How to reinforce the proficiency of SPF testing

Renoux, Pascale[1]; Clemente, Dáphine Antônia[2]; Piquemal, Pascale[1]; Pouradier, Florence[3*]

[1] L'Oréal Research & Innovation, Chevilly-Larue, France; [2] L'Oréal Research & Innovation, Rio de Janeiro, Brazil; [3] L'Oréal Research & Innovation, Clichy, France

* Florence, Pouradier, 9 rue Pierre Dreyfus 92110 Clichy, France, +33149180480, florence.pouradier@rd.loreal.com

Abstract

As photoprotection is about public health, sunscreen labelling must be sincere and reliable. Performance on UVB or UVA protection (expressed through SPF or UVA-PF) and Water-resistance is determined through standardized methods. Considering the importance of controlling the quality of the performed determination, all these standards have to use sunscreen formulation reference to validate the test, on each of the volunteers included in the clinical test. To reinforce this approach, we chose to also include regularly in our photoprotection studies well-known formulas, representative of marketed products.

Through this approach, we can collect a large amount of data, from different laboratories, worldwide. The objective of this study was to investigate how these data can be exploited, beyond the one-point approach test by test, to reinforce our global quality process and leverage knowledge on testing of the same formula across different world populations.

This approach allows us to control both the instantaneous performance and the continuous performance of the laboratory (quality assurance by different properties: bias, stability, and repeatability), ensuring the robustness of the SPF determination. The analysis of the resulting database highlights the consistency of the value obtained from different population across the world on which this standardized determination is performed.

Keywords: SPF; *in vivo*; Quality; Sincerity; Claim;

Background

The ultraviolet radiation from sunlight is a modifiable environmental hazard that has harmful effects on the skin, such as cancer, whose prevalence increases worldwide.

Consequently, sunscreens have been recommended as a form of protection against sunlight, with protection increasing with higher sun protection factor. As photoprotection is about public health, the sunscreen labelling must be sincere and reliable.

Several claims are generally used for a sun protection product such as the Sun Protection Factor (SPF), the UVA Protection Factor (UVA-PF) and the Water Resistance. The evaluation of these claims currently requires the use of standardized in vivo methods (e.g., ISO 24444 for SPF [1]) and seeks to measure the response of the human body after the sunscreen application, with exception for UVA-PF for which a standardized in vitro method exist.

Considering the importance to control the quality of the performed determination, all these standards recommend the use of sunscreen formulation standards to validate the test, on each of the volunteers included in the trial. As an interesting initiative, one of these reference sunscreens (P2) was used through meta-analysis to leverage knowledge on SPF method comparison or impact of the time of year, gender, age or Fitzpatrick's skin type of the volunteers on the measured SPF [2].

To continue our role of improving sunscreen testing confidence and to avoid gaps in research and knowledge regarding safety, efficacy, and public perception we choose to also include regularly in our photoprotection studies well-known formulas, representative of marketed products. These formulas are used as indicators to characterize the quality of the study and thus the reliability of the value obtained for the new formula evaluated in the same test. Through this approach, we can collect a large amount of data, from different labs, worldwide.

The objective of this paper is to describe how this data can be exploited, beyond the one-point approach test by test, to monitor the quality of our main studies and to reinforce our global quality process, as well as to leverage knowledge on testing the same formula across different world populations.

Methods

To ensure the reliability of the sun protection values of our main projects, and as part of the CROs quality control, we decided to include reference formulas representative of our catalogue in some in vivo SPF studies, besides the standards formulas recommended by the

specific norm. The 3 chosen formulas are representative from the market and present different level of protection.

These reference formulas have been more regularly included in our studies since 2019, providing a huge database. As part of the quality process, the application of these references was standardized and a unique batch that is used worldwide at the same time.

These tests are part of the quality process for the SPF testing laboratories that our Group developed and deployed [3]. Statistical analysis of this database is performed for each study and time by time, both to identify a criterion which reinforces the SPF determination and to evaluate the impact of the studied population on the obtained results.

Since the purpose of this paper is not to compare existing methods each other's, we filtered our database on the SPF tests performed according to the international standardized method ISO24444 (version 2010 and 2019) only.

Results

Thanks to our data base built over the years, several in vivo SPF evaluations were gathered for the 3 different sunscreen products that had been tested in different laboratories worldwide, as described in Table 1 below:

Table 1. Description of the database on the 3 reference formulas tested worldwide.

Labelled category	Data	Number of CROs	Countries
Medium Protection	227 individual data from 42 studies	7	Europe (4 countries) Canada Singapore
High protection	1279 individual data from 231 studies	15	Europe (7 countries) Canada Brazil China Japan Singapore
Very high protection	1931 individual data from 360 studies	12	Europe (7 countries) Canada Brazil China Japan Singapore

Using statistical methods this information is further condensed to assess the conformity of the study results against the internal standard criteria. For success criteria, inspired by BIPEA's approach and referring to the ISO 13528 standard [4], we used the "z-score" which is continuously updated by considering the global data from a formula, to calculate an accepted range. Z-scores are a way to compare results to a "normal" population, i.e., to the mean of a normally distributed random variables. It gives an idea of how far from the mean a data point is and is measured as a number of standard deviations from the mean. The z-score (z) is calculated from the laboratory result (x) based on the assigned value (x_{pt}) and half the tolerance value (VT/2): $z=(x-x_{pt})/(VT/2)$. ISO13528 standard defines that its absolute value greater than 2.0 is equivalent to a warning signal, while an absolute value greater than or equal to 3.0 is considered an action signal. The z-score revealed to be a relevant and useful criterion to assist the study sponsor in validating the reliability of the obtained results.

This analysis made available for assessment purposes, such as:

1) Alert of the necessity of a training of a specific CRO
2) Comparison of several laboratories' results, to evaluate their analytical performance on the same homogeneous samples.
3) To support a decision making with the desired confidence, given the quality of the data.
4) Determination if the data obtained from a specific study has the right quality to support the intended claim.

Internally, we have decided that when a study presents a z-score out of the acceptable limit, an investigation and exchange with the CRO in charge of this study is started. The main goal is to try to understand which parameter may have caused the unexpected result such as:

- How much equipment was used and whether it was calibrated.
- Who were the technicians who conducted the study and whether they are validated within the framework of our quality process.
- If an unexpected event occurred during the study
- And more globally how the CRO explained that result

If after this investigation, the CRO continues to present continuously too high absolute value for z-scores, we dismissed it temporarily. Then, new application and reading trials are started for each of the technicians and a new ring study is performed. If the results are acceptable, the CRO is reinstated. Example of such event can be visualized in Figure 1.

Sometimes it happened that we noticed a study was presenting an unacceptable z-score and at the same time the CRO informed us they have obtained an overestimated or underestimated result from recent BIPEA campaign. In that case the CRO is temporarily immediately dismissed, and an action plan is put in place to revalidate this CRO.

Additionally, by comparing the same formula tested worldwide across different populations, we can evaluate the impact of the geographical area of the SPF testing on the obtained results. The international standardized method ISO24444 aims to be used wherever the inclusion criteria for the test subjects can be meet (Fitzpatrick phototype = I, II or III in the previous ISO24444:2010 version, Individual Typology Angle ITA° >28° in the current version ISO24444:2019). With the usual ring tests, it is not so obvious to distinguish the effect of the testing laboratory to the effect of the tested population. With our database, even if some geographical areas are represented by one or two laboratory(ies) only, the repetition of the testing and the quantity of the data collected allow us to investigate the population effect.

For the 3 formulas, we can investigate the geographical area effect by comparing results from Canada, France and Romania, plus Brazil and Poland for formula “high” and “very high” protection. For the other countries, we do not have enough data from them to be relevantly included in the analysis. When it was sufficient to obtain enough data to be included in the analysis, we artificially merged China, Japan and Singapore together to build an “Asia” area. Other European countries are not included in a global “Europe” area because they will be completely hidden behind the numerous data from France, Poland and Romania. Future studies will allow to complete this analysis by country.

To avoid the results being biased by under- or over-estimated values, only the studies with a z-score < 2.0 were included in the analysis. Data are visualized in Boxplots and the geographical area effect is investigated by ANOVA and T-tests to estimate the p-value, plus Bayesian approach to estimate the effect-size.

For the “medium Protection” formula, the SPF value is (in average +/- SDT) 24 +/- 4. No statistical difference is observed between Canada (North America), France and Romania (Europe) (Figure 2).

For the “high protection” formula, the SPF value is (in average +/- SDT) 31 +/- 4. No statistical difference is observed between Asia, Brazil (South America), Canada (North America), France, Poland and Romania (Europe) (Figure 3).

For the “very high protection” formula the SPF value is (in average +/- STD) 69 +/- 6. We observed a higher SPF value in Canada (North America) than in Asia, Brazil (South America) and Romania (Europe) for this formula (with a moderate effect-size) (Figure 4). We can hypothesis it may be an artifact due to the unbalanced sampling across the countries in our database (n=10 for Canada vs n=125 in Romania) as extreme average results are more prone to be observed in small sampling. Moreover, the estimated difference is in the same range than the variability of the method for this high level of SPF (maximum estimated difference is observed between Canada and Romania Δ=-8.43, corresponding to 12% of the mean). Thus, we hypothesize this difference not being clinically relevant. Our database should be reinforced to be able to confirm or reject this hypothesis.

Conclusion

The sun protection field is continuously moving for higher and higher reliability. However, it does not replace internal quality control or an assessment of compliance with standards. Moreover, data quality assessment is an important part of the overall quality management system.

The approach described in this paper, allows us to control both the instantaneous performance and the continuous performance of the laboratory, ensuring the robustness of the SPF determination. The analysis of the resulting database highlights the consistency of the value obtained from different population across the world on which this standardized determination is performed.

Acknowledgements: The authors express their sincere gratitude to Hicham Nocairi for the statistical support, and Remi Moyenin for the support in data management, as well as all the study managers in the L’Oréal R&I hubs worldwide who help collecting these data.

Conflict of Interest Statement: None

References:

1. ISO 24444: 2019 Cosmetics–Sun Protection Test Methods – In vivo Determination of the Sun Protection Factor (SPF)

2. Alejandria, M., et al. (2019). "Disparate SPF Testing Methodologies Generate Similar SPFs. II. Analysis of P2 Standard Control SPF Data."

3. Piquemal, P., et al. (2022). “Sincere and reliable SPF: a quality process that can be deployed worldwide.” Poster. IFSCC congress, London 2022.

4. ISO13528:2015 Statistical methods for use in proficiency testing by interlaboratory comparison

Figure 1. Evolution of the z-score in the last 3 years for one of the testing laboratory, and the "high protection" formula. This example illustrates a period of over-estimation which observed clearly via a z-score > 2.0 for 5 successive studies (January-May 2020). It was followed by investigation and re-validation (June 2020) to reinstate the CRO (from September 2020) while monitoring that the z-score level remained in the acceptance range for the subsequent studies.

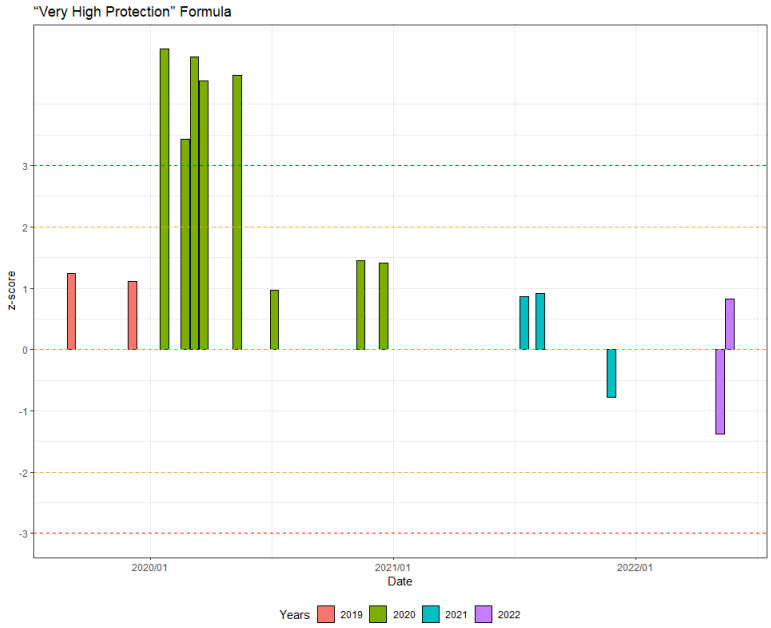

Figure 2. Effect of the geographical area of testing for the "medium protection" formula. No significant difference between the 3 studied countries.

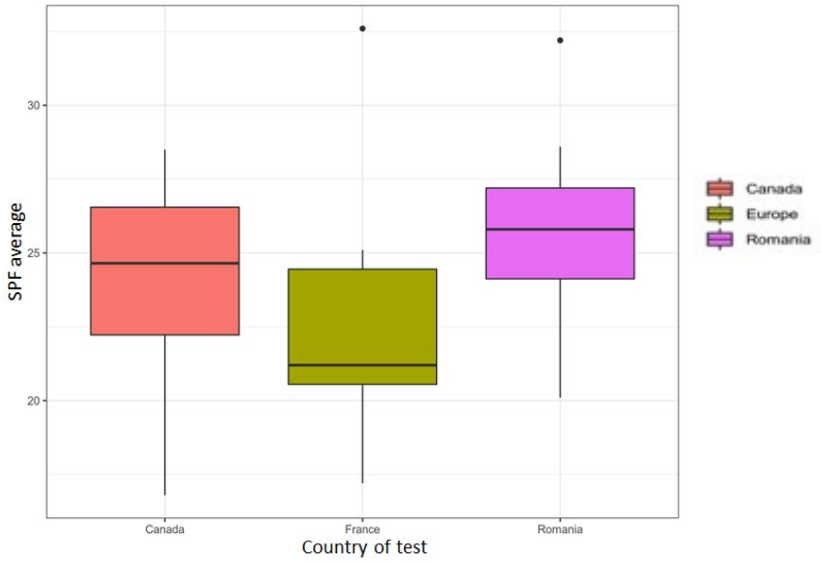

Figure 3. Effect of the geographical area of testing for the “high protection” formula. “Asia” is obtained by merging data from China, Japan and Singapore. “Europe” is obtained by merging data from France, Germany and Italy, while Poland and Romania data are kept separate to avoid the dataset being too much unbalanced. No significant difference between the studied countries.

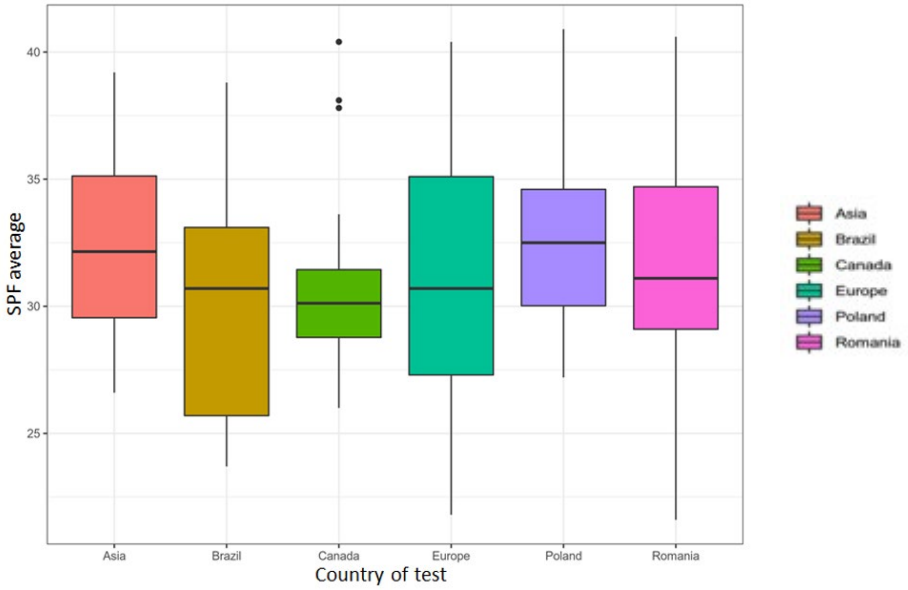

Figure 4. Effect of the geographical area of testing for the “very high protection” formula. “Asia” is obtained by merging data from China, Japan and Singapore. “Europe” is obtained by merging data from France, Germany and Italy, while Poland and Romania data are kept separate to avoid the dataset being too much unbalanced. Significant higher SPF is obtained in Canada vs Asia, Brazil and Romania with a moderate effect-size. Clinical relevancy of this difference should be further investigated.