J. Cosmet. Sci., **70,** 181–196 (July/August 2019)

Disparate SPF Testing Methodologies Generate Similar SPFs. II. Analysis of P2 Standard Control SPF Data

MELANNIE ALEJANDRIA, ANDREW MARRA, GLENN ROBERTS, and MICHAEL CASWELL, *Consumer Product Testing Company, Inc., Fairfield, NJ (M.A., A.M., G.R., M.C.)*

Accepted for publication May 26, 2019.

Synopsis

In the original scientific publication evaluating sunscreen methodologies, Garzarella and Caswell showed there to be no clinically significant or statistically significant difference in the average Sun Protection Factor (SPF) of a sunscreen formulation between any of three methodologies, Food and Drug Administration (FDA) Final Monograph, Australia/New Zealand, and European Cosmetics Association (COLIPA) International, suggesting that any differences in methodology were insignificant in the resulting SPF determined. These three major older methodologies have coalesced into two methodologies, 2011 FDA-Final Rule and ISO 24444, so that current sunscreen SPF testing is mostly 2011 FDA-Final Rule and ISO 24444. Another approach to evaluating the impact of methodological differences in sunscreen testing is to compare data on a control standard or reference sunscreen. If the difference between the two SPF values of P2 is statistically significant for the two different methodologies, then this would present evidence for a clinically significant difference in the SPF value between the two methodologies. For 2011 FDA-Final Rule, the expected SPF of P2 is 16.3 ± 3.43; for ISO 24444, the expected SPF of P2 is 16.1 ± 2.42. Using least squares average and standard error on 952 observations, the 2011 FDA-Final Rule SPF of P2 is 15.4 ± 0.12; using least squares average and standard error on 1,551 observations, the ISO 24444 SPF of P2 is 15.6 ± 0.10. The data described herein indicate no clinically significant nor statistically significant difference between the SPF average of P2 using the 2011 FDA-Final Rule methodology versus that using ISO 24444 methodology. Further statistical analysis indicates that the average SPF of P2 is independent of solar simulator type, time of year (month), age of subject, gender of subject, or Fitzpatrick Skin Phototype of subject. A statistically significant negative correlation was found between a subject's SPF of P2 and the subject's unprotected minimal erythemal dose. The implications of this relationship on SPF testing are explored.

INTRODUCTION

The three major older methodologies for sunscreen testing, Food and Drug Administration-Final Monograph (FDA-FM) method (1,2), Australia/New Zealand (Aus/NZ) method (3), and COLIPA International (International) method (4,5) have coalesced into two methodologies, 2011 FDA-Final Rule (6) and ISO 24444 (7). Examining data from sunscreen testing, Garzarella and Caswell (8) showed there to be no significant difference in the

Address all correspondence to Michael Caswell at mcaswell@cptclabs.com

average Sun Protection Factor (SPF) of a sunscreen formulation between any of the three older methodologies. Their data suggested that any differences in methodology were insignificant in the resulting SPF determined.

Another approach to evaluating the impact of methodological differences in sunscreen testing is to compare data on a control standard or reference sunscreen. A control standard, P2, P3, or P7, must be run concurrently on every subject undergoing SPF testing, regardless of which method is being used. P2, with actives of 7% ethylhexyldimethyl PABA (Padimate O) and 3% benzophenone-3, is a control standard for 2011 FDA-Final Rule (6) and for ISO 24444 (7). Although the 2011 FDA-Final Rule does not refer to this control standard as P2, for simplicity in this document, this control standard will be referred to as P2. For 2011 FDA-Final Rule, the expected SPF of P2 is 16.3 ± 3.43 (6); for ISO 24444, the expected SPF of P2 is 16.1 ± 2.42 (7) (Table I). If the difference between the two SPF values of P2 is statistically significant for the two different methodologies, then this would present evidence for a clinically significant difference in the SPF value between the two methodologies.

The data described herein indicate no statistically significant difference between the SPF average of P2 using the 2011 FDA-Final Rule methodology versus that using ISO 24444 methodology. Further statistical analysis indicates that the average SPF of P2 is independent of solar simulator type, time of year (month), age of subject, gender of subject, or Fitzpatrick Skin Phototype of subject. However, consistent with the hypothesis of Damian et al. (9), the data presented herein clearly show a statistically significant negative correlation between a subject's SPF of P2 and the subject's unprotected Minimal Erythemal Dose (MED). As a subject's unprotected MED decreases, the subject's SPF of P2 increases. The implications of this relationship on SPF testing are explored.

METHODS

The standard control, P2, was obtained from Cosmetech Laboratories, Inc., Fairfield, NJ. The lot numbers used were 1902D, 1902E, 1902F, 1902G, 1902H, 1902I, and 1902J.

The clinical trials were approved by Allendale IRB, Old Lyme, CT, Dr. Robert Staab, Chairperson. The trials were conducted at Consumer Product Testing Company, Inc., located in Fairfield, NJ, according to the principals of the Declaration of Helsinki (as amended) (10) and the Belmont Report (11) and according to ICH GCP and the Standard Operating Procedures at Consumer Product Testing Company, Inc. Each potential subject was apprised of the risks and benefits of the research clinical trial before conducting any procedures. Once the potential subject consented to participate, their consent was captured through the execution of an Informed Consent Document by the potential subject

Table I
Expected SPF Values of the Control Standard for ISO 24444 and for 2011 FDA-Final Rule

			Reported Herein			
Method	Mean	Standard deviation	Least square average	Standard error	Mean	Standard deviation
ISO 24444	16.1	2.42	15.6	0.10	15.7	2.40
2011 FDA	16.3	3.43	15.4	0.12	15.4	2.57

followed by its execution by a staff member. The subject was given a copy of the fully executed Informed Consent Document.

ULTRAVIOLET RADIATION (UVR) SOURCE

Xenon Arc Solar Simulators from Solar Light Company, Philadelphia, PA (150 or 300 W), were used as the source of UVR (12). The spectral output for the 150 and the 300 W were essentially identical (13,14). The lamp output was measured with a UV intensity meter (Model PMA2100, Solar Light Company) thrice daily. Solar simulators were equipped with 1-mm UG11 and WG320 filters, providing a spectral output in the ultraviolet range (290–400 nm) comparable with that of natural sunlight and meeting both 2011 FDA-Final Rule and ISO 24444 standards (see example spectral output for single-port solar simulator (Figure 1 and Table II) and for multiport with six light guides (Figure 2 and Table III). Irradiation beams were a minimum of 1 cm^2 with a beam uniformity of 10%, and they exhibited less than 20% time-related fluctuation. All solar simulators were calibrated and adjusted to deliver energies within 10% variance.

SPF DETERMINATION

An MED is defined by ISO 24444 (7) as "the lowest dose of UVR that produces the first perceptible unambiguous erythema with defined borders appearing over most of the field of UV exposure, 16 to 24 h after UV exposure." An MED is defined by 2011 Final Rule (6) as "The quantity of erythema-effective energy (expressed as Joules per square meter)

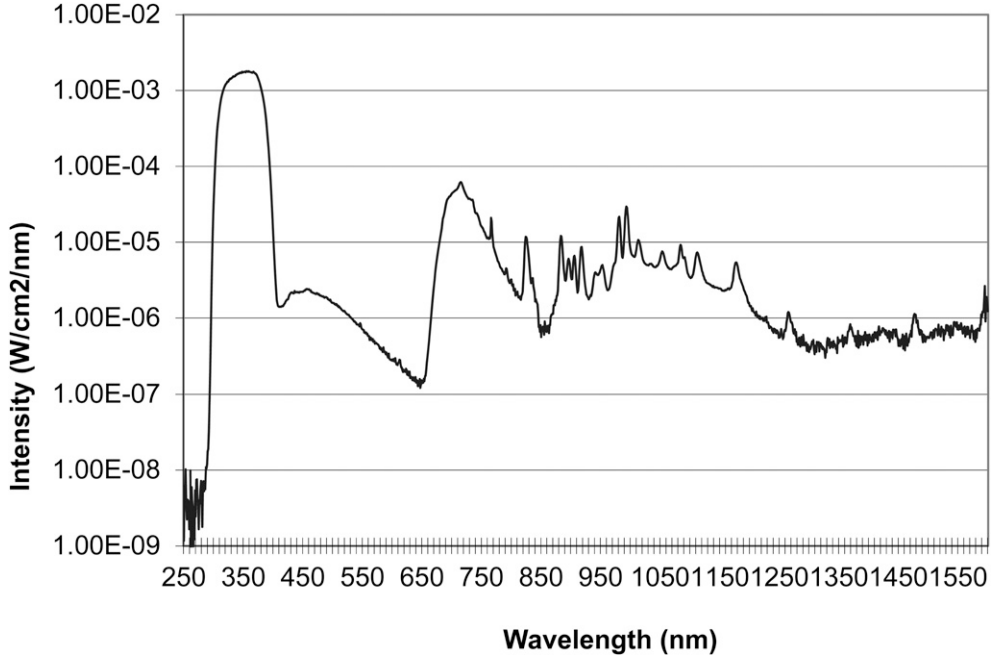

Figure 1. Typical spectral output of a 150-W xenon arc single-port solar simulator.

Table II
Typical Spectral Output of a 150 W Xenon Arc Single-Port Solar Simulator

Total irradiance (250–1,600 nm)	1.20E-01	W/cm^2
UVC irradiance (250–290 nm)	2.02E-07	W/cm^2
UVB irradiance (290–320 nm)	1.18E-02	W/cm^2
UVA irradiance (320–400 nm)	1.02E-01	W/cm^2
UVA2 irradiance (320–340 nm)	2.88E-02	W/cm^2
UVA1 irradiance (340–400 nm)	7.36E-02	W/cm^2
Visible + NIR irradiance (400–1,600 nm)	5.55E-03	W/cm^2
%UVC	0.0002%	–
%UVB	9.81%	–
%UVA	85.55%	–
%Visible + NIR	4.63%	–
SED	12.2	s
Erythemal effective irradiance	8.21E-04	W/cm^2
% Erythemal contribution of FDA (June 2011)/ISO 24444 SPF method (November 2010)		
<290 nm (<0.1%)	0.02%	
290–300 nm (1.0–8.0%)	4.77%	
290–310 nm (49.0–65.0%)	57.71%	
290–320 nm (85.0–90.0%)	87.74%	
290–330 nm (91.5–95.5%)	93.32%	
290–340 nm (94.0–97.0%)	95.48%	
290–400 nm (99.9–100.0%)	99.98%	
%UVA2/Total UV (≥20%) (320–340 nm/290–400 nm)	25.25%	
%UVA1/Total UV (≥60%) (340–400 nm/290–400 nm)	64.45%	
Total irradiance 250–1,400 nm for FDA (<1,500 W/m^2)	1,196	
Total irradiance 250–1,500 nm for ISO (<1,600 W/m^2)	1,197	

required to produce the first perceptible, redness reaction with clearly defined borders." Although these two definitions vary slightly, the implementation of the definitions is essentially identical. Each evaluator of erythema was qualified through training and evaluation testing.

A subject's SPFi, defined as the ratio of the MED on protected skin (MEDp) to the MED on unprotected skin (MEDu), was calculated for each subject as follows:

$$\mathrm{SPFi} = \frac{\mathrm{MEDi}\left(\text{protected skin}\right)}{\mathrm{MEDi}\left(\text{unprotected skin}\right)} = \frac{\mathrm{MEDpi}}{\mathrm{MEDui}}$$

SUMMARY OF DATA

A total of 2,607 observations encompassing 664 subjects were collected from February 2016 to September 2017. Of these observations, 104 observations (3.99%) on 86 subjects were collected from clinical trials labeled as invalid. An invalid observation occurs when all test sites exhibit an erythema score of at least 1, no test sites exhibit an erythema score of 1, or the erythema scores do not follow the irradiation sequence (2011 FDA-Final Rule method only). These 104 observations were not included in the data analysis. The resulting sample consisted of 2,503 valid observations encompassing 652 subjects. The arithmetic average SPF value for all 2,503 observations was 15.6 ± 2.5, before stratifying by the method used and before incorporating any statistical model of the data. Within the

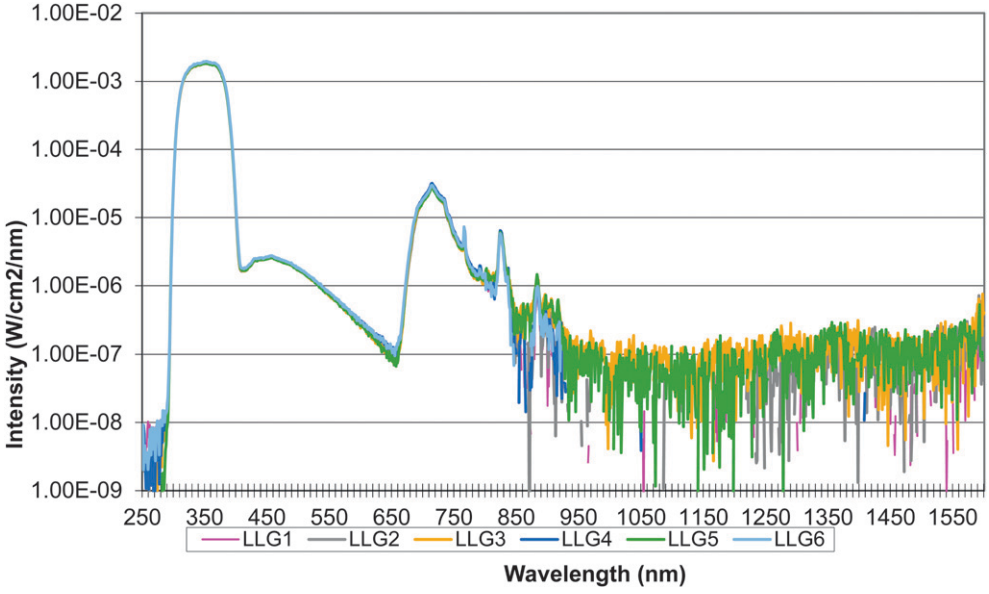

Figure 2. Typical spectral output of six light guides from a 300-W xenon arc single-port solar simulator.

2,503 observations, 952 (38.03%) of the observations were evaluated using 2011 FDA-Final Rule methodology, whereas the remaining 1,551 (61.97%) observations were evaluated using ISO 24444 methodology. The distribution of all 2,503 observations involved each calendar month of the year, with the most observations collected in June (288, 11.51%) and the fewest observations collected in January (101, 4.04%). Within the 2,503 observations, 291 (11.63%) observations were evaluated using single-port solar simulators, whereas 2,212 (88.37%) observations were evaluated using multiport solar simulators. Of the 652 subjects, 470 (72.09%) were female and 182 (27.91%) were male. The distribution of subjects based on their Fitzpatrick Skin Phototype was 44 (6.75%), 377 (57.82%), and 231 (35.43%) for Skin Phototypes I, II, and III, respectively (15,16). Ages of the subjects ranged from 18 to 70 years, with a median age of 51 years and a mean age of 47.6 years.

STATISTICAL METHODS

To determine any possible factors that may impact an observation's SPF value, a linear mixed-effects model using restricted maximum likelihood estimation was created. Fixed-effect predictors included the standard protocol used during the evaluation (i.e., the method used): the month when an observation was evaluated, the type of solar simulator used, the subject's Fitzpatrick Skin Phototype, age, gender, and unprotected MED value. The response variable was the observed SPF value. To address the within-subjects variability of the data, a random effect was assigned for each subject. As an assessment of the model's predictive capability, the coefficient of determination [(pseudo)-R^2 value] was calculated using the methodology proposed by Nakagawa and Schielzeth (17). For any pairwise hypothesis tests, statistical significance was achieved at the 95% confidence level ($p < 0.050$) using the t-distribution, under the assumption that the distribution of the

Table III
Typical Spectral Output of Six Light Guides from a 300-W Xenon Arc Multiport Solar Simulator

Total irradiance (250–1,600 nm)	1.22E-01	1.21E-01	1.20E-01	1.26E-01	1.20E-01	1.27E-01	W/cm^2
UVC irradiance (250–290 nm)	3.15E-07	1.96E-07	1.15E-07	1.91E-07	2.79E-08	2.86E-07	W/cm^2
UVB irradiance (290–320 nm)	1.24E-02	1.24E-02	1.21E-02	1.25E-02	1.24E-02	1.25E-02	W/cm^2
UVA irradiance (320–400 nm)	1.08E-01	1.07E-01	1.06E-01	1.11E-01	1.06E-01	1.12E-01	W/cm^2
UVA2 irradiance (320–340 nm)	3.20E-02	3.18E-02	3.12E-02	3.25E-02	3.16E-02	3.29E-02	W/cm^2
UVA1 irradiance (340–400 nm)	7.60E-02	7.51E-02	7.51E-02	7.89E-02	7.46E-02	7.94E-02	W/cm^2
Visible + NIR irradiance (400–1,600 nm)	1.64E-03	1.67E-03	1.76E-03	1.87E-03	1.76E-03	1.79E-03	W/cm^2
%UVC	0.0003%	0.0002%	0.0001%	0.0002%	0.0000%	0.0002%	
%UVB	10.13%	10.23%	10.08%	9.93%	10.28%	9.90%	
%UVA	88.52%	88.39%	88.46%	88.58%	88.26%	88.68%	
%Visible + NIR	1.35%	1.38%	1.46%	1.49%	1.46%	1.42%	
SED	11.9	11.9	12.2	11.8	11.7	11.8	s
Erythemal effective irradiance	8.38E-04	8.42E-04	8.22E-04	8.48E-04	8.52E-04	8.48E-04	W/cm^2
% Erythemal contribution of FDA (June 2011)/ISO 24444 SPF method (November 2010)							
<290nm (<0.1%)	0.04%	0.02%	0.01%	0.02%	0.00%	0.03%	
290–300 nm (1.0–8.0%)	4.45%	4.65%	4.57%	4.52%	4.83%	4.49%	
290–310 nm (49.0–65.0%)	55.88%	56.14%	56.01%	55.85%	56.84%	55.62%	
290–320 nm (85.0–90.0%)	87.03%	87.15%	87.08%	86.92%	87.46%	86.79%	
290–330 nm (91.5–95.5%)	93.02%	93.15%	93.04%	92.91%	93.28%	92.84%	
290–340 nm (94.0–97.0%)	95.39%	95.46%	95.39%	95.29%	95.57%	95.25%	
290–400 nm (99.9–100.0%)	99.96%	99.98%	99.99%	99.98%	100.00%	99.97%	
%UVA2/Total UV (≥20%) (320–340 nm/290–400 nm)	26.59%	26.63%	26.37%	26.23%	26.63%	26.34%	
%UVA1/Total UV (≥60%) (340–400 nm/290–400 nm)	63.14%	63.00%	63.41%	63.69%	62.93%	63.62%	
Total irradiance 250–1,400 nm for FDA (<1,500 W/m^2)	1,220	1,209	1,201	1,258	1,202	1,266	
Total irradiance 250–1,500 nm for ISO (<1,600 W/m^2)	1,220	1,209	1,201	1,258	1,203	1,266	

sample mean followed a normal distribution. The *p*-values for fixed effects were determined using an Analysis of variance (ANOVA) of Type III Sums of Squares, with an adjustment to the denominator degrees of freedom using the Satterthwaite approximation. To counteract the possibility of increased Type I Error due to multiple hypothesis tests, the maximum expected proportion of false discoveries among the rejected hypotheses for all pairwise tests was maintained at 5.0% using the Benjamini–Hochberg procedure.

STATISTICAL SOFTWARE

The statistical software R (version 3.2.2 for Microsoft Windows; R Foundation for Statistical Computing, Vienna, Austria) was used for all data analyses (18). In addition to the base package preinstalled with the software, the packages "ImerTest" (19) and "ggplot2" (20) were also used for linear mixed-effects model analysis and graphical plots, respectively.

STATISTICAL ASSUMPTIONS

A linear mixed-effects model requires certain statistical assumptions concerning the data. Because both inference and point estimation/prediction were performed on the data, two major assumptions were tested:

- Homoscedasticity–the population variance between each category/group was approximately equal. Although linear mixed-effects models do not require this assumption to determine point estimates or to predict effects, it is required if inference is made. Since ANOVA was performed and standard errors were calculated, testing of the equal-variance assumption was necessary.
- Normality of the Residual Distribution–the distribution of the residual deviation between the observed values and the values predicted by the model was approximately normal. This applies to both the fixed effects and the random effects. Major deviations will produce erroneous standard errors and confidence intervals. For this reason, testing the normality of the residuals was performed. For mixed-effects models, if a violation to this assumption were to occur, a data transformation can be performed to mitigate the impact of the violation. However, mixed-effects models are robust against violations to this assumption, especially when the sample size is large.

The diagnostic tests indicated heteroscedasticity in the categorical predictors, particularly for the month in which an observation was evaluated (Levene's Median Test, $p = 7.335e\text{-}06$). This was expected, as the sample sizes of the observations varied across different months of the year, and homogeneity of the variance is a function of the sample size in a group. To investigate the magnitude of the equal-variance violation, box plots were produced for every categorical predictor. Upon further investigation, the box plots revealed that although heteroscedasticity was present, any violations to the homogeneity of the variance were minor in magnitude. In addition, as stated in the "Statistical Methods" section, the ANOVA of the data was conducted with an adjustment to the denominator degrees of freedom using the Satterthwaite approximation. This approximation is used whenever an assumption of equal variances cannot be made. Therefore, it can be argued that there was no adverse impact to the integrity of the statistical model, along with any inferences made from it.

Testing of the residual distribution revealed deviations in the normality assumption, for both the fixed effects and the random effects (Shapiro–Wilk's W Test, $p = 2.2e\text{-}16$ and $p = 1.306e\text{-}11$, respectively). Visual inspection through quantile-quantile plots also revealed left and right skewness for both types of effects (i.e., fat tails), respectively. However, it was determined that any violations to the normality assumption would also be minor in magnitude, mainly because of the large sample size of the dataset. In addition, Box-Cox Power Transformation Analysis was performed to determine if data transformation would mitigate the violation, and the analysis did not provide any recommended transformation (maximum lambda value of 1.23, [95% CI: (1.15, 1.35)]). For these reasons, it can be argued that using the original data will not adversely impact the integrity of the statistical model.

In addition to testing the above assumptions, sensitivity analysis was performed to determine the presence of collinear predictors, influential observations, and potential outliers in the data. The analyses concluded no presence of collinear predictors, no influential observations, and no presence of any potential outliers. As a result, all of the data were included in the analysis, and the statistical assumptions of the model were verified.

RESULTS

To determine how well the linear mixed-effects model predicted the data, a conditional coefficient of determination [(pseudo)-R^2 value] was calculated. The (pseudo)-R^2 value was calculated by incorporating the variance of both the fixed and random effects in the model. The results indicated that the model accounted for approximately 23% of the variance in the data [Conditional (pseudo)-$R^2 = 0.227$].

ANOVA of the linear mixed-effects model revealed a statistically significant impact on SPF values for P2 for time/month, the subject's Fitzpatrick Skin Phototype, and the observation's unprotected MED value (p-values of 0.003, 0.026, and 2.2e-16, respectively). There was no statistically significant effect when accounting for the methodology used, the type of solar simulator used, age of subject, or gender of subject (p-values of 0.068, 0.373, 0.126, and 0.657, respectively).

Further evaluation of each observation's unprotected MED value revealed a statistically significant negative correlation with the resulting SPF value (Pearson's product-moment correlation = −0.409, 95% CI: (−0.441, −0.376), $p < 2.2e\text{-}16$). For every 10 mJ/cm^2 increase in the unprotected MED, the predicted SPF value of an observation decreased by 1.55 [95% CI: (−1.70, −1.40)], when controlling for all other predictors. This linear trend can be seen in Figure 3.

Estimated population SPF values among each categorical predictor are presented in Figure 4. Observations evaluated in December yielded statistically significant greater SPF values when compared with those observations evaluated from January, February, August, or September, with the greatest change in SPF values occurring between December and January [mean difference = −0.904, 95% CI: (−1.436, −0.371)].

Although the ANOVA indicated a statistically significant impact for Fitzpatrick Skin Phototype ($p = 0.026$), pairwise comparisons revealed no statistically significant differences in mean SPF values among the three Skin Phototype groups. Pairwise comparison of Phototype I with Phototype II gave $p = 0.994$, of Phototype I with Phototype III gave $p = 0.312$, and Phototype II with Phototype III gave $p = 0.069$.

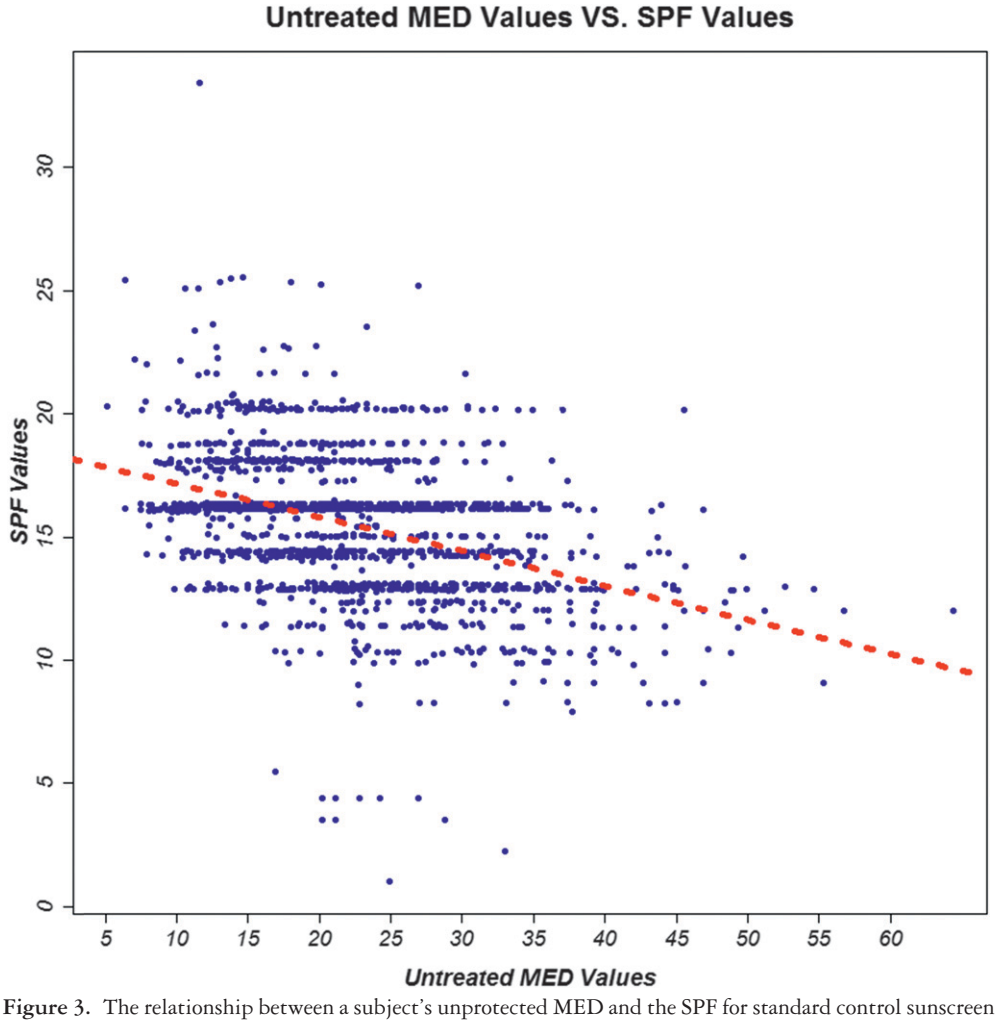

Figure 3. The relationship between a subject's unprotected MED and the SPF for standard control sunscreen P2, when controlling for all other predictors. The dashed red line represents a line with a y intercept of 18.579 and a slope of −0.155. The dashed red line has a Pearson's product-moment correlation of −0.409, with approximately 23% of the variance of the data explained by the linear mixed-effects model (Conditional (pseudo)-$R^2 = 0.227$).

DISCUSSION

This is the first report of an analysis of a large number of subjects, each tested with the same sunscreen. Although these data can suggest which factors are important to control in SPF testing, the data come from only one testing facility, Consumer Product Testing Company, Inc. The value for control standard from other testing facilities may differ slightly from that reported herein.

METHODOLOGY

No clinically significant or statistically significant SPF effect on P2 was found between the 2011 FDA-Final Rule methodology (952 observations) and the ISO 24444 methodology

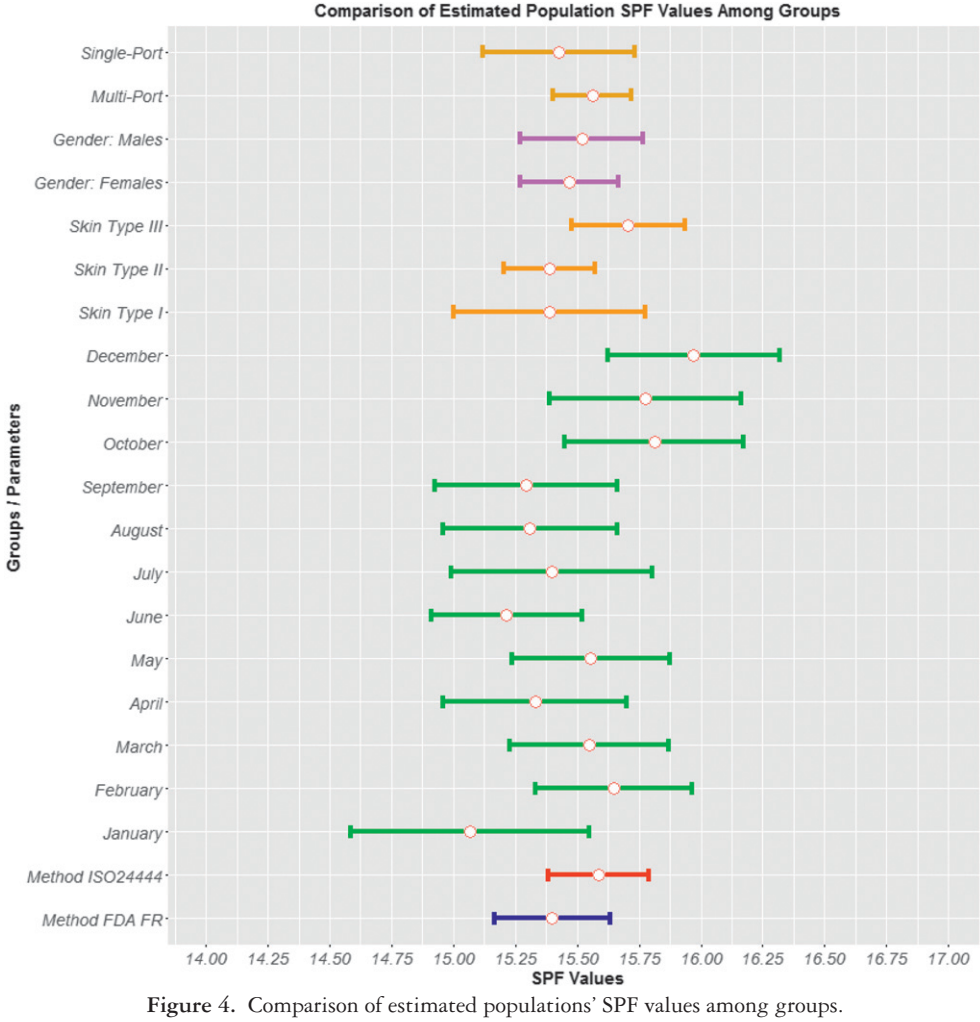

Figure 4. Comparison of estimated populations' SPF values among groups.

(1,551 observations) (Table I). For 2011 FDA-Final Rule, the expected SPF of P2 is 16.3 ± 3.43 and for ISO 24444, the expected SPF of P2 is 16.1 ± 2.42. The least square average SPF for P2 using the 2011 FDA-Final Rule methodology was 15.4 (standard error = 0.12) and that using the ISO 24444 methodology was 15.6 (standard error = 0.10). Unlike a simple arithmetic average of the data using no statistical model, the least square average incorporates the linear mixed-effects statistical model used prior and adjusts the estimated average based on covariates present inside the model. For this reason, the least square average with a standard error, rather than the simple unadjusted arithmetic average with standard deviation is a more appropriate estimate of the population SPF values. Nonetheless, the unadjusted arithmetic average SPF for P2 using the 2011 FDA-Final Rule was 15.4 (standard deviation = 2.57) and that using the ISO 24444 methodology was 15.7 (standard deviation = 2.40). This lack of difference between the 2011 FDA-Final Rule method and the ISO 24444 method is consistent with the earlier findings of Garzarella and Caswell (2013) between the FDA-FM (1,2), Aus/NZ (3), and International (4,5)

methodologies. There seems to be no clinically significant difference or statistically significant difference between the average SPF of P2 generated by different methodologies (Table IV), when examined by two different approaches. This suggests that slightly alternative methodologies may result in similar SPF values.

The SPF label value will differ between ISO 24444 and 2011 FDA-Final Rule because the 2011 FDA-Final Rule methodology, unlike the ISO 24444 methodology, subtracts the "*A*" value from the average SPF. The SPF label value then becomes the next lower integer after subtraction (6). The *A* value is the product of the upper 5% point of the t-distribution and the standard deviation, divided by $\sqrt{(n)}$, where n equals the number of subjects with valid data (minimum 10). This subtraction reduces the label SPF value to an integer that would, except under unusual circumstances, be different from the SPF value determined by ISO 24444, resulting in identical formulations labeled with different SPF values.

AGE AND GENDER OF SUBJECT

ANOVA revealed no statistically significant effect of age ($p = 0.126$) or gender ($p = 0.657$) on the SPF of P2. As a result, it cannot be ruled out that any effects on the SPF of P2 are likely due to chance alone. This lack of age effect on the SPF of P2 fails to support the age restrictions placed on subjects by 2011 FDA-Final Rule and by ISO 24444.

TYPE OF SOLAR SIMULATOR

There are two basic types of solar simulators (Solar Light Company) used in SPF testing (12), single-port solar simulators (150 and 300 W) and multiport solar simulators (150 and 300 W). ANOVA of the data for 291 observations evaluated following use of single-port solar simulators versus 2,212 observations evaluated following use of multiport solar simulators indicated no statistically significant differences ($p = 0.373$) in the SPF of P2. This is consistent with reciprocity holding for the wattage (13,14) and for the type of solar simulator. The type of solar simulator (Tables II and III) seems to have no significant effect on the SPF of P2 (13,14).

FITZPATRICK SKIN PHOTOTYPE

Fitzpatrick Skin Phototype was created to assist in the prediction of MED for phototherapy in a physician's office (15,16). Phototype was determined using either a subjective assessment based on the patient's phenotype (hair color, eye color, etc.) (15) and later on the subject's recollection of his burning and tanning response to sun exposure (16). Because these two different subjective methods were proposed to determine Fitzpatrick Skin Phototype, conflicting Fitzpatrick Skin Phototypes can be generated for the same person. Despite its inability to predict MED (21) and increasing limited relevance (22), Fitzpatrick Skin Phototype use has been expanded over the past 40 + years. Recently, however, Individual Typology Angle (ITA) (23) has been found to be a better predictor of MED and is included in ISO 24444 as an alternative to Fitzpatrick Skin Phototype for subject qualification.

Table IV
Comparison of SPF Testing Parameters ISO 24444 versus 2011 FDA-Final Rule

Parameters	ISO 24444		2011 FDA-Final Rule	
Source of UVR				
	% RCEE defined in different bands			
	λ range (nm)	RCEE (%)	λ range (nm)	Erythemal effective radiation (%)
Acceptance limits	≤290	<0.1%	≤290	<0.1%
%RCEE UVAII/	290–300	1.0–8.0	290–300	1.0–8.0
UVAI	290–310	49.0–65.0	290–310	49.0–65.0
	290–320	85.0–90.0–	290–320	85.0–90.0
	290–330	91.5–95.5	290–330	91.5–95.5
	290–340	94.0–97.0	290–340	94.0–97.0
	290–400	99.9–100	290–400	99.9–100
	UVAII 20% UVAI 60% of the total UV irradiance to ensure that appropriate amounts of UVA radiation are included		UVAII 20% UVAI 60% of the total UV irradiance to ensure that appropriate amounts of UVA radiation are included	
Sunscreen application				
Amount	2.00 ± 0.05 mg/cm^2		2 mg/cm^2	
Preparation	Test area may be cleaned with a dry cotton pad or equivalent		None	
Conditions	22 ± 4°C		None	
Application	Finger cot is optional		Finger cot required	
Spreading time	35 ± 15 s		None	
Dry time	15–30 min		≥15 min	

Table IV
Continued

Parameters	ISO 24444	2011 FDA-Final Rule
UV Exposures		
Progression of UV Dose	For the unprotected area and the protected areas, a minimum of five subsites centered on the expected SPF × MEDu shall be exposed with a maximum geometric progression of 1.25^n. The progression must be identical for the unprotected and protected areas. A maximum progression of 1.12^n must be used for expected SPF >25	Geometric progression (1.25^n) for the unprotected area For the protected areas, geometric series of five exposure where the middle exposure is placed to yield the expected SPF plus two other exposures placed around the middle exposure According the expected SPF (X) SPF <8: 0.64X, 0.80X, 1.00X, 1.25X, 1.56X SPF 8 to 15: 0.69X, 0.83X, 1.00X, 1.20X, 1.44X SPF >15: 0.76X, 0.87X, 1.00X, 1.15X, 1.32X
Reference sunscreen formulations		
Reference sunscreen formulations used	Expected SPF <20: P2 or P3 or P7 Expected SPF ≥20: P2 or P3 The same reference has to be tested on every subject in the same series of at least 10 subjects	P2 (Padimate O 7.0% + Oxybenzone 3.0%) with SPF 16.3 (SD: 3.43)
Calculations and results		
Number of test subjects	Minimum of 10, maximum of 25 Invalid data from up to five subjects are acceptable	At least 10 valid; no maximum Invalid data from up to three subjects is acceptable
Statistical criterion	95% confidence interval should fall within ±17.0% of mean SPF	None

First, the 2011 FDA-Final Rule requires that the subjects qualify based on the Fitzpatrick Skin Phototype using the subject's recollection of his burning/tanning response. Second, 2011 FDA-Final Rule requires that the subjects in a valid SPF test not be of all the same skin phototype. Although the ANOVA indicated a statistically significant impact for Fitzpatrick Skin Phototype, pairwise comparisons revealed no statistically significant differences in mean SPF values between the three Skin Phototype groups, I, II, and III. The data presented herein fail to support the 2011 FDA-Final Rule requirement that the subjects in a valid SPF test not be of all the same skin phototype. Fitzpatrick Skin Phototype has no significant effect on the SPF of P2.

ISO 24444 requires that the subjects qualify based on Fitzpatrick Skin Phototype using subjective assessment or have an ITA of greater than 28° (7). The data presented herein support the abandonment of Fitzpatrick Skin Phototype as a qualification for subjects in SPF testing, in favor of ITA° because ITA° is a much better predictor of unprotected MED (23).

SEASONAL VARIATION

During summer months exposure to UVR induces an increase in the MED of a population (24). One might, therefore, expect to observe a change in the SPF of the P2 standard during the course of the year. This was not found. ANOVA of the linear mixed-effects model revealed a statistically significant impact on SPF values for December. Pairwise comparisons yielded statistically significant different SPF values for P2 only between January and December, between June and December, between August and December, and between September and December.

UNPROTECTED MED

The unprotected MED value revealed a statistically significant negative correlation with the resulting SPF value (Figure 3). This is consistent with the hypothesis of Damian et al. (9), whose data indicated statistically significant higher SPF values in sunscreen tested on subjects with lower unprotected MEDs. The authors reported data on three sunscreen formulations and two standard controls, P3 with an SPF of 15.7 ± 2.0 (7) and P7 with an SPF of 4.4 ± 0.4 (7), that appear to support a higher SPF on subjects with lower MEDs. The data from the two standard controls included 17 subjects, whereas each of the sunscreen formulations had data from 10 or 12 subjects. An analysis on the published plot (Figure 3a in reference 9) suggested that as the subject's MED increases by 10 mJ/cm^2, the SPF of the P7 control standard decreases by 0.35 units. Similarly, the SPF of the P3 standard control decreased by 1.09 units for every 10 mJ/cm^2 increase in MED (Figure 3b in reference 9). The SPF of sunscreen A (SPF label is 15+) decreased by approximately 1.88 units for every 10 mJ/cm^2 increase in MED (Figure 4a in reference 9); the SPF of sunscreen B (SPF label is 15+) decreased by approximately 4.07 units for every 10 mJ/cm^2 increase is MED (Figure 4a in reference 9); the SPF of sunscreen C (SPF label is 30+) decreased by approximately 8.42 units for every 10 mJ/cm^2 increase in MED (Figure 4a in reference 9). Although these data indicate a negative relationship between MED and SPF, the number of data points is small.

Herein, we report data on 2,503 observations. In the data reported herein, the SPF of P2 (15.6 ± 2.5) decreases by 1.55 units for every 10 mJ/cm^2 increase in MED (dashed line in Figure 3). This data suggests that to maximize the SPF value of a sunscreen, one should use subjects with very low MEDs. These data also suggest that people with higher MEDs will have less protection from a sunscreen compared with people with lower MEDs.

This negative correlation between unprotected MED and SPF of a sunscreen might account for some of the variability found in data from different sunscreen testing laboratories. This correlation also suggests that sunscreen testing should include subjects with a range of unprotected MED values, rather than Fitzpatrick Skin Phototype or ITA°. For example, perhaps not more than three subjects below an unprotected MED of 15 mJ/cm^2 and at least three subjects above an unprotected MED of 40 mJ/cm^2. The purpose of such a requirement would be identical to the reasons for a variety of Fitzpatrick Skin Phototype in the 2011 FDA-Final Rule.

CONCLUSIONS

No clinically significant difference or statistically significant difference was found between the average SPF of P2 using the 2011 FDA-Final Rule methodology versus that using ISO 24444 methodology at this laboratory. Furthermore, the average SPF of P2 is independent of the type of solar simulator (multiport versus single-port), age of subject, gender of subject, or Fitzpatrick Skin Phototype of subject. The data clearly show a statistically significant negative correlation between a subject's SPF of P2 and the subject's unprotected MED.

REFERENCES

(1) Department of Health and Human Services, Food and Drug Administration, Sunscreen drug products for over-the-counter human use; final monograph, *Fed. Regist.*, 4, 27666–27693 (1999).
(2) Department of Health and Human Services, Food and Drug Administration, Sunscreen drug products for over-the-counter human use; proposed amendment of final monograph; proposed rule, *Fed. Regist.*, 72, 49070–49122 (2007).
(3) The Australian/New Zealand Standard™ (1998), Sunscreen Products - Evaluation and classification, (Standard # AS/NZS 2604).
(4) COLIPA, *International Sun Protection Factor (SPF) Test Method, February 2003 (Joint Conference on Harmonization, Colipa, JCIA, and CTFA SA)*.
(5) COLIPA, *International Sun Protection Factor (SPF) Test Method, May 2006 (Joint Conference on Harmonization, Colipa, JCIA, CTFA SA and CTFA)*.
(6) Department of Health and Human Services, Food and Drug Administration, Labeling and effectiveness testing; sunscreen drug products for over-the-counter human use, *Fed. Regist.*, 76, 35620–35665 (2011).
(7) *Cosmetics–Sun Protection Test Methods–In Vivo Determination of the Sun Protection Factor (SPF)*, ISO 24444:2010(E), accessed is November 15, 2010, https://www.iso.org/standard/46523.html.
(8) K. Garzarella, M. Caswell, Disparate SPF testing methodologies generate similar SPFs, *J. Cosmet. Sci.*, 64, 297–307 (2013).
(9) D. L. Damian, G. M. Halliday, R. S. Barneston, Sun protection factor measurement of sunscreens is dependent on minimal erythemal dose, *Br. J. Dermatol*, 141, 502–507 (1999).
(10) World Medical Association Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Subjects, accessed January 27, 2019, www.wma.net/Policy/Current Policies/WMA Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Subjects.
(11) The Belmont Report–Ethical Principles and Guidelines for the Protection of Human Subjects of Research, 1979, accessed January 27, 2019, www.hhs.gov/OHRP/Reg0075lations&Policy/TheBelmontReport.

(12) D. S. Berger, Specification and design of solar ultraviolet simulators, *J. Investig. Dermatol.*, **53**, 192–199 (1969).
(13) M. Caswell, C. Wood, G. Roberts, A. Martinez, No Difference Detected in SPF Determined Using a 300 Watt Solar Simulator Versus a 150 Watt Solar Simulator, 19th Annual Photomedicine Society Meeting, accessed March 3, 2010.
(14) A. Martinez, G. Roberts, K. Garzarella, M. Lutz, M. Caswell, Interoperability of 300 watt and 150 watt xenon arc solar simulators in sun protection factor and in UVA protection factor clinical testing, *Photodermatol. Photoimmunol. Photomed.*, **29**, 78–83 (2013).
(15) T. B. Fitzpatrick, Soleil et peau, *J. Aesthetic Med.*, **2**, 33–34 (1975).
(16) T. B. Fitzpatrick, The validity and practicality of sun-reactive skin types I through VI, *Arch. Dermatol.*, **124**, 869–871 (1988).
(17) S. Nakagawa, H. Schielzeth, A general and simple method for obtaining R2 from generalized linear mixed-effects models, *Methods Ecol. Evol.*, **4**, 133–142 (2013).
(18) R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, accessed on January 27, 2019, www.R-project.or).
(19) A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, ImerTest package: tests in linear mixed effects models, *J. Stat. Softw.*, **82**, 1–26 (2017).
(20) H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2009).
(21) E. D. Baron, R. S. Stern, C. R. Taylor, Correlating skin type and minimum erythema dose, *Arch. Dermatol.*, **135**, 1278–1279 (1999).
(22) B. K. Ho, J. K. Robinson, Color bar tool for skin type identification: a cross-sectional study, *J. Am. Acad. Dermatol.*, **73**, 312–313 (2015).
(23) A. Chardon, I. Cretois, C. Hourseau, Skin colour typology and suntanning pathways, *Int. J. Cosmet. Sci.*, **13**, 191–208 (1991).
(24) R. M. Sayre, D. L. Desrochers, C. J. Wilson, E. Marlowe, Skin type, minimal erythema dose (MED), and sunlight acclimatization, *J. Am. Acad. Dermatol.*, **5**, 439–443 (1981).