Cosmetics Europe
the personal care association

COSMETICS EUROPE:

GUIDELINES FOR THE EVALUATION OF THE EFFICACY OF COSMETICS PRODUCTS

MAY 2008

COLIPA GUIDELINES
Efficacy Evaluation of Cosmetic Products

SUMMARY

*** * * * ***

PREAMBLE

This document replaces the second edition (2001) of the Colipa Guidelines for the Evaluation of the Efficacy of Cosmetic Products.

These guidelines aim to assist the cosmetics industry to comply with the applicable European regulations for the efficacy evaluation of cosmetic products. Because methodologically sound research is essential for the efficacy evaluation, these guidelines provide an overview of established testing methodologies.

Different types of experimental studies can be used to provide data on the performance of cosmetic products, and it should be borne in mind that new techniques are continually evolving and being published. The current "state of the art" should, therefore, be taken into account. It may also be useful to take into consideration the various guidelines published elsewhere, e.g. EEMCO guidelines relative to instrumental clinical techniques, International Sun Protection Factor Test Method, international guidelines (e.g. ISO, CEN, ICH, etc.).

Cosmetic claim substantiation should be an integral part of product development and design and not follow a rigid pre-conceived standard superimposed after development only for the purpose of supporting the communication of product performance and benefits. Moreover, the industry should be encouraged to research and develop new and improved cosmetic effects for the benefit of all consumers, provided these benefits are accurately communicated and delivered by the product. Validated evaluation methodologies such as those described below provide an appropriate and effective tool to assess the validity of product efficacy while facilitating innovation and competition.

The use of studies substantiating product efficacy in product communication, as well as the use of other information sources that are not pertinent to this document (e.g. data found in the literature such as that available on key raw materials) are the subject of a separate Colipa guideline for Communication on Cosmetic Products (see bibliography).

Certain product efficacy claims, even if substantiated by established methodologies, may fall outside the scope of the Cosmetics Directive. These guidelines do not address the scope of claims not permitted on cosmetic products, and users are advised to consult the borderline product manual of the European Union (see bibliography).

In accordance with the requirement to include claim substantiation in the cosmetic Product Information package, kept on file by the person placing the product on the market, a short summary of the technical data supporting the effect should be accessible to control authorities. The summary of claim substantiation in turn may be derived from data generated using the methodologies outlined below. The requirements for the documentation of the efficacy of cosmetic products in Europe are covered in the Colipa guidelines on Product Information Requirements in the European Union (see bibliography).

I. MAIN METHODOLOGICAL APPROACHES

Depending on the intended use of the cosmetic product in course of development, it is possible to use and combine several experimental approaches:

- the sensorial approach (sight, touch, olfaction) by consumers themselves or experts;

- the instrumental approach which favours specific criteria measured using *in vivo*, *ex-vivo* or *in vitro* approaches which do not reproduce normal conditions of the use of products, but allow objective analysis of specific activities taken out of context or attempt to replicate key parts of the product use cycle under controlled conditions.

Experimental design of studies is a large and complex subject and for optimal results has to rely on knowledge and awareness of statistical principles in design and analysis of the study, including appropriate consultation with an experienced statistician. This is to ensure that the studies achieve scientifically valid conclusions with the minimum number of subjects.

I.1 Evaluation on Human volunteers

I.1.1 Sensorial tests

These tests are based on an appreciation of product performance made through the senses of either panellists or of experts. They give information mainly on observed or perceived parameters.

a) Auto Evaluations

- **Use tests by consumers**
  A use test evaluates the consumers' perception of product efficacy and cosmetic properties based on parameters that they can observe or feel. They must be conducted on a sufficient number (see Statistical Guidance) of people.

  There are two main types of use tests:
  - Blind use tests are product tests without providing any information such as brand, decor, communication which could influence the consumers' judgement and alter their perception of the effect of the product alone.
  - Concept use tests are product tests combined with elements of communication that help to check whether the concept, the communication and the effect of the product as perceived by the consumers match up; information from concept use tests are used to complement that contained in the product efficacy dossier.

- **Sensorial-evaluation tests by trained expert panels**
  The sensorial evaluation enables a profile of the product to be drawn up according to predefined criteria. They must be conducted with the help of a panel of trained experts, following a well-defined protocol, with precise sensorial criteria.

b) Evaluation by professional experts

- **Tests under medical supervision**
  These tests in relation to the cosmetic benefits of a product are performed under the control of a physician. The parameters are evaluated by clinical

observation and/or scoring. They can be quantified by comparison with initial results or with an untreated control or a placebo or a reference product.

- **Tests under the control of other professionals**
  The tests can be conducted by a suitably qualified professional. Examples include: paramedical practitioners, hairdressers, aestheticians or other professional experts. The above would evaluate the performance of a product in terms of tactile and visual appreciation against a previously established scale.

Auto-evaluation by volunteers themselves can be associated with these tests (evaluation by professional experts) in order to assess that they perceive the expected effects.

I.1.2 Instrumental tests

These tests are performed with instruments that can precisely measure given parameters, according to a defined protocol, following the application of a product on human subjects.

- **Laboratory instrumental tests**
  These tests are performed under the control of a technician trained in the skilled use of the apparatus. The measurements are made on subjects in controlled laboratory conditions/environment: for example measurements of hydration, roughness, firmness, elasticity of the skin or measurements such as sun protection factors, UVA protection factors of filtering products, etc.

- **Instrumental measurements associated with an evaluation by professional experts**
  These measurements are made under the control of a suitably qualified professional (see section I.1.1 (b)) and use precise criteria: e.g. trichogram analysis and its derivatives for hair formulae, measurement of hydration and of the skin's mechanical properties, measurement of cutaneous fold/crease, centimetre measurements, colorimetric tests, etc.

I.2 *Ex Vivo* / *In Vitro* Tests

Ex Vivo (latin: *"off the living")*: relates to phenomena observed in the laboratory on a biological substrate taken from a living organism, without modification to the intrinsic properties of the substrate. For example *ex vivo* tests may correspond to instrumental tests which are conducted in the laboratory on keratin supports such as isolated hair fibres or hair tresses which have been cut from the human head in order to measure their mechanical properties, their surface properties or their colour in conditions that allow for the isolation of any effects emanating from the scalp. Other examples of *ex vivo* studies would include skin microflora and tape strips of skin. These tests can be generally quantified and comparative (with and without a specific ingredient, reference product, etc.).

In Vitro (latin: *"in glass"*): relates to phenomena observed in the laboratory, in artificial media (e.g. in test tubes or other containers such as culture dishes). *In vitro* tests are generally made in order to give prominence to performances which can be provided by ingredients or finished products which can be best demonstrated in this way. They can be comparative and their results may be quantified. *In vitro* tests may be used as screening instruments during product development or to illustrate an ingredient's mode of action.

They may also be used to demonstrate mechanisms relating to the finished product provided that a correlation with actual product usage can be demonstrated.

They can also be used to establish the efficacy of a finished product, in the instance where they have been correlated with a reference *in vivo* method. *In vitro* data may be used without reference to an *in vivo* method, but support for product efficacy should not be based purely on such data.

The substrates used may be biological (e.g. hair maintained artificially in order to study its growth kinetic, cell cultures, reconstructed skin, etc.) or artificial (e.g. glass or quartz or plastic plates and various containers).

Remark: The presentation of the various types of tests described above in section I is not restrictive and does not exclude tests which may correspond to other experimental approaches, which must nevertheless satisfy the general principles applicable to all scientific procedures.

II. GENERAL PRINCIPLES FOR ALL TESTS

Studies must be relevant and comprised of methods which are reliable and reproducible. The studies should follow a well-designed and scientifically valid methodology according to good practices. The criteria used for evaluation of product performances should be defined with accuracy and chosen in compliance with the aim of the test.

Studies conducted on volunteers should naturally respect ethical rules and products tested should have previously undergone a safety investigation. Human studies should be conducted on the target population when necessary, defined by strict inclusion/exclusion criteria.

Depending on the aim of the study, tests can be open, single- or double-blind.

Ex vivo/in vitro tests must be conducted under standardized conditions and their protocols must refer to published and/or “in house” validated methods. Clear descriptions of the methodology will be documented, as well as the statistical analysis of the data. These tests should be conducted in a controlled environment.

A study protocol must be drawn up and validated by the parties involved. This is essential to enable the study manager/promoter to monitor the study and the experimenter to carry out the test in order to ensure its quality.

The test laboratories must have standardized operating procedures. The equipment must be the subject of documented maintenance adapted to its use. Whatever the type of study, it is important that the person conducting the study:

- has the appropriate qualifications;
- has the training and experience in the field of the proposed study; and
- is respected for ethical quality and professional integrity.

A study monitoring system must be set up in order to ensure that the protocol and the operating procedures are correctly followed.

Data processing and the interpretation of results must be fair and should not overstep the limits of the test's significance. Data recording, transformations and representation in tabular or graphical form should be transparent or clearly explained

if complex. It should not be designed to overstate the effect(s) measured. Appropriate statistical analysis of the data should be performed.

III. INFORMATION WHICH SHOULD APPEAR ON TEST PROTOCOLS

When pertinent, the following information should appear on test protocols:

III.1 General Information

III.1.1 Study objective

The study objective must be stated clearly.

III.1.2 Product tested and reference product (if used)

- Type of product (e.g. skin cream).
- Quantity of product applied if applicable.
- Product to be tested and reference product(s) (if used).

The product(s) to be tested must be correctly identified. They must indicate a date of manufacture, and their storage conditions must be in line with data on product stability. A use-by date may possibly be mentioned for the requirements of the test.

The product(s) can be prepared extemporaneously (in the case of mixtures or the making up of solutions) and their use must be consistent with the test objective. The preparation must be adapted to the tests, and not bias them in any way. After the test, a sample of the product tested and the reference product should be retained for at least six months under suitable conditions by the investigator and/or the promoter.

III.1.3 Test procedure

- Timetable.
- Study location.

III.1.4 Data management – Data processing – Analysis of results

The methods of collecting data (questionnaire, observation notebooks, laboratory books, diaries, electronic CRF [1] etc.) are indicated. Details must be given regarding the management of electronic data (single or double capture of data input; control to assure the coherence of data, etc.).

Calculations carried out and the statistical analysis used to meet the defined test objective must be specified. Statistical methods (statistical tests chosen, alpha risk and software used) should be indicated. The data obtained on the reference product(s) should help to validate the study and/or provide a comparison with the product studied.

III.1.5 Equipment and reagents

- Description, specification and identification of equipment (including the commercial reference of the model).
- Usage conditions.
- Relevance of the measurement in relation to the study objective.

[1] Clinical Research File

The reagents should be properly identified and kept in an appropriate manner.

III.2 Specific Information

III.2.1 Evaluation on human volunteers

III.2.1.1 Product tested

The safety of using the product under the protocol conditions must be established.

III.2.1.2 Volunteers

- Inclusion and exclusion criteria: demographic criteria, criteria linked to the study**.**
- Number: justification of the number of subjects based on statistical and/or methodological expertise (background data). It is possible to include more subjects to allow for subject drop-outs.
- Training (time period, validation, etc.) and number of trained panelists for sensorial evaluation tests by experts.

III.2.1.3 Methodology

- Experimental design: randomized study; single or comparative test; subjects used as control or not; open, single- or double-blind test; etc.
  - Sensorial evaluation tests by experts: discriminative or filing, ranking, monadic or comparative test, etc.
- Evaluation parameters: definition of efficacy criteria adopted.
- Product application methods: quantity of product applied if applicable, frequency of use, time of products' application / application areas, restrictions for use.
- Chronology of examinations, measurements.
- Evaluation methods
  Relationship between methodology and effects to be assessed.
  In the case of novel methods, indication of information sources which confirm their relevance.
  - Use tests by consumers: method (interview, correspondence, telephone, etc.) and format (evaluation form, questionnaire).
  - Sensorial evaluation tests by experts: design of presentation of products to trained panelists and notation method (types of scale).
  - Scoring done by a suitably qualified health or professional expert. These can be visual scores using scales drawn up, the evaluation can be direct or based on supports such as hair tresses, colour make-up testers, photographs, tactile scores (softness) or other sensorial scores (sniff-tests) etc. These scales and supports must be described.
  - Auto-evaluation by volunteers themselves, using questionnaires, closed questions, notes, multiple choice etc.
  - Instrumental methods using measuring devices, whether requiring standardized environmental conditions or not (temperature, humidity, light) for example: phototrichogram, imperceptible water loss, SPF determination. The equipment used must be described. The operations to be performed can be explained in detail or be succinct, referring to technical procedures or to publications.

III.2.2 *Ex vivo* / *in vitro* tests

III.2.2.1 Substrate

The relevance of the substrate must be explained. It should be described in detail. The substrate must be standardized, identified and preserved in an appropriate manner.

For a reactive system, its nature and origin must be specified, as well as the method of obtaining/preparing it.

III.2.2.2 Methodology

The number of subjects in the sample/test and the number of tests must be specified.
The test planning should be explained (timetable defined):
- Pre-treatment (define and set out the chronology if necessary).
- Frequency and duration of treatments/organization of measurements.
- Randomization of treatments if several products are being tested.
- Quantity of product tested and justification.
- Incubation.
- Measurements carried out.
- Frequency and timing of analyses.

IV. INFORMATION WHICH SHOULD APPEAR ON TESTS REPORTS

When pertinent, the following information should appear on test reports

IV.1 General Information

IV.1.1 Identification

- The sponsor of the study.
- The organisation in charge of the assessment and the address of the laboratories where the tests actually take place.
- The person responsible for testing (if applicable, the identification and qualifications of the investigator).
- If appropriate, other investigators involved.
- The product(s) tested: type of product, formula number, batch number or code, etc.
- Issue date of the report.

IV.1.2 Objective of the test

IV.1.3 Test schedule

- Starting date.
- Finishing date.

IV.1.4 Methodology

- Summary of protocol (if necessary, the detailed protocol will be appended to the report).
- Documentation of any deviation from the protocol.

IV.1.5 Statistics

- Definition of method employed.
- Outcome of statistical analysis.
- If not stated in the report, justification.

IV.1.6 Results

- Presentation of results.
- Methods for analysing and interpreting results.
- Individual data can be given in appendix.

IV.1.7 Discussion

IV.1.8 Conclusion

IV.1.9 Signatures of the persons responsible for testing

- Technician(s).
- Investigator.
- Quality assurance.
- Person responsible for the statistical analysis or statistician, if appropriate.

IV.1.10 Summary of the report

IV.2 Specific information

IV.2.1 Evaluation on human volunteers

- Panel:
  - justification of panel choice with regard to specific effects' assessment; and
  - demographic criteria.

- Drop-outs (withdrawals, interrupted tests):
  - size of sample analyzed; and
  - consideration of drop-outs with justification (as far as possible).

IV.2.1.1 Use tests by consumers

- Panel:
  - Socio-demographic criteria
- Presentation of results:
  - wording of questions for which responses confirm effects relevant to the claim;
  - assessment method used (nominal, ordinal or visual analogical notation scale); and
  - if justified, consideration of extraneous factors.

IV.2.1.2 Sensorial evaluation tests by trained expert panels

- Presentation of results:
  - choice of presentation of results (e.g. spider profile, principal component analysis, etc.);
  - analysis of the inter-variability of the panel; and
  - list of criteria assessed.

IV.2.1.3 Evaluation by a professional expert and Instrumental tests

- Presentation of results:
  - for quantitative data: number of subjects, median, standard deviation, percentages;
  - for qualitative data: absolute or relative frequency (percentages);
  - method used to assess the observed effect;
  - interpretation of results, taking into account the expected normal range for measured values, the expected magnitude of outcome, the variability of individual reactions; and, if justified, consideration of extraneous factors.

IV.2.2 *Ex Vivo* / *In Vitro* tests

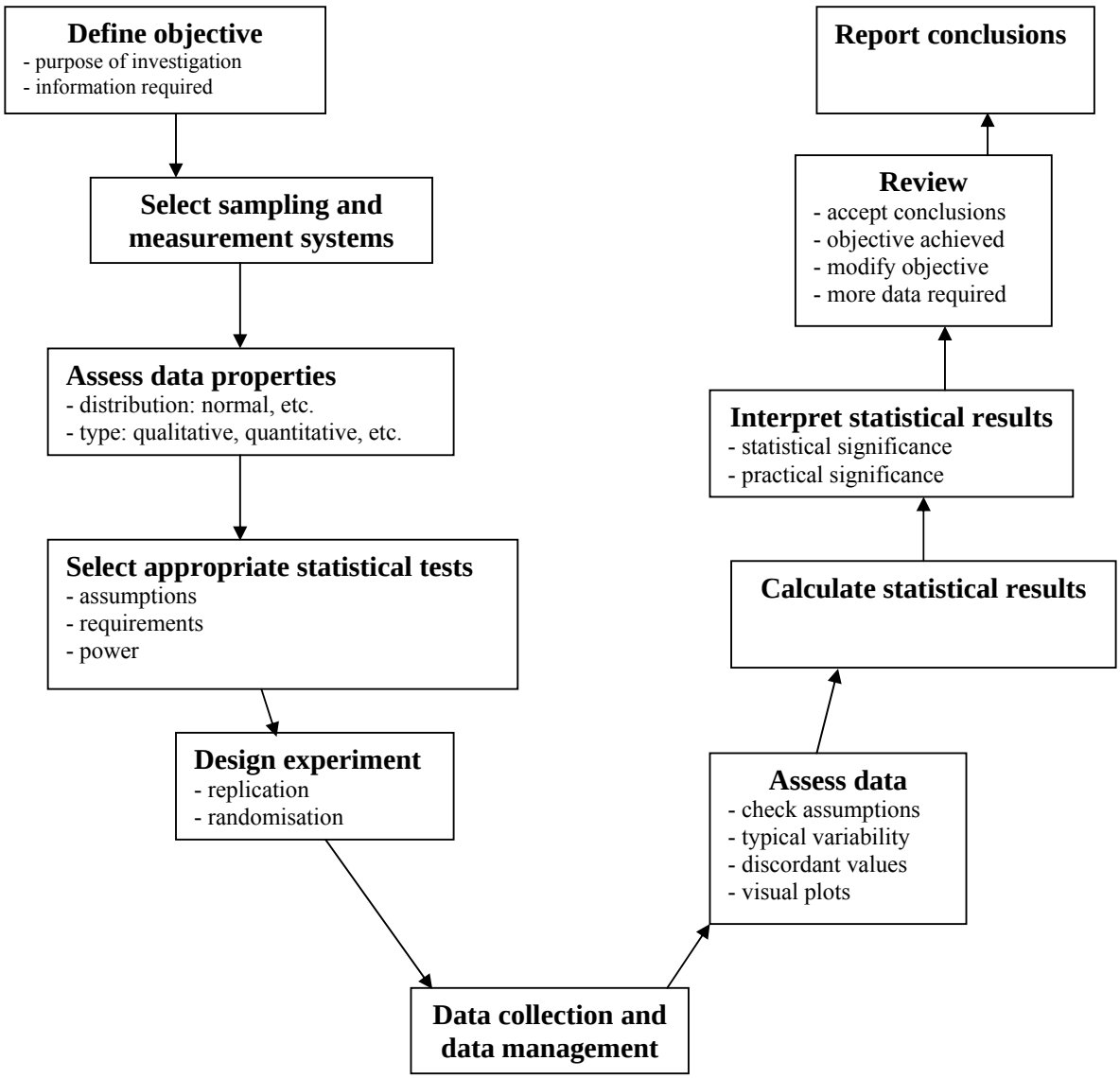
- Presentation of results:
  - results recording; and
  - interpretation of results, in particular with reference to the performance and limitations of the method used.

Remark:
The indications above in sections III and IV are given as examples; they might not all be relevant depending on the test under consideration and they are not exhaustive; they illustrate the need to include in the test protocols and reports all useful information that can assure the reliability of the study.

STATISTICAL GUIDANCE

Statistics is the science of using data to increase the probability of making correct decisions. Generally, inferences are made about populations based on data obtained by sampling from that population. This appendix describes some of the general principles of statistics but is not a substitute for training in statistical methods. If in doubt, refer to a suitable text or seek assistance from suitably qualified personnel.

The purpose of applying statistical methods is to get an objective assessment of the information contained in the data. The flowchart below summarizes the steps you should go through to ensure an effective statistical approach to data analysis:



General principles:

1) Sources of variation in data

All measurements are subject to variation. There are two types: special cause or common cause, either of which may be systematic or random. They have different properties. Special causes of variation are factors known to affect the measurement e.g. concentration of reagent. These effects can be estimated or eliminated by good experimental design. Common causes of variation are random, uncontrolled or uncontrollable effects e.g. measured value is different from true value because of the variability inherent in the measurement method. If variation is systematic this will introduce bias in the data which may make it impossible to derive sound conclusions from your results. All measurements are subject to random error. Random errors cause the measured values to vary without any particular pattern of deviation.

2) Study design

Before the study can be designed, you must define the study objective, what information is required to test it and how you wish to analyse the data. You may need to loop round the flowchart iteratively until the design is optimised. There are many possible designs that could be considered – it is important to choose the design that is most appropriate to address the study objective. Also when designing a study, it is important to minimise possible bias. Randomisation, pairing and blocking are techniques to minimise this.

3) Sample Size & Power

The size of study required will depend on the magnitude of the effect you wish to detect, the variability of the data and the power of the study. In general, the smaller an effect you wish to detect, the larger your study needs to be (all other factors holding constant). The power of a study is the probability that e.g. it will detect a difference of the magnitude specified if it truly exists. It is typical to size studies based on 80% or 90% power. However, for exploratory or pilot studies a smaller power can be chosen.

The number of subjects /size of a study should always be large enough to provide a reliable answer to the questions addressed (i.e. have sufficient power). The number is usually determined by the primary objective of the study through a formal sample size calculation or by a justification based on statistical and/or methodological expertise (background data, former study, etc.).

4) Data Management

Poor data collection and recording can affect the results of the analysis. Processes must be in place covering data entry, data manipulation and data transfer to ensure high data quality. Data should be recorded to adequate levels of resolution required for analysis. Check that data is not truncated or rounded before recording and record it to appropriate statistically significant figures.

5) Making decisions

During the study design phase, you will have generated an hypothesis (e.g. null hypothesis: no difference between treatments versus alternative hypothesis: there is

a difference between treatments) that you wish to test. Now you have generated your study data, we ascertain whether there is sufficient evidence from the data to reject the null hypothesis in favour of the alternative hypothesis. The decision whether to reject the Null hypothesis or not, is based on the value of an appropriate test statistic calculated from the data and compared with a critical value of the statistic and this results in a p-value. A p-value is the probability of obtaining the value observed or one more extreme when there is in fact no difference.

Typically a significance level of 5% is chosen (2.5% in case of one-sided). This is the benchmark against which the p-value generated from the hypothesis test is compared. Obtaining results with p-values below 0.05 indicate that the risk of these differences having happened by chance alone is small i.e. less than 5%.

It is also good practice to calculate confidence intervals for your results to present with the p-values. A confidence interval gives an indication of the reliability with which the statistic based on the sample, estimates the true value from the population. Typically 95% confidence intervals are presented.

It is important to appreciate that you may obtain results that are statistically significant i.e. with p-values less than 0.05, but the results may not be of practical or clinical significance because for example the difference you have detected is so small to be of no practical or clinical relevance.

6) Statistical Method

There are many different statistical techniques. To analyse the data it is important to use a statistical method which is appropriate to the purpose of the analysis, to the data type and to the data interdependency. Using the incorrect technique will mean the conclusions drawn are not sound.

References

Statistical Methods in Medical Research. Fourth Edition. P Armitage, G Berry and JNS Matthews (2001)

Practical Statistics for Medical Research, DG Altman (1991)

Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. Box GEP, Hunter WG & Hunter JS (1978)

ICH E9 Statistical Principles for Clinical Trials

ANNEX 2

BIBLIOGRAPHY

Cosmetics Directive 76/768/EEC and amendments.

IFSCC Monograph N°1, Principles of product evaluation: objective sensory methods, 1987.

Declaration of Helsinki, adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964; amended by the 29th World Medical Assembly, Tokyo, Japan, October 1975, the 35th World Medical Assembly, Venice, Italy, October 1983 and by the 41th World Medical Assembly, Hong Kong, September 1989.

J. Serup, EEMCO guidance for the assessment of dry skin (xerosis) and ichthyosis: clinical scoring Systems, Skin Research and Technology 1995, 1,109-114.

G. E. Piérard, EEMCO guidance for the assessment of dry skin (xerosis) and ichthyosis: evaluation by stratum corneum strippings, Skin Research and Technology 1996,2, 3-11.

E. Berardesca, EEMCO guidance for the assessment of stratum corneum hydration: electrical methods, Skin Research and Technology 1997, 3,126-132.

G. E. Piérard, EEMCO guidance for the assessment of skin colour. Journal of the European Academy of Dermatology and Venerology 1998, 10, 1-11.

G. E. Piérard, EEMCO guidance to the *in vivo* assessment of tensile functional properties of the skin, Skin Pharmacology and Applied Skin Physiology 1999, 12, 352-362.

J. L. Lévêque, EEMCO guidance for the assessment of skin topography. Journal of the European Academy of Dermatology and Venerology 1999, 12, 103-114.

M. Lodén, Efficacy testing of cosmetics and other topical products, IFSCC Magazine, 2000, 3, vol.3, 47-53.

G.E. Piérard, EEMCO guidance for the *in vivo* assessment of skin greasiness. Skin Pharmacology and Applied Skin Physiology 2000, 13, 372-389.

Colipa, Guidelines for the Evaluation of the Efficacy of Cosmetic Products, Second Edition, 2001.

V. Rogiers, EEMCO guidance for the assessment of transepidermal water loss in cosmetic sciences. Skin Pharmacology and Applied Skin Physiology 2001, 14, 117-128.

R. Marks, J.L. Lévêque, R. Voegeli, The essential Stratum Corneum, Ed. Martin Dunitz, 2002

J.L. Parra, EEMCO guidance for the *in vivo* assessment of skin surface pH. Skin Pharmacology and Applied Skin Physiology 2003, 16, 188-202.

G.E. Piérard, EEMCO guidance for the efficacy assessment of antiperspirants and deodorants. Skin Pharmacology and Applied Skin Physiology 2003, 16, 324-342.

G.E. Piérard, EEMCO guidance for the assessment of hair shedding and alopecia. Skin Pharmacology and Applied Skin Physiology 2004, 17, 98-110.

The science of Hair care. Second edition. Edited by Claude Bouillon, John Wilkinson, Talor & Francis, 2005

CTFA South Africa, Colipa, JCIA, CTFA, International Sun Protection Factor (SPF) test method, May 2006.

Commission Recommendation of 22 September 2006 on the efficacy of sunscreen products and the claims made relating thereto (notified under document number C(2006) 4089) (2006/647/EC)

Ethnic Skin and Hair , Informa. Edited by Enzo Berardesca, Jean-Luc Leveque, Howard I. Maibach, 2007

Cosmetic Product Information Requirements in the European Union – Amended Guidelines for the Cosmetic Industry Based on Article 7a of the 6th Amendment to the EU Cosmetics Directive

http://ec.europa.eu/enterprise/cosmetics/html/cosm_borderline_docs.htm

Colipa Guidelines for Communication on Cosmetic Products (draft)

COSMETICS EUROPE IS THE EUROPEAN TRADE ASSOCIATION REPRESENTING THE INTEREST OF THE COSMETICS, TOILETRY AND PERFUMERY INDUSTRY

COSMETICS EUROPE – THE PERSONAL CARE ASSOCIATION
AVENUE HERRMANN-DEBROUX 40, 1160 BRUSSELS
T. +32 2 227 66 10, F. +32 2 227 66 27
WWW.COSMETICSEUROPE.EU