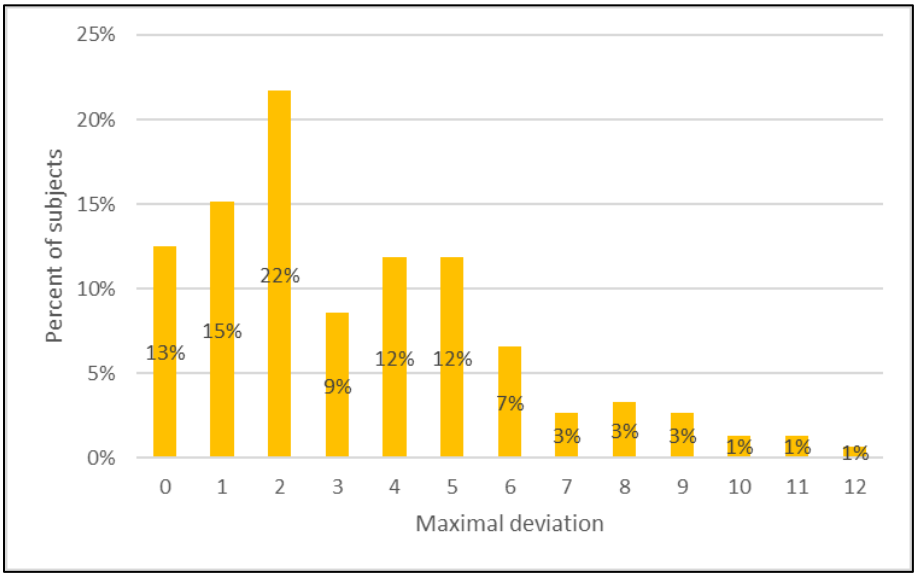

FIGURE 3: The distribution of the computed maximal deviations.

This changeability can be explained either by a variation in the question-and-answer understandings or by real evolutions of skin sensitivity in a short period of time. In both cases, the subjects with completely different answers are difficult to apprehend. We split the subjects in three groups, depending on the repeatability of their answers:

- the reliable panel: it gathered all the subjects whose maximal deviation was lower than or equal to three. This group was made up with 91 subjects.

- the acceptable panel: it tolerated higher variations for high sensitive grades. The maximal deviation had to be lower than or equal to 3, 4, 5 or 6 if the median clinical skin sensitivity index was lower than 3, between 4 and 5, between 6 and 7 or higher than 8 respectively. 14 additional subjects were included in this category.
- The unreliable panel: it grouped all other subjects, the ones who gave completely different answers. 50 subjects were rejected.

Instrumental skin sensitivity index based on hyperspectral image analysis. First, we focused on the reliable panel. We created a training set with 98 randomly selected subjects. The 21 remaining subjects were used as the testing set. Obviously, the two acquisitions of the same subject were included in the same set.

Once the model was trained, we validated its performances on new samples (the testing set). We reached a really high level of correlation between the MLP predictions and the ground truths: $R^2 = 0.9726$ [Figure 4]. It proved that our model was able to generalize.

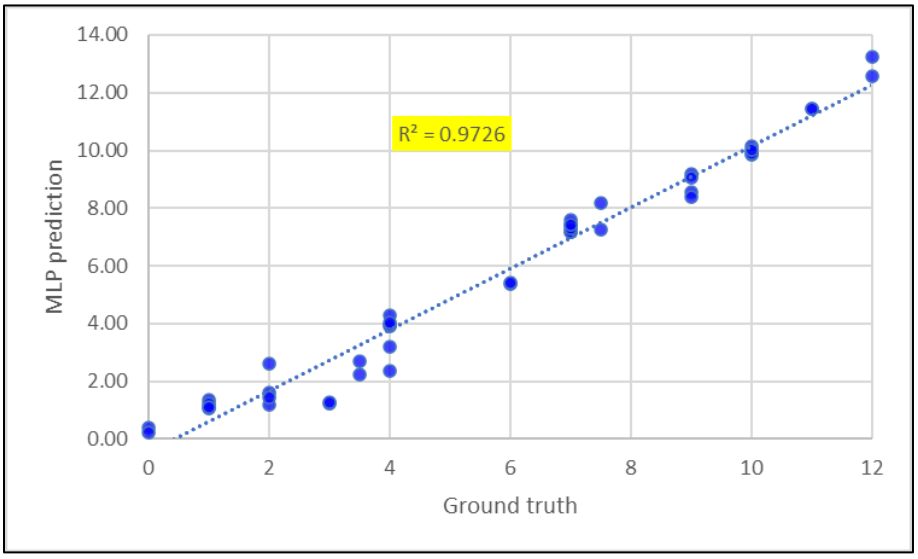

FIGURE 4: The correlation between the MLP predictions and the clinical index scores on the test set.

As SpectraFace® images were repeated twice, we also verified that our MLP model is robust to small positioning variations. The correlation between the two predictions was also really high: $R^2 = 0.9958$ [Figure 5].

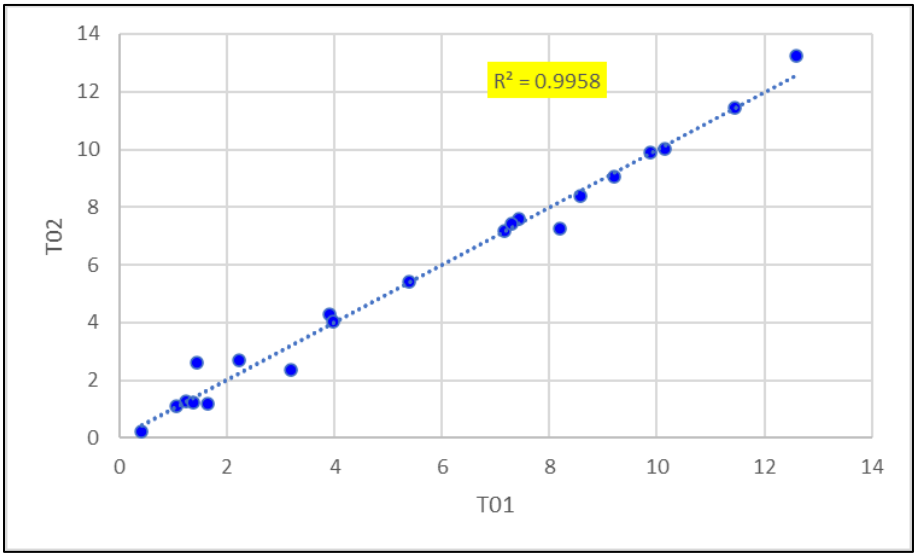

FIGURE 5: The correlation between the MLP prediction at two timepoints on the test set.

The maximal observed absolute difference between the instrumental index and the clinical index was equal to 1.77. That was satisfying, as the intrinsic variation acceptability in the ground truths was set to 3. It implies that, to underline a real product effect, the computed variation in sensitivity must be higher than 3.

We showed that our model predicts continuous skin sensitivity grades that allows small variations follow-up over time. In order to make the index interpretations easier, it is also possible to group the predicted grades into skin sensitivity severity levels:

- No sensitivity ($0 \leq \text{index} < 2$)
- Low sensitivity ($2 \leq \text{index} < 6$)
- Medium sensitivity ($6 \leq \text{index} < 9$)
- High sensitivity ($\text{index} \geq 9$)

The confusion matrix [Figure 6] proved that our model is precise: the accuracy equals to 0.83. We can note that our MLP slightly tends to under-estimate the sensitivity, in particular for lower scores. It could be improved in the next steps by enhancing the reliable panel.

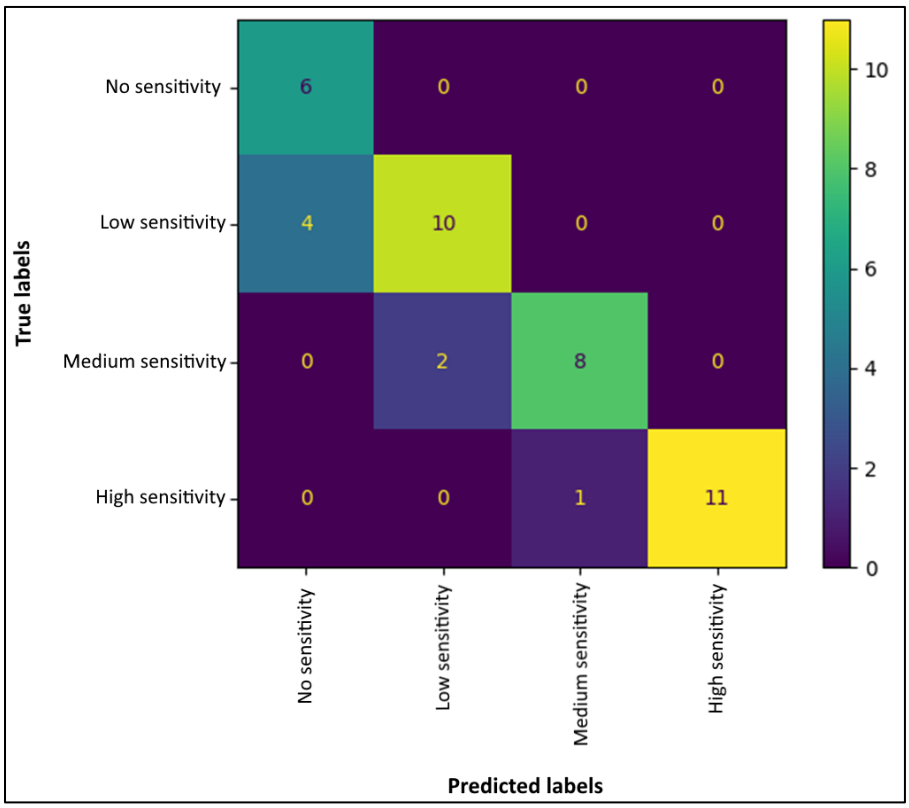

FIGURE 6: The confusion matrix for sensitivity group classification.

Afterwards, we enlarged the training set with the acceptable panel. We trained again the MLP, and we tested it on the same testing set as previously. The performances were affected. The determination coefficient between the predictions and the ground truths went from 0.9726 to 0.7950. It means that the added subjects were not reliable enough to bring relevant information for our model training. Thus, we considered the model trained on the reliable panel as the validated one.

Discussion. The introduced skin sensitivity indices offer new possibilities for clinical studies. Going from binary classifications to multi-grade scales provide a finer approach that can help to apprehend slight variations in skin sensitivity. It is critical for better understanding the skin sensitivity reactions and for assessing skincare product soothing performances.

In addition, our tools are not expert dependent. They make the sensitivity assessment more accessible. So, they should ease the implementation of clinical studies. Especially, the clinical index is computed with only four simple questions, making the process easy, quick and cheap. The questions can be asked even remotely, by phone calls. It is particularly

convenient for subject recruitment. It is now possible to select the subjects to be included according to their skin sensitivity to focus the clinical study on a specific population.
Next, the instrumental index is an innovative approach. It makes the most of hyperspectral acquisitions and neural networks. No similar tool already exists. It proposes a completely objective and repeatable estimations of skin sensitivity level, based on skin color parameters, homogeneity and hemoglobin rates. The assessment cannot be biased anymore by wording of subjective feelings.

Conclusion. In this article, we have showed that a sting test could not be correlated enough with the expert evaluation of skin sensitivity. We introduced two groundbreaking indices. The first one is called clinical skin sensitivity index. It is based on the subject ’s self-perception transformed into a score. It should be used to have a quick and easy understanding of the subject ’s skin sensitivity level. The other one, the instrumental skin sensitivity index based on hyperspectral image analysis, is more complex. It predicts the sensitivity from parameters computed from hyperspectral acquisitions. It has the advantage to be completely objective and robust to small variations in the subject ’s position in the SpectraFace®. Both indices can be used in a complementary way.
In the next steps, a third clinical study will be conducted to increase the dataset and validate the index qualities in the context of a soothing product performance evaluation. A special care will be granted to self-perception repeatability to select the most reliable subjects.

Acknowledgments. The financial support of this project is shared by both involved organizations. Symrise initialized the research and brought its expertise about skin sensitivity. The Newtone internal R&D team led the technical construction of the indices, supported and advised about data acquisitions and wrote this article. We thank I.E.C and Dermatech for the data collection management. We also thank Charlotte Peucat from Newtone Technologies for having positioned all the ROIs and managed the parameter computations.

Conflict of Interest Statement. NONE.

References.

1. Grove GL, Soschin DM, Kligman AM (1984) Adverse subjective reactions to topical agents. Cutaneous Toxicity: 203–211.
2. Symrise CICS (Cosmetic Ingredients consumer study) database 2012 – 2018.
3. Morizot F, Guinot C, Lopez S, Fur I, Tschachler E, Wood C (2001) Sensitive skin: analysis of symptoms, perceived causes and possible mechanisms. Cosmet Toil; 115: 83–89.
4. Misery L, Loser K, Ständer S (2016) Sensitive skin. Journal of the European Academy of Dermatology and Venereology, 30:2-8.
5. Guinot C, Malvy D, Mauger E, Ezzedine K, Latreille J, Ambroisine L, Tenenhaus M, Préziosi P, Morizot F, Galan P, Hercberg S (2006) Self-reported skin sensitivity in a general adult population in France: data of the SU. VI. MAX cohort.Journal of the European Academy of Dermatology and Venereology 20.4: 380-390.
6. Prasad R, Verma SK, Dua R, Kant S, Kushwaha RA, Agarwal SP (2009) A study of skin sensitivity to various allergens by skin prick test in patients of nasobronchial allergy. Lung India: official organ of Indian Chest Society, 26(3):70.
7. Primavera G, Berardesca E (2005) Sensitive skin: mechanisms and diagnosis. International journal of cosmetic science, 27(1):1-0.
8. Muizzuddin N, Marenus KD, Maes DH (1998) Factors defining sensitive skin and its treatment. American Journal of Contact Dermatitis, 9(3):170-5.
9. Gevaux L, Gierschendorf J, Rengot J, Cherel M, Séroul P, Nkengne A, Robic J, Trémeau A, Hébert M (2021) Real-time skin chromophore estimation from hyperspectral images using a neural network. Skin Research and Technology, 27(2):163-77.
10. Seroul P, Hébert M, Jomier M (2014) Hyperspectral imaging system for in-vivo quantification of skin pigments. IFSCC.
11. Seroul P, Hébert M, Cherel M, Vernet R, Clerc R, Jomier M (2016) Model-based skin pigment cartography by high-resolution hyperspectral imaging. Journal of Imaging Science and Technology, 60(6):60404-1.
12. Nkengne A, Robic J, Seroul P, Gueheunneux S, Jomier M, Vie K (2018)

SpectraCam®: A new polarized hyperspectral imaging system for repeatable and reproducible in vivo skin quantification of melanin, total hemoglobin, and oxygen saturation. Skin Research and Technology, 24(1):99-107.

13. Weatherall IL, Coombs BD (1992) Skin color measurements in terms of CIELAB color space values. Journal of investigative dermatology, 99(4):468-73.

14. Murtagh F (1991) Multilayer perceptrons for classification and regression. Neurocomputing, 2(5-6):183-97.