Scientific Committee on Consumer Safety

SCCS

OPINION ON

Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2)

Submission I

The SCCS adopted the final Opinion by written procedure on 30 July 2018

ACKNOWLEDGMENTS

SCCS members listed below are acknowledged for their valuable contribution to the finalisation of this Opinion.

For the Preliminary Opinion

SCCS members
Dr U. Bernauer
Dr L. Bodin
Prof. Q. Chaudhry (SCCS Chair)
Prof. P.J. Coenraads (SCCS Vice-Chair)
Prof. M. Dusinska
Dr J. Ezendam (Rapporteur)
Dr E. Gaffet
Prof. C. L. Galli
Dr B. Granum
Prof. E. Panteri
Prof. V. Rogiers (SCCS Vice-Chair)
Dr Ch. Rousselle
Dr M. Stepnik
Prof. T. Vanhaecke
Dr S. Wijnhoven

SCCS external experts
Dr A. Simonnard
Dr N. von Goetz

For the Final Opinion

SCCS members
Dr U. Bernauer
Dr L. Bodin
Prof. Q. Chaudhry (SCCS Chair)
Prof. P.J. Coenraads (SCCS Vice-Chair)
Prof. M. Dusinska
Dr J. Ezendam (Rapporteur)
Dr E. Gaffet
Prof. C. L. Galli
Dr B. Granum
Prof. E. Panteri
Prof. V. Rogiers (SCCS Vice-Chair)
Dr Ch. Rousselle
Dr M. Stepnik
Prof. T. Vanhaecke
Dr S. Wijnhoven

SCCS external experts
Dr A. Simonnard
Prof. W. Uter
Dr N. von Goetz

All Declarations of Working Group members are available on the following webpage:
http://ec.europa.eu/health/scientific_committees/experts/declarations/sccs_en.htm

This Opinion has been subject to a commenting period of a minimum eight weeks after its initial publication (from 22 December 2017 until 26 February 2018). Comments received during this time were considered by the SCCS.
For this final Opinion, comments received resulted in the following main changes: new SCCS comments have been added, in particular with regard to sections 4.2.1, 4.2.3.5, 4.2.3.6, 4.3.3.1, and 4.3.3.4. The wording of the discussion and conclusion points has been slightly modified, without changing the SCCS conclusion.

1. ABSTRACT

The SCCS concludes the following:

- *In light of the methodology provided, does the SCCS consider QRA2 adequate to establish a concentration at which induction of sensitisation by a fragrance ingredient unlikely to occur?*

  The "QRA2 final report" together with the supplementary information received shows that a lot of progress has been achieved since the initial publication of the QRA. However, it is not yet possible to use the QRA2 to establish a concentration at which induction of sensitisation of fragrance is unlikely to occur. Several aspects of the methodology are not clear and the scientific rationale behind the methodology needs to be better described. These aspects have been highlighted in this Opinion.

- *Does the SCCS have any further scientific comments with regard to the use of QRA2 methodology to determine, in particular regarding applicability, development and improvements?*

  A number of additional considerations and refinements have been incorporated to the proposed methodology. However, explanation of certain methodological approaches and assumptions, as well as a description of uncertainties is lacking, the provision of which would enhance understanding of the methodology. These aspects have been highlighted in the SCCS comments under each section with the aim to provide pointers for improvement. If shaped up properly, this could be a useful methodology not only for risk assessment of fragrance allergens, but potentially also for other cosmetic ingredients.

Keywords: SCCS, scientific opinion, Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2), Regulation 1223/2009

Opinion to be cited as: SCCS (Scientific Committee on Consumer Safety), Opinion on Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2), preliminary version of 24-25 October 2017, final version of 30 July 2018, SCCS/1589/17

About the Scientific Committees

Two independent non-food Scientific Committees provide the Commission with the scientific advice it needs when preparing policy and proposals relating to consumer safety, public health and the environment. The Committees also draw the Commission's attention to the new or emerging problems which may pose an actual or potential threat.
They are: the Scientific Committee on Consumer Safety (SCCS), the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER) and are made up of independent experts.

In addition, the Commission relies upon the work of the European Food Safety Authority (EFSA), the European Medicines Agency (EMA), the European Centre for Disease prevention and Control (ECDC) and the European Chemicals Agency (ECHA).

SCCS

The Committee shall provide Opinions on questions concerning health and safety risks (notably chemical, biological, mechanical and other physical risks) of non-food consumer products (for example cosmetic products and their ingredients, toys, textiles, clothing, personal care and household products such as detergents, etc.) and services (for example: tattooing, artificial sun tanning, etc.).

Scientific Committee members

Contact:
European Commission
Health and Food Safety
Directorate C: Public Health, country knowledge and crisis management
Unit C2 – Country knowledge and Scientific Committees
Office: HTC 03/073
L-2920 Luxembourg
SANTE-C2-SCCS@ec.europa.eu

ISSN ISBN

Doi ND

TABLE OF CONTENTS

2. MANDATE FROM THE EUROPEAN COMMISSION

Background

Skin sensitisation is induced by exposure to certain chemicals and might lead to allergic contact dermatitis. Contact allergy to fragrance ingredients is a common, significant and relevant problem in Europe. Therefore it is a topic of high interest for consumers, industry and Regulatory Authorities.

A model for dermal sensitisation quantitative risk assessment (QRA) had been developed and implemented by the International Fragrance Association (IFRA). The methodology relied on thresholds, no effect or low effect levels, established in healthy human volunteers and/or in animal experiments. It included a set of safety factors applied for inter-individual differences, for vehicle effects and for use considerations, to derive a so-called *"acceptable exposure level"*. The exposure to an allergen in different product types should be below this level. The QRA has been evaluated by the Scientific Committee on Consumer Products (SCCP) in 2008 (SCCP/1153/08) which was of the opinion that:

"*The model has not been validated and no strategy of validation has been suggested. There is no confidence that the levels of skin sensitizers identified by the dermal sensitization QRA are safe for the consumer."*

However, the SCCP also concluded that:
"From a scientific point of view, models like the dermal sensitization QRA approach may, after refinement and validation, in the future be applicable for risk assessment of new substances to suggest a safe level of exposure prior to incorporation into products. In such cases an independent post-marketing surveillance system would be essential.
Aggregated exposures must be incorporated in the dermal sensitization QRA model. Validation must be performed employing a broad range of different chemicals and data from substantial clinical investigations."

Later in 2012, in the context of the opinion on Fragrance Allergens (SCCS/1459/11), the SCCS reiterated that:
"For substances for which there are no clinical data of concern, models such as the dermal sensitisation QRA approach may, after refinement and validation, be used to suggest a safe level of exposure prior to incorporation into products. However, aggregated exposures must be incorporated in the dermal sensitisation QRA model".

Following the SCCS opinion of 2012, the International Dialogue for the Evaluation of Allergens (IDEA) project was established to improve the risk assessment of fragrance allergens through the collaboration of academia and industry scientists together with clinicians. The IDEA project has worked on the development of the QRA, in particular focusing on reviews of the uncertainty factors and introducing dermal aggregate exposure as replacement for the original individual product exposure assessment for fragrance ingredients. This new quantitative risk assessment methodology was named QRA2 and further developments were introduced following the critical review performed by the Joint Research Centre (JRC) in 2015. The consolidated QRA2 was successively submitted to the Commission services in October 2016.

Terms of reference

(1) In light of the methodology provided, does the SCCS consider QRA2 adequate to establish a concentration at which induction of sensitisation by a fragrance ingredient unlikely to occur?

(2) Does the SCCS have any further scientific comments with regard to the use of QRA2 methodology to determine, in particular regarding applicability, development and improvements?

3. OPINION

During the commenting period of the preliminary Opinion on QRA2, the Applicant provided the SCCS with a document intended to supplement the original submission of the "*IDEA Project Final Report on QRA2*". This supplement contained not only a reply to the comments made by the SCCS, but also additional information and new Figures and Tables. This supplement can be found in Appendix 1. This new information has been evaluated by the SCCS and additional SCCS comments referring to this document are included in the Opinion whenever needed.

3.1 Scope and aim of QRA2

From the Final Report on QRA2, 2016:
The scope of the skin sensitisation QRA as presented here is the evaluation of the risk to consumers of the induction of contact allergy presented by fragrance ingredients in cosmetics and other household consumer products. The original risk assessment methodology (QRA1) was implemented by IFRA (International Fragrance Association; www.ifraorg.org) into standards on the first three ingredients in 2006. While IFRA membership accommodates about 90% (by volume) of the fragrances produced globally and used in consumer products, there are a number of product types and exposures to fragrance ingredients that are not under the scope of IFRA and therefore not covered by the IFRA Standards (e.g. aromatherapy, drugs and topical treatments, massage and spa therapies, occupational exposure, natural exposure, foods, etc.).

The aim of skin sensitisation QRA is the prevention of induction of contact allergy (primary prevention). If induction is prevented, elicitation will not occur. QRA is intended to deliver an output specifically in relation to induction. Elicitation thresholds are likely to be lower compared to induction thresholds. At present, the relationship between the potency of an allergen, the induction thresholds, and the ability of the substance to elicit responses has not been characterised (ECHA, 2012). In part this is due to the fact that elicitation thresholds depend not only on the intrinsic potency of a sensitiser, but also on the susceptibility of the exposed individual. This latter aspect being a function not only of potency, but also of the severity of the induction process (Hostynek and Maibach, 2004; Friedmann, 2007). Typically, substance-specific elicitation thresholds can only be derived from clinical studies using volunteers who are sensitised to the substance in question. Many examples of such work have appeared in the literature (e.g. Fischer *et al.*, 2009) and it has been suggested that the variation between the thresholds for contact allergens may be rather less than that for induction (Fischer *et al.*, 2011).

SCCS comment
Ideally, SCCS would prefer a QRA methodology that can be used both for primary and secondary prevention, to ensure protection of consumers with an existing fragrance allergy as well. However, SCCS agrees that it is challenging to define safe levels to prevent elicitation responses for the previously mentioned reasons. Therefore, the QRA2 methodology will be evaluated only for its adequacy to prevent induction of sensitisation, which may not necessarily mean prevention of elicitation of allergic reactions. Thus, in circumstances where (many) consumers have already become sensitized, data which indicate safe elicitation levels in humans would take precedence over theoretical models such as the QRA concerning presumably safe levels for induction (memorandum SCCS/1567/15).
Importantly, the use of fragrances not covered by IFRA standards (10% by volume) is not included in the aggregate exposure assessment and subsequent risk characterisation. Hence, the methodology cannot be considered adequate to ensure full primary prevention of fragrance contact allergy.

SCCS comment to the additional information (Appendix 1)

The reply from the Applicant (Appendix 1) confirms the concern of the SCCS that the IFRA standards derived from QRA2 do not cover the full market for fragrance compounds. This remains an uncertainty in QRA2.

3.2 General principle of QRA2

From the Final Report on QRA2, 2016:
The quantitative risk assessment methodology outlined in many publications (for instance WHO, 2004; ECHA, 2012; ECETOC, 2009) is the cornerstone of health-based exposure limits and used extensively by governmental agencies and industry. Safety assessments for chemicals that possess the ability to cause sensitisation by contact with the skin have traditionally been conducted using an *ad hoc* comparative risk assessment technique (Robinson *et al*., 1989). Since it is known that the general principles of quantitative risk assessment can also be applied to induction of skin sensitisation, an alternative and potentially better quantitative risk assessment approach for skin sensitisation was developed (Robinson *et al.,* 2000) and described in a series of papers (Farage *et al.*, 2003; Felter *et al.,* 2002; Felter *et al*., 2003; Gerberick *et al*., 2001; Griem *et al*., 2003). This Quantitative Risk Assessment (QRA) methodology was subsequently described for use with fragrance ingredients (Api *et al*., 2008). The skin sensitisation QRA approach follows the same four steps outlined above for general toxicology risk assessment. It is implicit that the conduct of the full skin sensitisation QRA is necessary only for those ingredients identified as dermal sensitisers.

The different phases of risk assessment (as described in detail in WHO, 2004) are as follows:

Hazard Identification
This involves the use of experimental data to determine the skin sensitisation potential of the fragrance ingredient. Historically, this has involved a murine Local Lymph Node Assay (LLNA) or the use of other assays such as the guinea pig maximization test or Buehler guinea pig test (Kimber *et al*., 2003, ECETOC 2003). Moving forward it will rely on the integrated assessment of data based on a weight of evidence analysis using all available data, including non-animal test methods.

Dose–Response Assessment or Hazard Quantification
The dose response for induction of skin sensitisation, from a previously executed LLNA, is used to identify an EC3 value (Estimated Concentration required to result in a threshold positive response; i.e. a Stimulation Index = 3). The EC3 value is used to define the relative sensitisation potency. A good correlation between the EC3 and the NOAEL in the Human Repeat Insult Patch Test (HRIPT) has been established (Gerberick *et al*., 2001; Basketter *et al*., 2005; Api *et al*., 2014).

Exposure Assessment
The amount of fragrance ingredient that remains on the skin under the conditions of product use in terms of quantity per unit area (e.g. $\mu g/cm^2$) is assessed. Exposure to the fragrance ingredient is determined using habits and practice data for consumer product use, human parameters data, the level of perfume in the finished product and the level of the individual fragrance ingredients in the perfume.

Risk Characterisation
The data from the previous steps are used to determine an acceptable exposure level to a fragrance ingredient against which the real-life exposure of consumers to that fragrance ingredient in a specific product type can be compared. The acceptability or unacceptability of real-life exposures can then be determined.
In developing a methodology for quantitative risk assessment for skin sensitisation of fragrance ingredients based on the above approach new terms were adopted. "No Expected Sensitisation Induction Level" (NESIL) and "Sensitisation Assessment Factors" (SAFs)

replaced the terms NOAEL and uncertainty factors, generally used in toxicological risk assessments. The Acceptable Exposure Level (AEL = NESIL/total SAFs) is equivalent to the 'reference dose (RfD)' used in general toxicology. These terms have been adopted to take into account unique elements of quantitative risk assessment for skin sensitisation and are described in detail in the sections within this dossier.
The overall skin sensitisation QRA2 is presented in Figure 1 and its use in conjunction with aggregated exposure is shown in Figure 2 and is detailed in the remaining sections of this report.

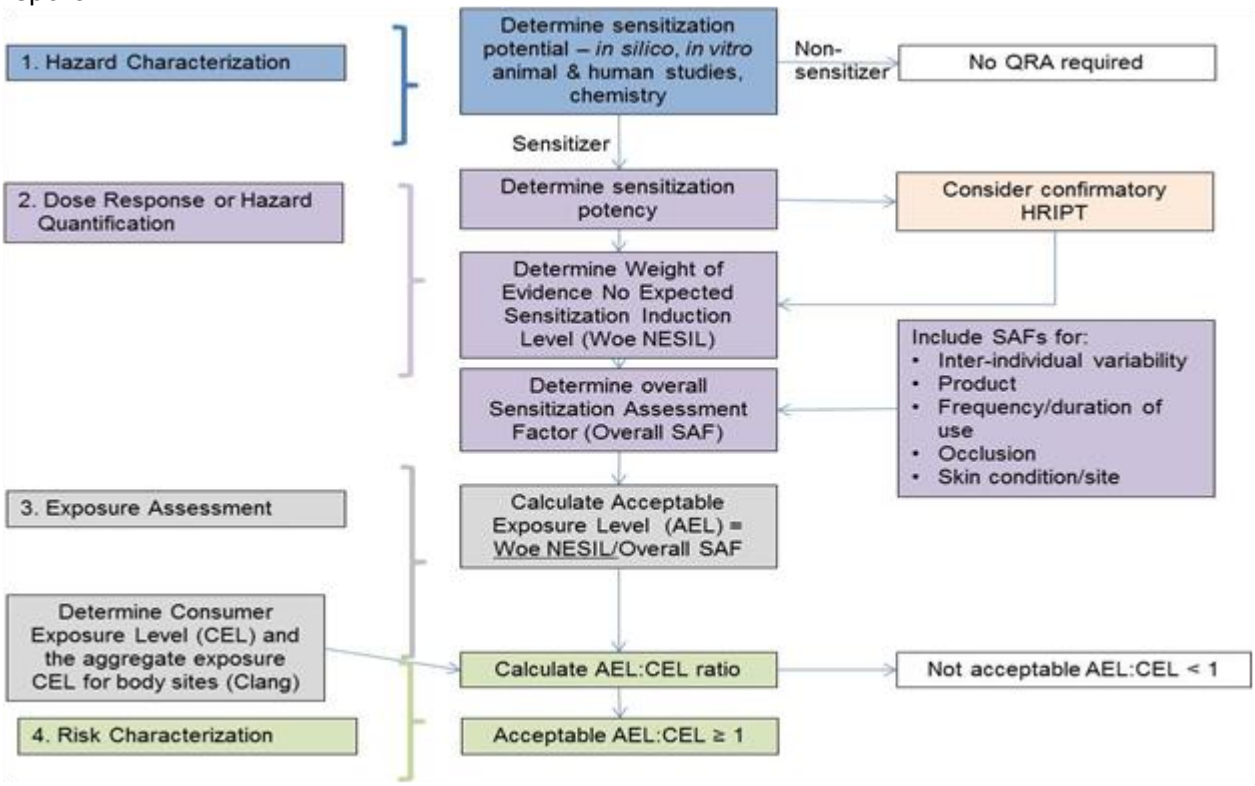


Figure 1: Skin Sensitisation QRA2 for Fragrance Ingredients (taken from Final Report on QRA2, 2016)

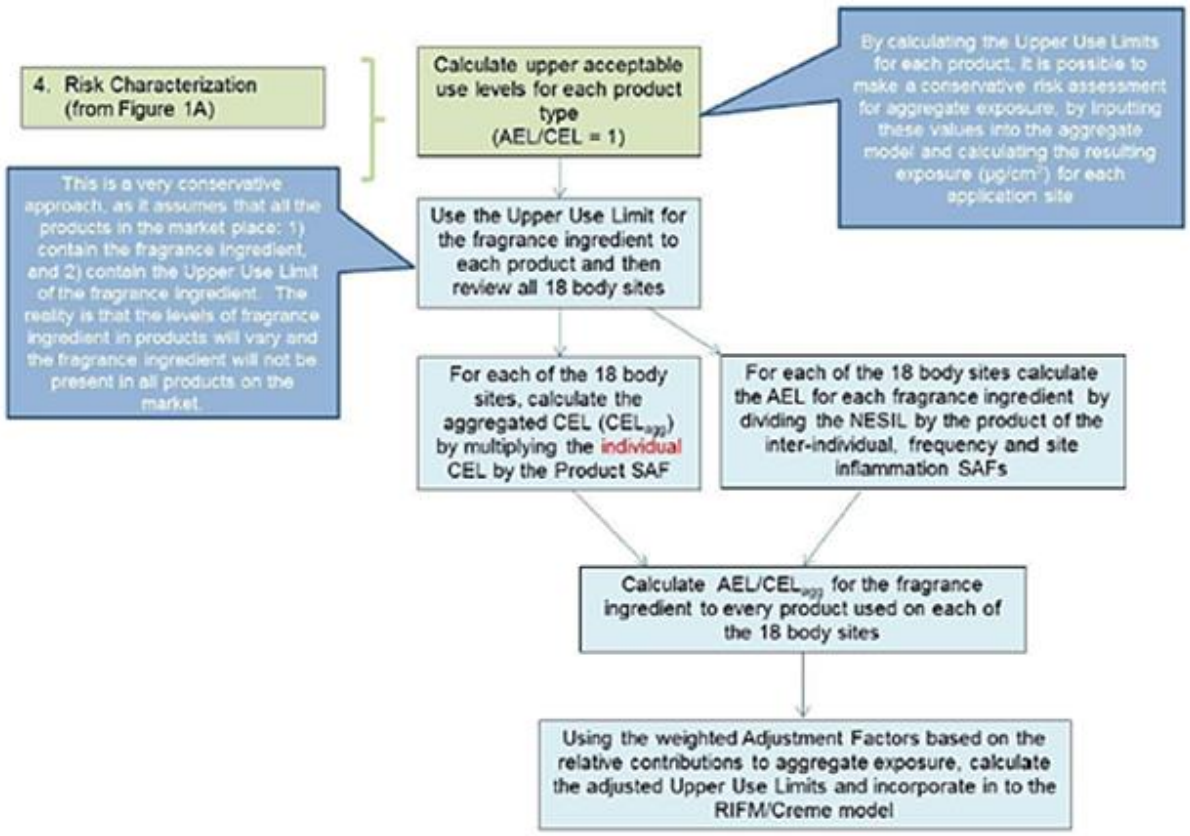


Figure 2: Use of QRA2 with Aggregate Exposure for Skin Sensitisation for Fragrance Ingredients (taken from final report on QRA2, 2016)

SCCS comments

In general, the way in which the QRA2 approach is presented in Figure 1 as well as the description of how the methodology is used in conjunction with aggregate exposure in Figure 2 is unclear. Furthermore, the relationship between the figures is unclear. An explanation is needed if an assessment can be performed using only the steps in Figure 1 or if the steps in Figure 2 are always part of QRA2. Since aggregate exposure is more the rule for fragrances than the exception, the SCCS suggests modifying Figure 1 so that SAFs are divided from the start into exposure-related (product-specific), location-related (site-specific) and general (hazard-specific) SAFs and be fed into the process at the respective places. In such an assessment it does not make sense to calculate "overall SAFs", but the SAFs should be used as weighting factors at each step in the process of aggregation. The figure is not always in line with the text in the final report. To be able to evaluate the adequacy of the methodology, the SCCS needs figures and explanatory texts that are clear and consistent.

Other aspects that should be reconsidered in the description of the methodology are:

- In Figure 1 the determination of the total SAF is done at the level of the hazard assessment. Since product-specific SAFs are defined, this would pose a problem in an aggregate exposure assessment if different SAFs would apply for the different products. It would seem easier to apply the product-specific SAFs to the single exposures that are later aggregated.
- The caption of Figure 2 suggests that Figure 1 refers to single product exposure, but there is a box for exposure assessment that reads "determine… and the aggregate CEL for body sites (Clang)" in Figure 1. In the document text 'Clang' is not explained, nor is it explained which CEL (aggregated over body sites or not aggregated) is used for deriving the risk ratio. Whenever aggregate exposure is mentioned in the final report, it

should be made clear if aggregation over body sites is meant or aggregation over products.
- Further: "calculate acceptable exposure level" above is coloured in grey in Figure 1, but is still part of the hazard quantification and not exposure assessment, and therefore should be violet.
- "Upper use level" is not such a good denomination, because "use" is more often related to use patterns of consumers. "Upper concentration levels" would be more appropriate.
- In Figure 2, the methodology needs to be further clarified. The figure seems to suggest how aggregate exposure can be calculated and further refined. First, in the green box at the top, upper acceptable use levels of fragrances in products are calculated for every single product type (AEL/CEL = 1). However, the text suggests that the "upper use levels" are defined by the full use of the AEL by one product. Then the explanation box suggests that these upper use limits can be used in a conservative approach to aggregate the exposure by inputting them in the Creme model and calculating the resulting exposure for each application site. Since CEL-upper use is equal to AEL for every single product, this would then result in risk as soon as two products are aggregated. It is not clear how this will help the risk assessment.
- The concept of categorisation of the products in broader product categories as well as the use of the product with the lowest upper use level in the aggregate exposure model should be explained in this context and represented in Figure 2 more clearly.
- The refinement of the aggregate exposure assessment is suggested in Figure 2, but a description of this approach (similar to "risk characterization" above) is lacking for the process. Furthermore it is coloured differently than the risk characterisation process above.

SCCS comment to the additional information (Appendix 1)
The Applicant provided in the Supplement (Appendix 1) a revision of Figure 1, summarising the process followed under QRA2. This Figure replaces the original Figure 1. In addition, the original Figure 2 was removed from the new submission. This original figure showed the product SAF as being applied to the individual product CEL rather than as part of the SAFs applied to the NESIL and this appeared to be incorrect. This means that all SCCS comments made in the preliminary Opinion for Figures 1 and 2 are redundant.
Applicant provided two additional new figures (2 and 3, see Appendix 1) that summarize the process for the calculation of the weighted adjustment factors. All three new figures have been evaluated by the SCCS and comments are summarized below.
- The change in Figure 1 from "upper use level" to "upper concentration level" UCL is helpful.
- The new Figure 1 more clearly denotes the general idea of the QRA approach. From the new Figure 1, it also becomes clearer that this approach is more similar to the derivation of maximum residue levels in the area of pesticides legislation than to the safety assessments of cosmetics.
- It is unclear to the SCCS why in the new Figure 2, the old product categories (A-F) are used. Figure 3 uses the new product categories (1-12), but the relationship between the two Figures 2 and 3 is not explained. It is recommended to integrate the figures in a new version of the Final report and provide a clear explanation for both Figures and their relationship.

3.2.1 Hazard Characterisation

From the Final Report on QRA2, 2016
Historically, several animal models have been used to determine the potential for a fragrance ingredient to induce sensitisation. Guinea pig tests (adjuvant and non-adjuvant) were used for many years to assess the inherent contact sensitisation potential of chemicals. These tests can assess potency to a certain extent or antigen cross-reactivity of structurally related chemicals. Later, the murine local lymph node assay (LLNA) was approved by the OECD. This not only determines the potential of an ingredient to induce

contact sensitisation, but also makes further use of these data for assessment of the relative sensitisation potency (EC3 value).
After a thorough review of the data, it was agreed at the IDEA workshops that an interspecies assessment factor for extrapolation from the LLNA to humans was not needed. Strictly speaking, the EC3 value is not a true NOAEL in mice; it provides an indication of potency that correlates very well with the NOAEL in the confirmatory HRIPT. However, given the caution used to ensure that the selected dose levels avoid the induction of skin sensitisation in panellists, in most cases these HRIPTs do not determine maximum no effect levels. This may impact on the quality of the correlation between the LLNA EC3 value and experimental HRIPT NOAELs. The true maximum HRIPT NOAEL is generally somewhere well above the dose levels chosen for this confirmatory test and for ethical reasons, is not determined in the QRA process. The HRIPT, according to strict and harmonised criteria (McNamee *et al*., 20008; Politano and Api, 2008), is used to confirm the 'no effect level' based on the total amount of material applied to the skin expressed as a dose per unit area (e.g. µg/cm^2).

SCCS comment
SCCS acknowledges that several studies demonstrate that on average, EC3 values correlate relatively well with human NOAELs derived from HRIPT and HMT studies. However, different publications show that for some chemicals this correlation is not so perfect. When comparing classifications of fragrances based on EC3 values and human data according to GHS potency categories, the fragrances included in Api et al. (2015) and ICCVAM (2011) show that 6 (11%) and 7 (24%) fragrances, respectively, are classified as other sensitisers based on EC3 values (>2%) but as strong sensitisers based on human data (≤500 µg/cm^2). Likewise, when comparing classification of EC3 values based on the CLP classification, the fragrances included in Api et al. and ICCVAM show that 7 (13%) and 8 (28%) of the fragrances, respectively, are classified as weak-moderate sensitisers based on EC3 values (≥1-≤100%) but as strong sensitisers based on human data (≤500 µg/cm^2). Thus for a significant proportion of the fragrances, the LLNA EC3 value is higher than the human threshold.
A recent publication has shown that applying an interspecies factor is required to ensure that the sensitisation threshold determined in the LLNA does not underestimate the human threshold (Bil et al., 2017).
The SCCS would therefore suggest the inclusion of an interspecies SAF in the absence of human data that could overrule the LLNA EC3.

SCCS comment to the additional information (Appendix 1)
As stated above, SCCS suggests the inclusion of an interspecies SAF in the absence of human data that could overrule the LLNA EC3. The SCCS view is that, based on the available literature for fragrances (Basketter et al., 2018), an interspecies SAF of 3 might be appropriate. However, as expressed in the SCCS Notes of Guidance (9th revision, 2016), when qualitative/quantitative toxicokinetic differences are observed between test animals and humans, e.g. from relevant toxicokinetic data for rat and/or humans (SCCS/1443/11, SCCS/1479/12), the interspecies toxicokinetic default factor can be reduced or enhanced (case-by-case evaluation). Also, other data that may support the necessity of an interspecies SAF may be used in this case-by-case evaluation. For example, data on reaction chemistry or physical-chemical properties of the test chemical may be useful in such an evaluation (Roberts & Api, 2018).

3.2.2 Dose-response or hazard quantification

3.2.2.1 No expected sensitisation induction level (NESIL)

From the Final report on QRA2, 2016
The NESIL is defined as the quantitative threshold exposure level that is considered not to induce skin sensitisation in humans. A Weight of Evidence (WoE) approach is used to

determine each NESIL. WoE introduces a scientifically more valid means for estimating the allergenic potency of a substance for its risk assessment than approaches used in the past. WoE has the advantage over formerly used risk assessment practices, by specifically addressing the elements of exposure-based risk assessment that are unique to the induction of dermal sensitisation while being consistent with the principles of general toxicological risk assessment. WoE is used increasingly by regulatory authorities both in Europe and in the USA (where it is commonly called 'systematic review'). As such, it is a clear improvement over an earlier risk management strategy used by industry, under which each specific fragrance ingredient identified as an allergen was limited to the same concentration across all skin contact product types categoried as either 'leave-on' or 'rinse-off' (Api et al., 2008). The determination of the NESIL, expressed as a dose per unit area (e.g. µg/cm²) is explained in detail by Api et al. (2008) with the scientific rationale to support use of this dose metric described by Kimber et al. (2008).
Briefly, there are several criteria that can assist in determining the NESIL. All the data that are available for a chemical should be considered. Quantitative Structure-Activity Relationship (QSAR) models or in silico models and read-across to data for structurally/mechanistically related chemicals that are determined to be suitable analogues of the chemical of interest can be important. An assessment of all the historical animal and human data is also essential.
For many fragrance raw materials sufficient test data (laboratory animal and human) already exist to allow estimation of skin sensitisation potential and potency classification. These data provide information permitting the establishment of a NESIL.
For newly developed ingredients, information to assess potency, (which is an essential requirement of the QRA), may need to draw on non-animal experiments. Recently, significant advances in the use of non-animal test methods in hazard classification of ingredients have been made. The development of non-animal methodologies to provide information to estimate potency is an area of extensive ongoing research both within the fragrance industry and other sectors.

Human data
Human sensitisation testing is not used in this process to determine hazard, but rather it is used to confirm the lack of sensitisation in the relevant species at a fixed exposure level that has been identified as highly unlikely to induce sensitisation. Human repeat insult patch testing (HRIPT) methodology has a long history of development. In every method a number of potential induction exposures are followed by a rest period and then a challenge exposure. Test volunteers are typically healthy adults who are enrolled without restriction as to sex or ethnicity. The test most typically conducted for confirming the absence of sensitisation responses under consumer relevant conditions is the HRIPT (McNamee *et al.,* 2008).
In HRIPTs, the size of the test population is important with regard to interpretation of findings. The sample size of test subjects must be sufficient so that results are likely to be valid for the population at large, yet small enough to be logistically feasible to conduct the study. For ethical reasons, a HRIPT is only conducted to confirm a dose level that is considered on the basis of solid evidence to be unlikely to cause reactions in the participating volunteers. Despite running many LLNAs and confirmatory HRIPTs, we are not aware of any false negative results (i.e. negative in the LLNA and confirmatory HRIPT, but clinical case reports of positive patch tests). There are certainly materials where there are potency differences between LLNA and HRIPTs (Api *et al*, 2014). A number of factors are incorporated in the protocol to further increase the sensitivity and reliability of the test (e.g. exaggeration through possible minor skin irritation of a test ingredient and use of occluded patches) (McNamee *et al.*, 2008).
To eliminate potential variations in methodology, the industry standard protocol (Politano and Api, 2008) has been adopted as the optimal approach to generate confirmatory human data for use in QRA.
It is generally agreed that HRIPT should not be conducted for hazard identification. Thus, a HRIPT is only conducted to confirm a dose level that is considered to be a NOAEL, where there is adequate data to support that the chosen dose will not result in the induction of

skin sensitisation. A high degree of caution is used to ensure that the dose levels chosen for these tests will not produce reactions in the panellists. The HRIPT is conducted following Good Clinical Practices (GCP), with full informed consent and review by an external ethical review board. RIFM has conducted 71 HRIPTs since 2005 (the first Standards based on the QRA were issued in 2006) on over 7,000 volunteers with only 24 reactions (0.3%) which includes 12 reactions with one material. As such the confirmatory HRIPT is used. It is pertinent to note that the critique by Basketter (2009) related to the testing of formulations and indicated that probably the only ethical use of the HRIPT was to confirm the NOAEL for a substance. He concluded that where there is a specific rationale for testing, for example, to substantiate a no-effect level for a sensitising chemical or to ensure that matrix effects are not making an unexpected contribution to sensitising potency, then rigorous independent review may confirm that an HRIPT is ethical and scientifically justifiable. The possibility that sensitisation may be induced in volunteers dictates that HRIPTs should be conducted rarely and in cases where the benefits overwhelmingly outweigh the risk.

SCCS comment

SCCS has expressed several times its ethical concerns on conducting human skin sensitisation tests, such as the HRIPT (SCCNFP, 2000; SCCP, 2008; SCCS, 2015). One of the concerns is that exposure levels used in the test may themselves cause sensitisation in healthy volunteers. ECHA has expressed similar concerns (ECHA 2016). Although the current experience with the risk on sensitisation due to this 'confirmatory' HRIPT indicates that it is low (0.3%), it is not absent. More importantly, one fragrance material was responsible for 12 of the 24 cases, showing that for this specific material the selected HRIPT concentration was not safe. Both the final report and the accompanying literature do not describe the procedure of arriving at the concentration that has been considered safe to test in the confirmatory HRIPT. According to Api (2008), this concentration is based on hazard assessment data from animal tests, but guidance on how the animal data are extrapolated to the concentration used in the induction phase of the HRIPT is not provided. This should be clarified.

It is questionable why a confirmatory HRIPT is needed. The HRIPT dose is selected based on the scientifically reasoned expectation that none of the 100 volunteers will be sensitised. This dose can then be used for the NESIL. Moreover, when a confirmatory HRIPT in 100 subjects yields the (expected) result of no sensitised individual (i.e. 0%), there is, based on statistical considerations, a confidence interval to be considered. This implies that for a sample of 100, a confidence interval of 95% would include up to 3 individuals (i.e. 3%) who still could have been sensitised (Gefeller, 2013).

In line with ECHA (ECHA 2016), the SCCS emphasises that testing human volunteers for HRIPT data is strongly discouraged. However, where good quality data are already available they should be used as appropriate in well justified cases.

SCCS comment to the additional information (Appendix 1)

The replies of the Applicant on several aspects of the HRIPT have no effect on the ethical concerns of the SCCS concerning the HRIPT. Concerning the conventional sample size of 100, the upper 95% confidence limit of 3% to a zero event rate is a biostatistical fact, quantifying uncertainty, and has nothing to do with the general validity of HRIPT results with respect to actual exposure and sensitisation conditions. Moreover, the additional evidence provided by HRIPT beyond LLNA data appears insufficient. The sole exception could be a false-negative LLNA, where a HRIPT would – contrary to its intention – detect a relatively higher sensitisation potential in terms of inducing one or few human volunteers. The procedure to select dose levels in the HRIPT provided in the Supplement need to be included in a revised version of the Final report on QRA2 (2016).

3.2.2.2 Weight of Evidence Approach (WoE) for determining NESIL

From the Final Report on QRA2, 2016

Historical data for determining the sensitisation potential of an ingredient may be of variable quality and robustness. Therefore, WoE is used, which takes account of all the available

data for the identification of a NESIL, to form the basis of an exposure-based quantitative risk assessment process. WoE approach will be reviewed in the next phase of the IDEA project. When deriving a NESIL, expressed as a dose per unit area, there may be cases where the level derived from a LLNA EC3 value is significantly higher or lower than the level derived from the No Observed Effect Level (NOAEL) obtained in a previously conducted HRIPT or HMT or from read-across or QSAR data. In these cases, a WoE approach may be helpful in deriving a scientifically sound NESIL.

Guideline 1
From experimental investigations and on the grounds of basic immunological considerations, the quantity of chemical per unit area of the skin (e.g. µg/cm²) is considered as the most appropriate "dose metric" for skin sensitisation.

Guideline 2
A NOAEL from a well-run HRIPT will be given precedence over NOAELs from other tests that were conducted in human volunteers (e.g. HMT, earlier precursors to the HRIPT such as the Modified Draize Test), regardless of the NOAELs indicated from those other tests. A well run HRIPT is one which follows the protocol described by Politano and Api (2008) or which is more severe than this in accordance with the critical factors described by McNamee *et al.* (2008).

Guideline 3
Where a Lowest Observed Effect Level (LOAEL; i.e. the lowest dose per unit area which resulted in sensitisation) from other human tests exists (e.g. HMT), which is lower than the NOAEL from the HRIPT, *it will be considered* unless there is some reason to disregard such a LOAEL. In some instances, the conduct of a confirmatory HRIPT to substantiate a NESIL may be warranted.

Guideline 4
In the absence of a NOAEL from a HRIPT, a NOAEL from a different predictive human test (e.g. HMT) can be used to set the NESIL, provided that it is supported by an EC3 value from an LLNA conducted according to OECD Guideline TG 429 (OECD, 2002).

Guideline 5
Adjuvant tests in animals (e.g., GPMT, mouse ear swelling test (MEST)) and non-adjuvant tests in guinea pigs (e.g. Buehler) shall not be used as primary sources for defining NESILs in this context. They may be used to contribute information to determine the potency classification, according to the guidelines provided in the ECETOC, 2003 Technical report No. 87, and be incorporated in a WoE approach.

Guideline 6
When only LLNA data are available, then a confirmatory HRIPT should be considered. A cautious approach should be used for selection of the dose level of fragrance ingredient in the conduct of any such confirmatory HRIPTs including consideration of data on similar ingredients. Under exceptional circumstances (e.g. low volume of use, low use level) the EC3 value (or weighted average where more than one study exists; limited to two significant figures), can be used to define a NESIL or a default NESIL can be applied, based on potency considerations (Gerberick *et al*., 2001). This requires expert judgment.

Guideline 7
A NOAEL from a well-run HRIPT will (even if higher) take precedence over all other NOAELs (including LLNA EC3 values). When there is a significant discrepancy between a HRIPT NOAEL and a LLNA EC3 value (e.g. around an order of magnitude or more), further consideration in setting the NESIL will be required. A LLNA EC3 value that exceeds a NOAEL determined by a HRIPT will not be used to define the NESIL. If the HRIPT NOAEL is the lowest NOAEL available, it takes precedence in deriving the NESIL. Additional sources of data such as guinea pig studies, evaluated as described in ECETOC technical report No. 87,

may provide additional evidence for the purposes of establishing a potency classification. In addition, data elucidating species differences, e.g. studies on metabolism (in the skin), skin penetration and vehicle effects should be considered.

Guideline 8
Data from diagnostic patch test studies cannot be used directly in a WoE approach for the determination of NESILs for the induction of contact allergy to fragrance ingredients. These studies are useful in helping to determine the need for additional data, for example indicating where current exposures to a fragrance ingredient may be a source of clinically relevant positive reactions. The absence of relevant positive reactions following testing in dermatology clinics could be interpreted as evidence that current exposures to the fragrance ingredient are safe.

SCCS comments
Guideline 2 clearly describes that data from a well-run HRIPT are given precedence over other human studies – with the exception of HRIPT studies that are considered more severe in accordance with the critical factors described by McNamee et al. (2008). However, it is not clear if these critical factors are included in the HRIPT protocol of Politano and Api (2008). This should be clarified.

In regard to Guideline 3, it is understandable from a precautionary point of view that LOAEL from other human tests takes precedence over the HRIPT NOAEL if this LOAEL is lower. It is, however, unclear what criteria would be used to disregard such a LOAEL and clarification is needed on the reasoning for disregarding a LOAEL from other human studies.

The proposal in Guideline 4 is appropriate but it is unclear how it will work in the cases where the NOAEL is not supported by the respective EC3 value. This needs clarification.
In regard to Guideline 5, the SCCS agrees that the animal tests mentioned are not designed to provide potency information and for that reason cannot be used as stand-alone methods to derive a potency estimate.

According to Guideline 6, a confirmatory HRIPT should be considered when only LLNA data are available. As expressed before, the SCCS questions the need for a confirmatory HRIPT. The HRIPT dose is selected based on the scientifically-reasoned expectation that none of the 100 volunteers would be sensitised. Therefore, the LLNA data together with an interspecies SAF could instead be used for deriving NESIL. Whilst the SCCS strongly discourages testing with human volunteers, any already available good quality data could be used if appropriate.

Guideline 7 seems to be redundant as it overlaps with Guidelines 2, 5 and 6. Overall, all the Guidelines go in the same direction that data from a well-run HRIPT always takes precedence over other available data, either animal or human. As mentioned before, guidance is lacking on how to proceed where HRIPT NOAEL exceeds LLNA EC3.

In regard to Guideline 8, the SCCS agrees that diagnostic patch test data cannot be used to determine the NESIL. These data may be useful to validate the effectiveness of QRA2 to prevent induction of skin sensitisation. If clinical evidence shows that the QRA2 has failed to prevent induction, the accuracy of the QRA2 methodology should be re-evaluated.

SCCS comment to the additional information (Appendix 1)
The Applicant provided additional clarification of the guiding principles followed in the derivation of the NESIL. According to the Applicant, this new explanatory text needs to be seen as an update of, and replacement for, the original guidance on the WoE approach to derive a NESIL. The SCCS comments made in the preliminary Opinion on the WoE guidelines are therefore redundant and SCCS evaluated the new clarification.
According to the Applicant, in the derivation of the NESIL all available data should be taken into consideration in a weight of evidence approach. The new Figure 4 in Appendix 1

provides a list of (*in silico, in vitro, in vivo)* test methods of which data can be considered in establishing a NESIL. However, not all methods listed provide quantitative data that can be used to derive a NESIL. Some methods can only be used for hazard identification. Furthermore, the explanatory text does not provide any basic rules of interpretation of the test methods listed to derive this NESIL. To conclude, clear explanation of the WoE procedure still needs to be provided.
In the new text, the HRIPT still takes precedence over all other data. Furthermore, the Applicant states that "*Adjustments of thresholds derived from any source other than human to derive a NESIL should be made in the process of derivation of the NESIL, i.e. on the hazard side in the QRA approach and not by application of a generic interspecies adjustment factor to derive the AEL*". We kindly refer back to our SCCS comment in 4.2.1. that suggests a case-by-case evaluation on the necessity of an interspecies SAF, taking into account relevant data to support the derivation.

3.2.3 Sensitisation Assessment Factors (SAFs)

From the Final Report on QRA2, 2016:
Sensitisation Assessment Factors (SAFs) are uncertainty factors are necessary to extrapolate from experimental to real-life exposure scenarios. A detailed explanation of the SAFs originally used in QRA1 was provided in the paper by Api *et al*. (2008). A review of current data supporting the SAFs was conducted by Basketter and Safford (2015).

3.2.3.1 Inter-individual variability

From the Final Report on QRA2, 2016:
A confirmatory HRIPT is a major contributor to the WoE for quantifying the NESIL. This test is carried out on 100 or more healthy volunteers of both sexes and spanning a wide range of ages (18-70). Therefore, the result of the HRIPT and, thus, also the NESIL already implicitly covers a good deal of the variability between individuals. However, these are healthy volunteers and an additional SAF value may be needed. The uncertainty factor or SAF for inter-individual variability allows for possible variations in the sensitivity of individuals within the human population compared to this small sample of subjects in the HRIPT. These factors include genetic effects, sensitive subpopulations, existing disease states, age, sex and ethnicity. While all of these parameters are potentially important, some have more influence than others with respect to the endpoint of skin sensitisation. For example, genetic effects, sensitive subpopulations (including polysensitised individuals) and inherent skin condition are more influential than age, sex, ethnicity and pre-existing disease states (Basketter and Safford, 2015; Api *et al*., 2008; Felter *et al*., 2002; Robinson, 1999). There is little evidence to suggest that subjects with diseased skin (e.g. atopic eczema, psoriasis) have more intrinsic sensitivity to skin sensitisers.
The conclusion from the IDEA Workshops is that to account for differences in sensitivity of individuals within the human population, not accommodated in the NESIL, a SAF of 10 should be applied. (Note: Uncertainty relating to skin state – e.g. presence of irritant dermatitis – is addressed in the section on skin condition)

SCCS comment
The scientific reasoning for a SAF of 10 to account for inter-individual variability has not been made transparent in the QRA2 final report, but more discussion is provided in Basketter & Safford (2016). The available information from human studies (HMT and HRIPT) suggests that human variability of susceptibility to induction of skin sensitisation is likely to span 3-4 orders of magnitude. However, the underlying database seems to be rather weak. The authors further state that by the range of 3–4 orders of magnitude, the majority of variability would be covered (i.e. this would not represent the total variability). Although the SCCS acknowledges that there are limitations in the available human data, these data point to a quite high inter-individual variability. It is argued that some portion of variability might already be covered in the HRIPT (which, however, is performed in only 100 healthy

volunteers and which cannot be quantified). Despite the indications that inter-individual variability might be very high, a considerably lower factor of 10 is suggested as SAF to account for this variability, using the reasoning that most regulatory frameworks use a default factor of 10 for this. The SCCS is of the opinion that the step from an indication of a variability spanning several orders of magnitude to the proposed SAF of 10 deserves a better substantiated justification.

SCCS comment to the additional information (Appendix 1)
Further elaborations were provided to explain the suggested SAF of 10 for inter-individual variability; however, these did not clarify the SCCS concern. Although a default factor of 10 for inter-individual variability is suggested by many regulatory and conceptual frameworks, these also suggest to use substance-specific information to adjust or substitute the default factors (ECHA, 2012; WHO/IPCS, 2005). The available information in Basketter and Safford (2016) indicates that inter-individual variability of induction of skin sensitisation might be well above 10. Therefore it is not self-evident why a value of 10 is suggested in QRA2. A suggested way forward could be to introduce a separate chapter on uncertainties covering this issue (and also other issues on QRA2).

3.2.3.2 Products

From the Final Report on QRA2, 2016:
The consumer can be exposed to fragrance ingredients in many different product forms, which are of varying complexity ranging from aqueous media, simple ethanolic media to multi-phase creams. Under the experimental conditions of a confirmatory HRIPT, exposure to the fragrance ingredient is typically in a simple vehicle. In addition, some of the consumer product formulations may contain ingredients that are irritants or enhance penetration. It is noted however that new and previously overlooked data indicate that enhancement of penetration through the epidermis does not necessarily enhance sensitisation.
The product SAF should take into consideration the role of vehicle or matrix – predicted effect of product formulation versus the experimental conditions. Experimental evidence suggests that the matrix in which the sensitiser is presented to the skin may influence the degree of sensitisation. In considering the appropriate Product SAF it must be remembered that the most common solvents used in the HRIPTs for fragrance ingredients are DEP/ethanol. These solvents are considered to be optimal for the induction of sensitisation in an experimental situation. That said, the experimental data in both animals and humans which supports this is, at best, limited. Thus, for products based on these or similar solvents, a factor of 1 is considered appropriate to account for the matrix. For aqueous based products, (although it is considered possible that the sensitisation potential will be reduced based on observations in the LLNA), it is proposed to maintain a factor of 1 for these products since they are rarely purely aqueous, and will contain other ingredients such as surfactants, that help the product wet the skin.
For solid matrices such as talc or residues on clothing, it is considered that the allergen itself would migrate from the solid substrate to sweat and sebum on the skin. It would then become the matrix from which skin penetration occurs. Given the oily nature of sebum it is proposed to use a factor of 1 for such exposures. A significant factor in the induction of sensitisation is the rate at which the allergen migrates into the sweat/sebum and this should be appropriately factored into the exposure calculation.
It was agreed at the IDEA Workshops that a SAF of either 0.3 or 1 or 3 could be used on a case by case basis (e.g. 0.3 (inert objects with no direct contact, e.g. candles or detergent pods or no vehicle/matrix) or 1 (most products) or 3 (penetration enhancers greater than anticipated from the experimental condition).

SCCS comment
The products SAF is applied to take into consideration the influence of the product matrix on the degree of sensitisation. It is, however, not fully clear to the SCCS, why and when this SAF is applied. According to the final report, and the Basketter & Safford (2016) paper, the

SAF is used to account for the ingredients that enhance skin penetration. However, the final report mentions that "enhancement of penetration does not necessarily enhance sensitisation". This seems to be contradicting the need for the product SAF. In addition, since the CEL is based on the external dose and not on the epidermal dose, the enhancement of penetration does not seem to be relevant. In the final report on QRA2, this SAF is 1 for all product categories, so it is unclear for which products a SAF of 0.3 or 3 is applied.
The SCCS has not been able to evaluate the relevance of this SAF and needs further clarification on the scientific substantiation in this regard. Examples of product categories for which a SAF of 0.3 or 3 is applied would also be useful.

SCCS comment to the additional information (Appendix 1)
The Applicant provided an explanation of why a product SAF is needed. However, the text provided is still unclear on the scientific reasoning. Two different explanations on the need for this SAF are provided. First, it is explained that this SAF is needed to take into account the possible impact of differences between the product matrix and the vehicles used in either the HRIPT or the LLNA on the sensitising potential of the test substance, which relates to the hazard. The second explanation for the need for this SAF is on the possible influence of penetration enhancers in the product, which relates to exposure.
The Applicant states that in practice this SAF is not relevant for cosmetic products that are applied directly to the skin and are solvent-based or water-based. For those products the SAF is 1. For fragrances that are in a solid matrix, e.g. dry facial tissues, a lower SAF (0.3) is applied. There are no products mentioned for which a product SAF higher than 1 is applied Taken it all together, there seems to be no reason to assume that the product matrix has an impact on the skin sensitisation hazard or exposure. This is in line with the conclusion of the Applicant, that "unless a risk assessor has a reason to apply a different SAF, a value of 1 should be chosen". Therefore, the product SAF seems redundant and in order to simplify the method SCCS suggests excluding it from the assessment.

3.2.3.3 Occlusion

From the Final Report on QRA2, 2016:
As a conservative approach the worst case experimental conditions (full occlusion) were applied to all exposure situations and no correction (e.g. use of SAF smaller than 1) is introduced for non-occluded exposures/skin site. Occlusion of the skin (covering the area of application with a dressing) results in multiple effects, including increases in the hydration of the stratum corneum, skin temperature, microbial count, pH, and dermal irritation. The increase in hydration state, in particular, has been associated with increased dermal penetration. The standard test conditions of the HRIPT used to confirm the NESIL employ a series of 24-hour exposures under full occlusion (Politano and Api, 2008). Typically, exposure to fragrance ingredients in consumer products involves a considerably lower degree and duration of occlusion than this. Experimental data indicate that the sensitisation potential from partially occluded or non-occluded exposures may be lower than from full occlusion.

SCCS comment
The SCCS agrees that a SAF for occlusion is not needed.

3.2.3.4 Frequency/duration

From the Final report on QRA2, 2016:
With regard to the period/frequency of exposure, it is likely that many products will be used on a daily basis over extended periods of time. The experimental data from an HRIPT involves nine 24-hour exposures over a three-week period, and it has been questioned whether this is a valid simulation of for longer term use. There is limited experimental

evidence to show that sensitisation may be increased when the normal dosing regimens of predictive tests are extended over longer periods.
It was agreed at the IDEA Workshops that frequency/duration SAF of 3 is sufficient.

SCCS comment
The impact of frequency of exposure on the induction of skin sensitisation is not fully clear due to a lack of relevant studies. Basketter et al. (2006) showed in a human study that frequent exposures to a low concentration of the strong skin sensitiser PPD resulted in a higher rate of sensitisation than less frequent high dose exposure. This study supports the need for an SAF to account for frequent exposure. Ubiquitous use of fragrance materials in a broad range of consumer products makes frequent exposures likely, and the SCCS agrees that a SAF to account for this needs to be applied. No scientific rationale is provided supporting the factor of 3 that is assigned to account for this uncertainty. The SCCS is aware that other regulatory frameworks use higher factors for covering this type of duration extrapolation.

SCCS comment to the additional information (Appendix 1)
The SCCS agrees with the Applicant that the frequency SAF of 3 is needed to cover for uncertainties on the impact of frequent exposure to fragrances on skin sensitisation hazard. The lack of scientific knowledge on this aspect makes it impossible to exactly determine the factor needed to cover for this. A way forward is to assign a SAF of 3 if the uncertainties related to this are clearly described in a separate chapter on uncertainties.

3.2.3.5 Skin condition

From the Final report on QRA2, 2016:
There is little evidence from the scientific literature that particular skin areas of the body are inherently more prone to the induction of skin sensitisation than others. However, the presence of compromised/inflamed skin may have an effect. The HRIPT is conducted on non-inflamed and intact skin, whilst consumers in the population at large may have compromised/inflamed skin due to a number of factors. In addition, there is little evidence that compromising the skin barrier by physical or chemical means increases the potential for the induction of sensitisation. However, the generation of inflammation in skin, particularly from contact with irritant chemicals (such as sodium lauryl sulfate or skin with active irritant contact dermatitis), may increase sensitivity to skin sensitisers.

Table 1: Summary of skin condition SAFs based on Body Site (Table 3 in the Final report on QRA2)

Body Site	Additional definition for this study	Skin Condition SAF
Scalp		1
Face	Does not include: Eyes, Lips, Mouth, Behind Ears	3**
Peri-ocular	The eyelid and surrounding skin.	3**
Lips		3**
Intraoral	"Buccal" / "Inside Cheek"; Does not include: Lips	3**
Neck	Does not include: Behind Ears	3**
Behind Ears		1
Chest	Does not include: Axillae, Abdomen	1

Abdomen		1
Back	Does not include: Axillae	1
Axillae		10
Arms	Does include: Shoulder, Forearm, Upper arm; Does not include: Wrists, Hands, Palms, Axillae	1
Wrists		3**
Back of hand	Does not include: Palms, Wrists	3**
Palms		3**
Ano-genital		10
Legs	Does include: Buttocks, Thighs, Calves; Does not include: Feet	3**
Feet		3**

*In order to conduct the risk assessment considering aggregate exposure, the skin condition SAFs are aligned with the list of application sites from survey data. **Note: for practical purposes the number 3 approximates 3.16 or the half log of 10

It is recognised that certain skin sites are more prone to inflammation than others, and that the SAFs may therefore vary between sites.
A SAF of 1, 3 or 10 should be assigned based on the susceptibility of the skin site to inflammation. Table 3 (*Table 1 in this opinion*) details SAFs used for each skin site, and Table 4 (*Table 2 in this opinion)* provides the rationale for applying skin condition SAFs to various products.

Table 2: Rationale for Skin Condition SAF (Table 4 in the Final report on QRA2)

Product Type	Rationale for Skin Condition SAF
Deodorants & Antiperspirants of all types including fragranced body sprays	The SAF is 10 as these products are applied to the axillae where the skin is easily irritated due to a combination of factors including the unique environment of the axillae (humid, oil rich sebum production and site for perspiration). There may also be acute transient irritation due to product application or mechanical irritation. Shaving may produce an acute transient response.
Hydroalcoholic Products (eau de toilette, parfum etc.)	The area is the neck, wrists, antecubital fossa. Irritation from shaving may produce an acute transient response. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Eye Products (Includes: eye shadow, mascara, eyeliner, eye make-up)	The SAF is 3* because product is applied to the peri-ocular site and face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation

Product Type	Rationale for Skin Condition SAF
Body Creams, lotions	The SAF is 10 because the area is the entire body which may include areas of inflamed skin, i.e.: intimate regions and axillae. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Hand cream	The SAF is 3* because the product is applied to the hands. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Facial Cream (Moisturizing)/Facial Balm	The SAF of 3* has been attributed because the product is applied to the face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Women's Make up (Foundation)	SAF is 3* because the product is applied to the face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Make-up remover	SAF is 3* because the product may be applied to eyelids (peri-ocular region) and face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Lip Products	A SAF of 3* is applied because the site is applied to the lips (highly vascular and there is exposure to mucous membranes and possible exposure to dry or chapped lips). Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Hair styling aids (mousse, gels, leave in conditioners)	The SAF is 3* because when the product is applied to the hair there will also be exposure to the scalp and the palms of the hands. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Hair sprays	The SAF is 1 because it is applied to the scalp. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Shampoo	The SAF is 10 because the product is applied to the head (hair) and scalp with the hands and may also be used over the entire body as a shower gel. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Body wash/shower gels	The SAF is 10 because product may be used all over the body including intimate regions and axillae. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Conditioner (rinse-off)	SAF is 3* because the product is applied to the head (hair) and scalp with the hands. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Bar soap	The SAF is 10 because product may be used all over the body including the axillae and intimate regions. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation

Product Type	Rationale for Skin Condition SAF
Liquid soap	The SAF is 3* because product may be used on the hands and face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Face washes, gels, scrubs	The SAF of 3* has been attributed because the product is applied to the face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Bath gels, foams, mousses	The SAF is 10 because product may be used all over the body including intimate body regions and the axillae. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Toothpaste	The SAF is a 3*. The sites are the lips and mouth. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Mouthwash	The SAF is a 3*. The sites are the lips and mouth. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation

*Note: for practical purposes the number 3 approximates 3.16 or the half log of 10.

SCCS comment

The rationale provided for the skin condition SAF is not clear. Table 2 (Table 4 in the Final report on QRA2, 2016) is based on the product types used. In Table 2, the sentence "No additional contribution to skin condition is expected from product irritation" has been used for each product. This needs further explanation, because some matrices may contain ingredients that have irritant properties. In such cases, it would be more logical to cover these matrix effects in the product SAF, which according to the final report, only covers penetration enhancers. It would also be more logical to apply the skin condition SAF only for those skin sites that have more susceptibility to inflammation, irrespective of the products used.

The final report provides no scientific justification as to why certain body parts are considered more susceptible to inflammation, e.g. for axillae it is only mentioned because it is an intimate region without any further reasoning. Also, no explanation is provided why an SAF of 3 is assigned to a large proportion of the skin sites, which implies that many body sites are susceptible to inflammation. Furthermore, a SAF of 10 is used for product types that are used all over the body, without any justification why such products need such a high SAF. It is unclear why different SAFs have been used for bar and liquid soap, which both could be used all over the body. SCCS does not understand why face scrubs have the same SAF as face gels and face washes, whereas body scrub has been missed out altogether. A better description of the rationale for the SAFs needs to be provided in Table 2.

QRA2 uses 18 body sites, which is very specific and detailed. A rationale as to why so many different body sites are used in the methodology needs to be provided.

SCCS comment to the additional information (Appendix 1)

The Applicant provided further explanation on the skin condition SAF and provided a new Table 2 (Appendix 1). According to this Table, the SAF relates to the body sites where the products are applied to the skin. Hence, the SAF is not referring to certain skin conditions, but to the site of product application, which only partly entails certain skin conditions, like

increased hydration in intertriginous areas such as ano-genital or axillary. Therefore, SCCS recommends renaming this SAF to avoid any confusion.
The sentence "*No additional contribution to skin condition is expected from product irritation*" is still used in the revised Table 2. SCCS has already expressed that this sentence is unclear and confusing and suggests removing it. To be able to understand why this SAF is applied, a revision of the text in the Final report is clearly needed. From the explanation provided in Appendix 1 it appears that this SAF is applied to cover uncertainties related to several aspects that may increase the hazard, including irritants present in the product and sensitive skin. The SCCS agrees that a SAF of 3 for all products and a SAF of 10 for the most sensitive sites can be applied to account for these uncertainties.

The Applicant provided a rationale for the selection of the 18 body sites. This rationale is based on practicability: the 18 body sites represent the most detailed partitioning provided by the Kantar database, which is used also by the Crème global model. However, in addition to these practical considerations a scientific rationale is needed, so that these body sites can be used in an assessment. Such a scientific rationale may include differentiation according to skin properties, occlusion levels, product types etc. Also, it should be considered when and how exposure of different body sites may need to be aggregated (e.g. in the case for palms and back of the hands, which presumably drain to the same lymph nodes).

3.2.3.6 Defining SAF numbers

From the Final Report on QRA2, 2016:
The total SAF is calculated by multiplying the factors assigned to account for inter-individual variability, product effects, frequency of exposure and skin condition SAFs. As in other areas of toxicology, for each substance, careful consideration should be given to the appropriateness of applying a particular uncertainty factor (SAF).

SCCS comment
As already mentioned on page 9, the SCCS does not agree with a determination of the total SAF at the level of the hazard assessment. Since product-specific SAFs are defined, this poses a problem in an aggregate exposure assessment in the case that different SAFs apply for the different products. It would be more appropriate to apply the product-specific SAFs to the single exposures that are later aggregated.

SCCS comment to the additional information (Appendix 1)

According to the new information provided by the Applicant (new Figure 1), the SCCS understands that in the approach presented the overall SAF is applied for each product category separately, and only the upper concentration levels are aggregated to derive the final concentration limits for each product.

3.3 Exposure

3.3.1 Dose metric

From the Final Report on QRA2, 2016:
The measurement of exposure ("dose metric") recommended for use in skin sensitisation risk assessments for fragrance ingredients is dose/area (µg/cm2). There is a difference between the applied versus the delivered dose since there are factors that can affect the effective amount of ingredient delivered to the viable epidermis such as evaporation, binding/sequestration in the skin, metabolism (inactivation and activation). For the purposes of QRA, the applied dose is used as a conservative estimate of actual consumer exposure.

Based upon the understanding of the immunological mechanism involved, it is logical to assume that for an immune response to be initiated, a certain number of Langerhans cells (LC) are required to be activated in order to initiate the cascade of events leading to the threshold of induction for skin sensitisation being exceeded. This would suggest that for the induction of contact allergy, the application of an amount of allergen expressed as percent (weight/volume) is not as important as understanding both the dose applied and the surface area over which the allergen is applied. This has been thoroughly reviewed by Kimber *et al*. (2008) and has been established as an acceptable approach (Ter Burg, *et al*., 2010; ECHA, 2012).

3.3.2 Consumer exposure level (CEL)

From the Final Report on QRA2, 2016:

Estimation of the Consumer Exposure Level (CEL) is an essential element of the QRA. Below we discuss the use for this purpose of Creme probabilistic aggregate exposure model to assess this. It is important to understand how consumers are likely to be exposed to fragrance ingredients from their use of the consumer products. Exposure levels occurring under intended and foreseeable conditions of use, but not deliberate misuse are addressed. The calculation of consumer exposure must include parameters such as frequency, use practices (e.g. how a consumer actually uses the product), duration of use, amount of product used per application/use and level of fragrance in product.

There are limited consumer habits and practices data for children, which are inadequate for probabilistic modelling. In addition, there are data to show that children are not more susceptible to skin sensitisation than adults (Cassimos et al., 1980; Epstein, 1961). Skin sensitisation is linked to exposure. In the application of the QRA (Api et al., 2008), products designed for children (e.g. baby care consumer products, diapers) were considered in the SAF assignments.

The experimental evidence appears to show that young children are less easy to sensitise, so that a risk assessment for adults is conservative for children. A review on developmental immunotoxicology and risk assessment by Holsapple et al. (2004) concluded that current risk practices have been generally shown to be sufficient in protecting children (> 6 months old) and an additional safety factor is not needed to provide additional protection from that which is already achieved. Another review by Militello et al. (2006) finds that the risk of sensitisation appears to increase with age, which may be linked to an increase in exposure.

It should be noted that the CEL defined within this dossier addresses consumer products that are bought for personal use. Occupational/professional exposure is not included at this time because comprehensive habits and practices data are not available. It will be important to address occupational/professional exposure in the QRA approach when these exposure data become available. This is explored in recommendations for further refinement (see Section 3). Cross-reactivity appears to be an uncommon occurrence except with very closely related structures. When there are materials that cross-react, then the NESIL for the most potent material within the class is applied to all the materials. The levels of any such material cannot exceed the limit dictated by the QRA (i.e. the IFRA Standard on Rose Ketones).

In the approach described here, dermal aggregate exposure is considered after the QRA-derived Upper Limit for acceptable consumer exposure level (AEL/CEL ratio = 1) for the fragrance ingredient is estimated.

It is equally important to have accurate data on human parameters such as the body surface area over which the product is used. Skin penetration is not specifically addressed in measuring consumer exposure since the dose metric is unit weight applied per unit area of skin. As such, using a conservative approach, the applied dose is taken to be the delivered dose. In the case of reliable information on skin penetration rates the conservative approach can be modified.

Using these criteria, the data sources listed in Table 7 were used in the calculation of CEL. A hierarchy was established for selecting data based on quality and scope. When measured data for the same product type were available from more than one source, then the most

conservative value (i.e. the highest value) was used unless there was a sound scientific rationale for using data from another source.
Examples:

1. Hall et al. (2007) exposure study data were used in preference to the data published in Loretz et al. (2005) on the basis that the Hall et al. (2007) study participants used their own products rather than products supplied by the study investigator as in the CTFA study leading to more realistic use.
2. Cowan-Ellsberry et al. (2008) deodorant/antiperspirant data were used instead of those of Loretz et al. (2006) and Hall et al. (2007) because Cowan-Ellsberry et al. (2008) used measured 90th percentile exposure (amount) and surface area data and integrated them into a per diem exposure.

All of these sources of exposure data listed below use information of varying detail and completeness. This means that the robustness of the exposure data can also be different. For these reasons when evaluating a distribution of exposure data, the same percentile data point cannot be selected for each set of exposure data. For example, the 90th percentile was chosen from the Hall et al. (2007; 2011) and Loretz et al. (2006; 2008) exposure studies to define the most appropriate exposure level given the conservatism in the models. On the other hand, whilst the study conducted by Tozer et al. (2004) and Cano (2006) measured distribution of amount, frequency of use and surface area it was not overly conservative like the Hall et al. (2007; 2011) studies. On this basis it was more appropriate to choose a higher percentile from this study and therefore the 95th percentile was chosen. The individual references used to define the consumer exposure to different product types are detailed in Appendix 2. When introducing dermal aggregate exposure in the QRA, single point values for the habits and practices data are not used. The full distribution of exposure data were built in to the Creme RIFM Aggregate Exposure Model.

SCCS comment
Susceptibility to skin sensitisation of children should not be discussed under the exposure chapter of the final report on QRA2, because this relates to hazard.
Furthermore, it is unclear to the SCCS what "full distribution of exposure data" entails. Specifications are needed on which parameters are meant, and what use patterns are considered. It is also not clear whether this section only describes single-product exposures (since aggregate exposure is considered in the next chapter), and what the general purpose is of this chapter.
The Table in Appendix 2 (Table 7 in the final report) has the title "Summary of available habits and practices…product types". It is unclear whether this Table lists the parameters recommended by QRA2 for single-product use, or whether these data are also used for the "Creme RIFM Aggregate Exposure Model".
From the beginning of the chapter on exposure, one gets the impression that the Creme RIFM aggregate exposure model, which according to SCCS knowledge is a probabilistic model based on distributions for selected input data, represents an integral part of the QRA2 approach. However, Appendix 2 only lists point values for the input data and not distributions and only the next chapter lists different data on which the Creme RIFM Aggregate Exposure Model seems to be based. The methodology used for exposure assessment under QRA2 therefore needs to be clarified.

SCCS comment to the additional information (Appendix 1)
The SCCS appreciates that the Applicant agrees that susceptibility to skin sensitisation of children should be discussed under hazard considerations. We recommend revising this in a new version the Final report on QRA2.
From the new information provided by the Applicant the questions raised above by the SCCS are not answered. The previous SCCS comments and concerns remain.

3.3.3 Consideration of dermal aggregate exposure

From the Final Report on QRA2, 2016:

Consumers generally use several products each day, and some of these will be applied to the same skin site. If these products contain the same fragrance ingredients, then it becomes important to consider aggregate exposure when conducting the risk assessment for skin sensitisation. In order to incorporate dermal aggregate exposure in the QRA for ingredients, it is necessary to account for the products applied to each body site. The methodology reported here is focussed on assessment of exposure in cosmetics. It does not include aromatherapy, drugs and topical treatments, massage and spa therapies, occupational exposure, natural exposure, foods as the necessary data base is lacking still.
Since 2010, the Research Institute for Fragrance Materials has been developing a model to estimate the aggregate exposure to fragrance ingredients resulting from the use of consumer products. This model has now been modified for use in dermal QRA2 for sensitisation. Creme Global (www.cremeglobal.com) is their well-established partner in modelling exposure to cosmetics and foods, and their exposure methodologies are used by regulatory bodies such as SCCS (SCCS, 2014) and EFSA (Vilone *et al*., 2014) and a trade association (Cosmetics Europe, previously COLIPA; Hall *et al*. 2007, 2011; McNamara *et al*., 2007).
The Creme RIFM Aggregate Exposure Model is based on declared habits and practices data from 36,446 panellists across Europe and The United States of America (Kantar Database, 2011), also described in Comiskey *et al*. (2015) and Safford *et al*. (2015). Each panellist supplied diary data on which cosmetic products were used during the day for seven consecutive days, as well as information on the application sites of most products. The model uses probabilistic (Monte Carlo) simulations to allow full distributions of data sets, providing a more realistic estimate of aggregate exposure to individuals across a population (Comiskey *et al*., 2015; Safford *et al*., 2015) compared to a deterministic aggregate approach. An overview of the Creme RIFM aggregate exposure model is provided in the Appendix 4 in the final report. Output from the model provides dermal exposure as amount of product and/or fragrance per skin surface area (µg/cm^2) for different body areas for the highest use day for each consumer and also assumes a fragrance material is always present in every product, these assumptions are considered conservative. In order to select an appropriate percentile to use for risk assessment purposes, the probabilistic aggregate exposure model design is considered. The 95th percentile of exposure is used as standard in many domains of regulatory risk assessment, and is considered appropriate in this case, particularly in light of the conservative nature of the Creme RIFM aggregate model.
An example of such conservatism in the model is that dermal aggregate exposure is calculated using the assumption that the fragrance ingredient is present in all products at the QRA2 upper use level (concentrations). This leads to an aggregate Consumer Exposure Level (CEL_{Agg}) that exceeds the Acceptable Exposure Level (AEL) i.e. $AEL/CEL_{Agg} < 1$ in some instances.
As such, this section of the dossier is intended to explain the proposed methods of reducing the QRA derived upper use levels so that when aggregate exposure is considered, the AEL is not exceeded. The proceeding sections describe a method to reduce the fragrance concentrations in product types and categories based on their relative contribution to aggregate exposure.

SCCS comment
The SCCS notes that the title "Consideration of dermal aggregate exposure" is not explanatory enough and it is not clear what the difference in content is when compared to the preceding chapter of the final report on QRA2. Presumably the difference is that aggregate exposure has been considered here. Both chapter titles need to be changed to clarify this more explicitly.
As mentioned before, the methodology used for exposure assessment needs to be described in more detail, especially regarding which parameters are treated probabilistically and which as point values.
From Appendix 4 in the final report these parameters are not clear, e.g. how concentrations in products were treated by the model and how these distribution values have been derived.
It is important to point out that the Creme RIFM Aggregate Exposure Model uses the common methodology of probabilistic exposure assessment, which is currently not used or

recommended for cosmetics by the SCCS. This new emerging methodology needs further evaluation by the SCCS.

SCCS comment to the additional information (Appendix 1)
From the new information provided by the Applicant the questions raised above by the SCCS are not answered. The SCCS comments remain valid.

3.3.3.1 Deriving QRA2 upper use levels

From the Final Report on QRA2, 2016:
Initially the QRA upper use levels were calculated deterministically, based on the NESIL for the fragrance material, the total SAF for each product and application site (explained in the accompanying document) and the high percentile product exposure to each application site (Api et al., 2008). In the present proposal an example of such reverse calculations of the upper use levels were made for the fragrance Citral (**Error! Not a valid bookmark self-reference.**), using the following formula:

$$Upper\ Use\ Level\ (\%) = \frac{NESIL(\mu g/cm^2)}{1{,}000 \times Total\ SAF \times Exposure(mg/cm^2/day)} \times 100$$

Table 3: Derived QRA2 Upper Use Levels for Citral by Product Type (Table 8 in the Final report on QRA2)

	Citral NESIL = 1400 µg/cm²		
Product Type	**Proposed Total SAF for QRA2**	**Exposure mg/cm²/day**	**QRA2 product type upper use levels**
Deodorants and antiperspirants of all types including fragranced body sprays	300	9.10	0.05%
Hydroalcoholic products (eau de toilette, parfum etc.)	100	2.21	0.63%
Body creams, lotions	300	0.60	0.78%
Hand cream	100	2.60	0.54%
Facial cream (moisturizing)/facial balm	100	2.80	0.50%
Eye products (Includes: eye shadow, mascara, eyeliner, eye make-up)	100	2.17	0.65%
Women's make up (foundation)	100	0.92	1.52%
Make-up remover	100	0.90	1.56%
Lip products	100	11.80	0.12%
Hair styling aids (mousse, gels, leave in conditioners)	100	0.4	3.50%
Hair sprays	30	2.20	2.12%
Shampoo	300	0.17	2.75%
Body wash/shower gels	300	0.015	31.10%
Conditioner (rinse-off)	100	0.2	7%
Bar soap	300	0.2	2.33%
Liquid soap	100	0.2	7.00%
Face washes, gels, scrubs	300	0.15	3.11%
Bath gels, foams, mousses	300	0.01	46.67%
Toothpaste	100	1.27	1.10%
Mouthwash	100	1.00	1.40%

Based on these calculations, it was found that in many cases the upper use levels far exceeded realistic industry use levels (e.g. body wash/shower gel, 31.10%; Table 8 (*Table 3 in this Opinion)*) due to the assumption that some products are used evenly all over the

body leading to a reduced exposure per unit surface area which affords them a greater QRA2 upper use level. On the other hand, products that are assumed to be used on specific parts of the body (e.g. deodorants used on axillae, 0.05%; Table 8) their calculated QRA2 upper use levels are lower due to the reduced surface area with which they are applied. When the QRA2 upper use levels were input into the Creme RIFM Aggregate Exposure Model, it was found that many of the product types produced a CEL_{Agg} that exceeded the AEL for specific applications sites. This was due to product co-use and the fact that subjects in the habits and practices survey applied products in a way that is contrary to the QRA2 upper use levels assumptions e.g. shower gel used on palms and face only.
Moreover, the disparity in upper use levels between products (*cf*. bath gels and deodorants; Table 8) in the Creme RIFM aggregate exposure model resulted in all product types requiring a large reduction in upper use levels for specific application sites, despite the fact that only some products were driving the aggregate exposure. To rectify this issue, it was decided that the product type with the lowest upper use level from their designated product categories would be used in the aggregate exposure model for all products in their category, where products with similar exposure and SAF were grouped together (Table 9, *Table 4 in this opinion*).
Importantly, these categorised (lowest) upper use levels were considered to be more realistic in terms of proximity with industry use levels, based on expert judgment. Thus, each of the product types in their categories had the same (lowest) upper use level, and the exposure results from each individual product type were aggregated by product category. It should be noted that not all the product types are available in the Creme RIFM model, for example, eye products, make-up remover and bath gels. Using the conservative assumption that, for a given category, the upper use level is acceptable then for a given category the product types not in the model can be assumed to have the same low concentration.

Table 4: Upper Use Levels for Citral in Product Types and Product Categories (Table 9 in the Final report on QRA2)

Product Type	QRA2 product type upper use levels	Product Categorization	QRA2 categorized upper use levels
Deodorants and antiperspirants of all types including fragranced body sprays	0.05%	A	0.05%
Hydroalcoholic products (eau de toilette, parfum etc.)	0.63%	B	0.63%
Body creams, lotions	0.78%	C	0.50%
Hand cream	0.54%		
Facial cream (moisturizing)/facial balm	0.50%		
Eye products (Includes: eye shadow, mascara, eyeliner, eye make-up)	0.65%	D	0.12%
Women's make up (foundation)	1.52%		
Make-up remover	1.56%		
Lip products	0.12%		
Hair styling aids (mousse, gels, leave in conditioners)	3.50%		
Hair sprays	2.12%		
Shampoo	2.75%	E	2.33%
Body wash/shower gels	31.10%		
Conditioner (rinse-off)	7%		
Bar soap	2.33%		
Liquid soap	7.00%		
Face washes, gels, scrubs	3.11%		
Bath gels, foams, mousses	46.67%		
Toothpaste	1.10%	F	1.10%
Mouthwash	1.40%		

SCCS comment

In addition to the example of Citral case, a general description of the approach to calculate aggregate exposure is needed. The methodology for deriving the upper use levels needs to be better described. For example, a careful recalculation has shown that Table 3 (Table 8 in the Final report on QRA2, 2016) has listed the upper use levels derived for single products, whereas it is mentioned later in the text that "QRA2 upper use levels assumptions" encompass that "e.g. shower gel is used on palms and face only." It is not clear why this is an important factor for the differences between the product-specific calculation and the aggregate exposure calculation if no aggregation for body parts is done in Table 3. Furthermore, Table 3 should contain all assumptions used and the references for the exposure value, and also that shower gel is considered to be applied only on palms and face. There is also a need to better explain how the product categories A-F were derived. For example, it is not clear what "same exposure" means, and what is common between exposure to make-up remover and lip balm (apart from the obvious application to the face). It is stated that aggregation has been done for each product category separately (last paragraph). From the following sections, it seems that aggregation was also done for different categories (Figure 3; Figure 4 in final report on QRA2, 2016) and this needs to be clarified. Explanation is also needed on why SAFs are different for the same application site for conditioner (100), hair spray (30) and shampoo (300) when in the following chapter, the matrix is always SAF=1.
Many text passages are difficult to understand and do not seem logical; e.g. "in many cases the upper use levels far exceeded realistic industry use levels (e.g. body wash/shower gel, 31.10%; Table 3) due to the assumption that some products are used evenly all over the body leading to a reduced exposure per unit surface area which affords them a greater QRA2 upper use level". It is not clear why the assumption of an even spread of products on the skin is not realistic, and if it is not realistic, why it was not adjusted accordingly. Further down in the paragraph it is described that "subjects in the habits and practices survey applied products in a way that is contrary to the QRA2 upper use levels assumptions". It is not clear then why the QRA2 assumptions were not revisited after this reality check, rather than remediating in a somewhat arbitrary way with the "categorized upper use levels".
From a methodology point of view the SCCS notes that upper use levels for single products are of limited value for use in QRA2, since fragrances will always be applied in multiple products. Therefore, the product-specific "upper use levels" have little relevance, except to derive maximum levels for risk management. In the text, it should be clarified that they only serve as a starting point for assessing upper use levels based on an aggregate assessment. Also, in the QRA2 report, the upper use level approach has been described as part of the exposure assessment. Instead, they are a means for risk management, and the header or structure of the chapter needs to be changed accordingly.

SCCS comment to the additional information (Appendix 1)
The Applicant provided a new Table with new product categories (new Table 1, Appendix 1), together with an explanation of the determination of the adjustment factors (new Figure 3). SCCS has evaluated this new information and concluded that a better explanation of these new product categories is still needed.
As mentioned earlier, it is confusing that the new Figure 2 still refers to the previous product categories (A to F) whereas the new Figure 3 refers to the new product categories (1 to 10).
The Applicant provided more explanations regarding the exposure assessment with the CREME RIFM aggregate exposure model. This information is still too scarce, fragmented and not targeted enough. It is unclear whether the CREME RIFM model is an integral part of the QRA2 approach, and how the modelling is conducted in detail. A rationale is missing for the 25 product categories (called "products") used in the model that are further aggregated into 9 larger categories, and it is unclear why information on the ongoing work on the CREME RIFM model is included. This suggests that the model may be changed in the near future. The impact of this ongoing work on the current model is unclear and needs to be better explained.

3.3.3.2 Aggregate exposure risk assessment with upper limit use levels

From the Final Report on QRA2, 2016:
The categorised upper use levels were input into the Creme RIFM aggregate exposure model to estimate the 95th percentile CELAgg for each of body 18 application sites (Table 10; *Table 5 in this Opinion*). The AEL for Citral was calculated for each body application site. This first required the calculation of the total SAF, which is the summation of four SAFs: 1) inter-individual, 2) matrix, 3) frequency and 4) skin condition. The ratio of the total SAF to the NESIL for Citral was calculated to give the AEL (AEL = NESIL/Total SAF). Finally, the AEL/CELAgg could be calculated to determine if the ratio was above or below 1, where a value greater than 1 indicated that the CELAgg did not exceed the threshold set by the AEL.
It was found that four body application sites had an AEL/CELAgg below 1, which suggests that the Citral concentration (upper use level) should be lowered; lips, intra-oral region, palms and the axillae (Table 10). Lips had the lowest AEL/CELAgg (0.45), intra-oral region had the second lowest (0.48), followed by palms (0.63) and axillae (0.65). All other products had an AEL/CELAgg greater than 1. Therefore, the upper use level of Citral in the products applied to these application sites needed to be reduced such that their AEL/CELAgg were above 1.

SCCS comment
The title of Table 5 (Table 10 in the Final report on QRA2, 2016) needs to mention that this is an example using Citral. Also, "exposure risk assessment" should be changed to "exposure assessment".
In this approach aggregation is done for each of the 18 body sites separately. As mentioned earlier, the Final Report on QRA2 needs to explain the rationale for designating these different body sites.
SCCS acknowledges that a first step in aggregate exposure assessment is aggregation of all products that are used on the 18 specific body sites based on the products used during that day. In the Final Report on QRA2, aggregation over multiple body sites is not mentioned, although this may be a determinant in the risk on skin sensitisation as well. It is well understood that priming of the immune system takes place at the level of the lymph nodes. After dermal absorption, skin sensitisers will induce a local immune response in the skin followed by dislocation of Langerhans cells to the draining lymph nodes and finally followed by T cell priming in the lymph nodes draining the exposed area. It is possible that exposure of different body sites used in QRA2 will target the same lymph node, e.g. the hands are divided in three different body sites but presumably target the same lymph node if exposure is on one side of the body. Evidence for this is provided in a human study from Kligman (1966), who showed that four sequential exposures on different but adjacent sites on one extremity (arm or leg) was far more effective to induce sensitisation than four sequential exposures to each of the four extremities. Kligman concluded that "*bombardment of the same lymph node is superior to stimulation of four different nodal systems".*
Mirroring this process within the aggregated exposure assessment is not trivial, since this process is not fully understood. It may impact the risk assessment and may lead to an underestimation of the risk. SCCS finds it of great importance to address and discuss this uncertainty in the Final Report on QRA2 as well.

SCCS comment to the additional information (Appendix 1)
In response to questions regarding attribution of skin sites and aggregation (section 12 of Appendix 1) the Applicant provides the following: "*It is true that the events in regional (draining) lymph nodes [recognition by responsive T lymphocytes of processed antigen, and T lymphocyte activation and proliferation] are critical elements for the acquisition of skin sensitization, and also determine the level of sensitization that is induced. The exception to this rule is where the area of exposure to the inducing chemical allergen is less than 1cm^2, in which circumstance the area becomes important. It is well established that - under most conditions of exposure - the important metric in determining the extent of sensitisation that will develop is the dose per unit area of chemical allergen for single exposures, at a given Dose Per Unit Area (DPUA) above 1 cm^2, size of area is not a factor – so at that DPUA,*

applying something once, whether on a small or large area, will have (does have) the same effect.".
Regarding the lacking importance of exposed skin area per application beyond 1 cm², the SCCS points out that results referred to have been obtained using very potent sensitisers, such as DNCB (Friedmann, 2007). It is unclear whether the results are fully transferable to weaker allergens, such as most fragrances, where stochastic effects may increase sensitisation risk if a larger skin area is exposed, compared to a smaller one. This is certainly an area needing further research.
The new comments of the Applicant thus do not abate the concern of the SSCS that body sites draining to the same lymph nodes may need to be aggregated. The SCCS suggests to include a chapter describing uncertainties of this and other aspects of the methodology.

Table 5: AEL/CEL_{Agg} for Application Sites, Ordered from Lowest to Highest (Table 10 in the Final report on QRA2, 2016)

Application site	Inter-individual SAF	Matrix SAF	Frequency SAF	Skin Condition SAF	Total SAF	NESIL	AEL (NESIL/ Total SAF)	CEL_{Agg}	AEL/CEL_{Agg}
Lips	10	1	3	3	100	1400	14	31.1	0.45
Intra-oral	10	1	3	3	100	1400	14	29	0.48
Palms	10	1	3	3	100	1400	14	22.3	0.63
Axillae	10	1	3	10	300	1400	4.7	7.22	0.65
Back of Hand	10	1	3	3	100	1400	14	8.93	1.57
Face	10	1	3	3	100	1400	14	8.37	1.67
Neck	10	1	3	3	100	1400	14	6.35	2.2
Ano-genital	10	1	3	10	300	1400	4.7	1.61	2.9
Scalp	10	1	3	1	30	1400	46.7	9.77	4.78
Wrists	10	1	3	3	100	1400	14	2.8	5
Feet	10	1	3	3	100	1400	14	2.65	5.28
Peri-ocular	10	1	3	3	100	1400	14	2.36	5.93
Behind ears	10	1	3	1	30	1400	46.7	4.16	11.22
Legs	10	1	3	1	30	1400	46.7	2.15	21.72
Arms	10	1	3	1	30	1400	46.7	1.71	27.29
Chest	10	1	3	1	30	1400	46.7	1.52	30.7
Abdomen	10	1	3	1	30	1400	46.7	1.52	30.7
Back	10	1	3	1	30	1400	46.7	1.51	30.91

3.3.3.3 Use of aggregate exposure assessment for adjusting upper use levels

From the Final Report on QRA2, 2016:
In this section, the method of reducing the upper use levels in the product types that were applied to the four applications sites, whose AEL/CEL_{Agg} was less than 1 (lips, intra-oral, palms and axillae) is described.

SCCS comment
In the Final Report, several examples were given. In this Opinion, SCCS will only include the example for products applied to the lips.

3.3.3.4 Adjust Upper Use Levels in Products Applied to the Lips

From the Final report on QRA2, 2016:

For the case of adjusting the upper use level in products applied to the lips, there were four product categories to adjust (F, D, C, E), and therefore only four upper use level values (1.10%, 0.12%, 0.5%, 2.33%, respectively) to adjust. To adjust the upper use level in products applied to the lips, one must consider the contribution from those individual products categories to the overall aggregated exposure (Figure 4). Since not all product categories will have an equal contribution to aggregate dermal exposure it was necessary to approximate what their individual contributions were to total exposure. This allowed the upper use level concentration to be reduced by way of deriving weighting factors.
The approximate percentage contribution that each individual product category has on the aggregate exposure to an application site was calculated from their individual 95th percentile product category exposure. The 95th percentile exposure for each individual product category was divided by the sum of all 95th percentile product category exposures to an application site (see Table 11). It should be noted that the total sum of the individual product category exposures do not equate to the CELAgg but are used to approximate their relative contribution to the CELAgg. Importantly, sensitivity analyses have shown that individual product exposures are a good approximation of their contribution to aggregate exposure.
Finally, based on the relative contribution each product category has on the aggregate exposure, it was possible to calculate a weighting factor. In the example below, Category F products had a contribution of 84.7%, therefore the upper use level of Citral in this product category was reduced by 84.7%, by multiplying the upper use level of Citral in Category F products by a weighting factor of 0.15 (1 – 0.847). For Category E product types, whose contribution to CELAgg was 0.6%, the upper use level was reduced by 0.6% using an upper use level weighting factor of 0.99 (1 – 0.006).

When the weighted upper use levels were input into the Creme RIFM aggregate exposure model, it was found that the AEL/CEL_{Agg} was 1.9, which suggests that the upper use levels were reduced by more than was necessary. The AEL/CEL_{Agg} overshoot by a factor of almost 2 was caused by the high weighting factors, especially from Product Category F. In this instance it was necessary to incorporate a multiplication factor to appropriately reduce the individual product category weighting factors to produce an AEL/CEL_{Agg} that is closer to 1, thus:

Upper Use Level Weighting Factor = 1 – (Contribution × Multiplication Factor)

Using this method, it was found after several iterations that a multiplication factor of 0.776 (Table 12) provided appropriate upper use level weighting factors, which led to an AEL/CEL_{Agg} of 1.13 (Table 13). It should be noted that the adjustment factors produced an AEL/CEL_{Agg} that were in all cases above 1 (not equal to 1). The reason for this was that the probabilistic nature of the Creme RIFM model allows for standard error in the aggregate exposure estimates, thus an AEL/CEL_{Agg} that is slightly above 1 compensates for this.
Interestingly, the re-calculation of the AEL/CEL_{Agg} for all application sites with the adjusted upper use levels showed that the AEL/ CEL_{Agg} for the intra-oral region was found to be above 1 due to the reduced upper use level of Citral in product Category F. The AEL/CEL_{Agg} increased for all application sites, however the AEL/CEL_{Agg} for palms and axillae were still below 1 and thus required the upper use levels to be reduced in the products that contributed to their aggregate exposure

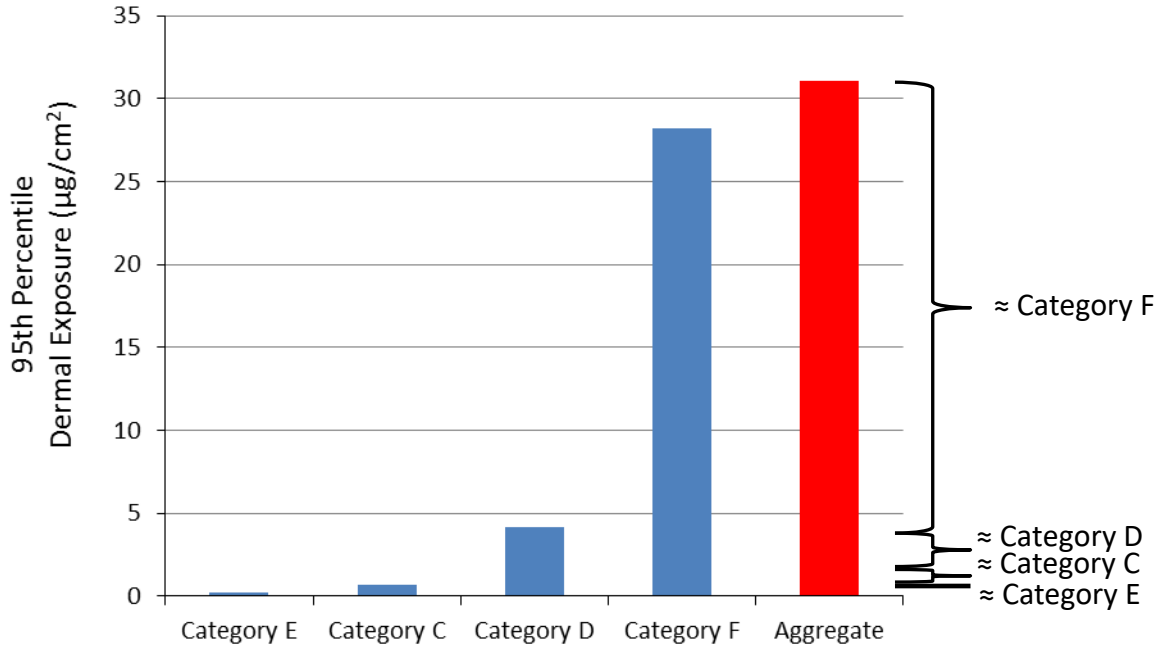


Figure 3: Illustration of Approximate Contribution of Product Categories to the CEL_{Agg} to the Lips (Figure 4 in the Final report on QRA2)

Interestingly, the re-calculation of the AEL/CEL_{Agg} for all application sites with the adjusted upper use levels showed that the AEL/ CEL_{Agg} for the intra-oral region was found to be above 1 due to the reduced upper use level of Citral in product Category F. The AEL/CEL_{Agg} increased for all application sites, however the AEL/CEL_{Agg} for palms and axillae were still below 1 and thus required the upper use levels to be reduced in the products that contributed to their aggregate exposure.

Table 6: Calculation of Approximate Percentage Relative Contribution to Aggregate Exposure from Individual Product Categories Applied to the Lips to Produce Upper Use Level Weighting Factors (Table 11 in the Final report on QRA2)

Product Category	**95th Percentile Dermal Exposure (µg/cm²)**	**Relative Contribution**	**Percentage Relative Contribution**	**Upper Use Level Weighting Factor**
F	28.2	28.2/33.3 = 0.847	84.7	1 - 0.847 = 0.15
D	4.2	4.2/33.3 = 0.126	12.6	1 - 0.126 = 0.87
C	0.7	0.7/33.3 = 0.021	2.1	1 - 0.021 = 0.98
E	0.2	0.2/33.3 = 0.006	0.6	1 - 0.006 = 0.99
Total	33.3	**1**	**100%**	**-**

Table 7: Calculation of Upper Use Level Weighting Factors Based on Product Category Contribution and Adjustment Factor (Table 12 in the Final report on QRA2)

Product Category	**Relative Contribution**	**Multiplication Factor**	**Upper Use Level Weighting Factor**

F	0.847	0.776	1 – (0.847*0.776) =0.34
D	0.126	0.776	1 – (0.126*0.776) = 0.9
C	0.021	0.776	1 – (0.021*0.776) = 0.98
E	0.006	0.776	1 – (0.006*0.776) = 1

The upper use level weighting factors for the product categories used on each application site were calculated and were used to adjust the Citral upper use levels (Table 19), collectively called the 'QRA2 aggregate adjustment factor'. It should be noted that in the present study, the adjustment factors were calculated based on the fragrance Citral. However, the same NESIL value was used to calculate the upper use levels for each product type and the AEL for each application site. Therefore, the adjustment factors calculated for Citral could also be used to adjust the upper use levels for all fragrances whose NESIL is known.

Table 8: Upper Use Levels for Citral in Product Types and Product Categories with Adjustment Factors (Table 19 in the Final report on QRA2)

Product Type	Citral NESIL =1400 µg/cm^2				
	QRA2 Upper use limit	**Product Categorization**	**QRA2 category**	**QRA2 aggregate adjustment factor**	**QRA2 aggregate exposure adjusted upper use level**
Deodorants & Antiperspirants of all types including fragranced body sprays	0.05%	A	0.05%	0.63	0.03%
Hydroalcoholic Products (eau de toilette, parfum etc.)	0.63%	B	0.63%	0.95	0.60%
Body Creams, lotions	0.78%	C	0.50%	0.47	0.23%
Hand cream	0.54%				
Facial Cream (Moisturizing)/Facial Balm	0.50%				
Eye Products (Includes: eye shadow, mascara, eyeliner, eye make-up)	0.65%	D	0.12%	0.88	0.11%
Women's Make up (Foundation)	1.52%				
Make-up remover	1.56%				
Lip Products	0.12%				
Hair styling aids (mousse, gels, leave in conditioners)	3.50%				
Hair sprays	2.12%				
Shampoo	2.75%	E	2.33%	0.57	1.33%
Body wash/shower gels	31.10%				
Conditioner (rinse-off)	7%				
Bar soap	2.33%				
Liquid soap	7.00%				
Face washes, gels, scrubs	3.11%				
Bath gels, foams, mousses	46.67%				
Toothpaste	1.10%	F	1.10%	0.34	0.37%
Mouthwash	1.40%				

SCCS comment

It should be clarified at the beginning of this section of the Final Report that the aim is to reduce the single-product upper use level to different types of "upper use levels" that are valid for the different products considered for aggregate exposure. According to the

definition of the upper use level earlier on, it cannot be reduced. Therefore, a change in nomenclature should be considered.
From a methodology perspective, it is not clear why a contribution of a product category of 84% should result in a reduction of 84% of the upper use level. This needs to be better explained as it implies that all product categories contribute equally to the joint upper use level. If this assumption is correct, then this would need a different algorithm.
Also, in the Final Report, the case study results for palms have been given and it is not clear why this is needed.

SCCS comment to the additional information (Appendix 1)
The additional information provided in Appendix 1 on the adjustment of upper concentration levels needs to be integrated in the Final Report on QRA2 (2016). In this example the old product categories are used in Table 6 and 7 and not the new product categories (1-12), this needs to be corrected in a new version as well.
The nomenclature used in the comment of the Applicant is not clear enough to allow an appraisal of the explanation. For example, it is not clear why the AEL is denoted as AEL_{agg}. For the AEL nothing is aggregated, but it is a site-specific AEL. Without the context of a new draft, the SCCS is unable to assess if the comments raised in the preliminary Opinion are addressed.

3.3.4 Planned work to further refine the QRA

From the Final report on QRA2, 2016:
QRA2 is an advance in the development of a robust and transparent risk assessment methodology for skin sensitisers compared to the original QRA procedure but further work is necessary in several key areas. The immediate priorities are:

- To complete the ongoing work to incorporate consideration of pro- and particularly pre-haptens into QRA2.
- Agreeing a protocol and conducting a critical evaluation of the effectiveness of QRA2 in minimising consumer sensitisation.

The IDEA project is also committed to identify and characterise non-animal tests as basis for conducting risk assessment in line with the requirements of the Cosmetics Directive

SCCS comment
The SCCS appreciates the Applicant's intention to carry out further work to refine the QRA methodology, and hopes that the comments provided in the preceding sections will be helpful in conducting a thorough revision of the whole methodology.

3.4 Discussion

The final report on QRA2 shows a lot of progress since the original launch of the QRA concept. The SCCS appreciates the Applicant's intention to carry out more work to further refine the methodology, and hopes that the comments provided in the preceding sections are helpful for a thorough revision and improved description of the methodology.
The SCCS analysis of the final QRA2 report has shown that, like other newly emerging methodologies, QRA2 still has a number of aspects that need either simplifying, adjusting, or explaining on scientific grounds. The Supplement provided by the Applicant during the public consultation period of the preliminary Opinion provided more, but also new information on the methodology. Not all SCCS comments were sufficiently addressed, meaning that concerns remain. The following is a brief summary of the SCCS analysis of the Final report on QRA2 as well as the supplementary information that has been provided in more detail under each individual section of the report:

- The methodology is primarily aimed at preventing induction of sensitisation, e.g. primary prevention. However, the methodology cannot be considered fully

comprehensive for primary prevention of fragrance contact allergy since fragrances not covered by IFRA standards will not be taken into account.

- In regard to QRA2, it is questionable why a confirmatory HRIPT is needed. The SCCS repeats its longstanding ethical concerns over human skin sensitisation tests, such as the HRIPT, because of the concerns that exposure levels used in the test may themselves cause sensitisation in healthy volunteers. Also at a confidence interval of 95% in a sample of 100, even the negative test results may still include up to 3 individuals (i.e. 3%) who could have been sensitised. Therefore, the SCCS strongly discourages new human testing for HRIPT data, but good quality existing data can be used.
- In the view of the SCCS, although the LLNA EC3 values correlate relatively well with human NOAELs, different publications show that for some chemicals this correlation is not so perfect. For a significant proportion of the fragrances, the LLNA EC3 value may be higher than the human threshold. Therefore, the SCCS suggests the inclusion of an interspecies SAF in the absence of human data that could overrule the LLNA EC3. The SCCS view is that, based on the available literature for fragrances, an interspecies SAF of 3 might be appropriate. A case-by-case evaluation on the necessity of an interspecies SAF is recommended. Data on toxicokinetic differences between test animals and humans as well as physical-chemical properties of the test chemical may be useful in such an evaluation.
- The WoE procedure to derive the NESIL still needs to be provided. Considerations taken to derive a WoE NESIL are unclear and need better explanation. Basic rules of interpretation of the test methods listed to derive this NESIL need to be provided.

- Scientific reasoning for the suggested SAF of 10 for inter-individual variability is still not provided. Although a default factor of 10 for inter-individual variability is suggested by many regulatory and conceptual frameworks, these also suggest the use of substance-specific information to adjust or substitute the default. The available information indicates that inter-individual variability of induction of skin sensitisation might be well above 10. Therefore it is not self-evident why a value of 10 is suggested in QRA2. A suggested way forward could be to introduce a separate chapter on uncertainties covering this issue (and also other issues on QRA2).

- The rationale behind the product SAF is still unclear and needs better explanation. Furthermore for the majority of cosmetic products a SAF of 1 is proposed. According to the Applicant, this SAF is not relevant to the majority of cosmetic products, for which the SAF is 1. The exceptions are products in a solid matrix for which a SAF of 0.3 is proposed. Therefore, the product SAF seems redundant and in order to simplify the methodology, SCCS suggests excluding it from the assessment.
- The impact of frequency of exposure on the induction of skin sensitisation is not fully clear. The SCCS agrees with the Applicant that the frequency SAF of 3 can be applied to cover for uncertainties on the impact of frequent exposure to fragrances on skin sensitisation hazard. The lack of scientific knowledge on this aspect makes it impossible to exactly determine the factor needed to cover for this. A way forward is to assign a SAF of 3 if the uncertainties related to this are clearly described in a separate chapter on uncertainties.
- The skin condition SAF seems to relate to the body sites where the products are applied to the skin. Hence, the SAF is not referring to certain skin conditions, but to the site of product application, which partly entails certain skin conditions, like increased hydration in intertriginous areas such as ano-genital or axillary. Therefore, SCCS recommends renaming this SAF to avoid any confusion. The SCCS agrees that a SAF of 3 for all products and a SAF of 10 for the most sensitive sites can be applied to account for these uncertainties.

- In QRA2 18 body sites are used. The rationale provided is based on practicability. In addition to these practical considerations a scientific rationale is needed, so that these body sites can be used in an assessment. Such a scientific rationale may include differentiation according to skin properties, occlusion levels, product types etc. Also, it should be considered when and how exposure of different body sites may need to be aggregated (e.g. in the case for palms and back of the hands, which presumably drain to the same lymph nodes).

- In regard to aggregate exposure assessment, clarifications are needed on "full distribution of exposure data" in terms of what parameters are meant, and what use patterns are considered, and whether this only refers to single-product exposures.
- An integral part of the QRA2 approach is the use of the CREME RIFM aggregate exposure model, which is a probabilistic model based on distributions for selected input data. The description of the methodology in relation to the use of the model for exposure assessment under QRA2 needs to be clarified, especially in regard to which parameters are treated probabilistically and which as point values. It is also important to point out that the CREME RIFM Aggregate Exposure Model uses the common methodology of probabilistic exposure assessment, which is currently not used or recommended for cosmetics by the SCCS and needs further evaluation by the SCCS.
- From a methodology point of view, the SCCS notes that upper use levels for single products can only be derived for fragrances on the basis of IFRA information. It would be useful to clarify whether the risk assessment part of the QRA2 can be used also for other cosmetic ingredients.
- In QRA2 aggregation is done for each of the 18 body sites separately, while aggregation over multiple body sites is not foreseen. There is uncertainty that this may impact the risk assessment, since exposure to certain adjacent body sites may lead to trafficking of Langerhans cells to the same lymph node, leading to a more effective priming of the immune system. This uncertainty needs to be better addressed in a separate uncertainty chapter.

SCCS recommends that the supplementary information provided is used to update and revise the Final report on QRA2 from 2016. A chapter describing the uncertainties related to the methodology would be important in this update as well.

4. CONCLUSION

- *In light of the methodology provided, does the SCCS consider QRA2 adequate to establish a concentration at which induction of sensitisation by a fragrance ingredient unlikely to occur?*

  The "QRA2 final report" together with the supplementary information received shows that a lot of progress has been achieved since the initial publication of the QRA. However, it is not yet possible to use the QRA2 to establish a concentration at which induction of sensitisation of fragrance is unlikely to occur. Several aspects of the methodology are not clear and the scientific rationale behind the methodology needs to be better described. These aspects have been highlighted in this Opinion.

- *Does the SCCS have any further scientific comments with regard to the use of QRA2 methodology to determine, in particular regarding applicability, development and improvements?*

  A number of additional considerations and refinements have been incorporated to the proposed methodology. However, explanation of certain methodological approaches and assumptions, as well as a description of uncertainties is lacking, the provision of which would enhance understanding of the methodology. These aspects have been highlighted in the SCCS comments under each section with the aim to provide pointers for improvement. If shaped up properly, this could be a useful methodology not only for risk assessment of fragrance allergens, but potentially also for other cosmetic ingredients.

5. MINORITY OPINION

/

6. REFERENCES

1. IDEA Project. Final report on the QRA2. Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients. September 30, 2016.

From the submission:

1. Api AM, Basketter DA, Cadby PA, Cano MF, Ellis G, Gerberick GF, et al. Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. Regul Toxicol Pharmacol. 2008;52(1):3-23.
2. Api, A.M., Basketter, D., Lalko, J., 2015. Correlation between experimental human and murine skin sensitisation induction thresholds. Cutaneous and ocular toxicology, 1-5.
3. Basketter, D.A., Lea, L.J., Dickens A., Briggs, D., Pate, I., Dearman, R.J. and Kimber, I. (1999). A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. Journal of Applied Toxicology, 19(4), 261-266.
4. Basketter D.A., Clapp, C., Jefferies, D., Safford, B., Ryan, C.A., Gerberick, F., Dearman, R.J. and Kimber I. (2005a). Predictive identification of human skin sensitisation thresholds. Contact Derm, 53:260-267
5. Basketter D, Safford B. Skin sensitization quantitative risk assessment: A review of underlying assumptions. Regul Toxicol Pharmacol. 2016;74:105-16
6. Cano, M.-F. (2006). Personal communication on exposure data for hydroalcoholic products for shaved skin.
7. Cassimos C, Kanakoudi-Tsakalidis F, Spyroglou K, Ladianos M, Tzaphi R. (1980) Skin sensitisation to 2, 4 dinitrochlorobenzene (DNCB) in the first months of life. J Clin Lab Immunol. 3: 111-113.
8. Comiskey D, Api AM, Barratt C, Daly EJ, Ellis G, McNamara C, et al. Novel database for exposure to fragrance ingredients in cosmetics and personal care products. Regul Toxicol Pharmacol. 2015;72(3):660-72
9. ECETOC (2003). Contact Sensitisation: Classification According to Potency. Technical Report No. 87. European Centre for Ecotoxicology and Toxicology of Chemicals. ISSN-0773-8072-87. Brussels, April 2003.
10. ECETOC (2009). Framework for the Integration of Human and Animal Data in Chemical Risk Assessment. Technical Report No. 104. Ecetoc, Brussels, January 2009. ISSN-0773-8072-104.
11. ECHA, (2012), Guidance on information requirements and chemical safety assessment, Chapter R.8: Characterisation of dose [concentration]-response for human health, Appendix 8-10, http://echa.europa.eu/documents/10162/13632/information_requirements_r8_en.pdf
12. Epstein WL (1961) Contact type delayed hypersensitivity in infants and children: induction of rhus sensitivity. Pediatrics 27: 51-53
13. Farage, M.A., Bjerke, D.L., Mahony, C., Blackburn, K.L. and Gerberick, G.F., (2003). Quantitative risk assessment for the induction of allergic contact dermatitis: Uncertainty factors for mucosal exposures. Contact Dermatitis, 49(3), 140-147.
14. Felter, S.P., Robinson, M.K., Basketter. D.A. and Gerberick, G.F. (2002). A review of the scientific basis for uncertainty factors for use in quantitative risk assessment for the induction of allergic contact dermatitis. Contact Dermatitis, 47(5), 257-266.
15. Felter, S.P., Robinson, M.K., Basketter. D.A. and Gerberick, G.F. (2002). A review of the scientific basis for uncertainty factors for use in quantitative risk assessment for the induction of allergic contact dermatitis. Contact Dermatitis, 47(5), 257-266.
16. Felter, S.P., Ryan, C.A., Basketter, D.A. and Gerberick, G.F. (2003). Application of the risk assessment paradigm to the induction of allergic contact dermatitis. Regulatory Toxicology and Pharmacology, 37, 1-10

17. Final Report on the QRA2 Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients of September 2016
18. Fischer, L.A., Menné, T., Avnstorp, C., Kasting, G.B., Johansen, J.D. (2009) Hydroxyisohexyl 3-cyclohexene carboxaldehyde allergy: relationship between patch test and repeated open application test thresholds. Br J Dermatol.: 161: 560-567.
19. Fischer, L.A., Menné, T., Voelund, A., Johansen, J.D..(2011) Can exposure limitations for well-known contact allergens be simplified? An analysis of dose- response patch test data. Contact Dermatitis.: 64: 337-342.
20. Friedmann, P.S. (2007). The relationships between exposure dose and response in induction and elicitation of contact hypersensitivity in humans. Br J Dermatol, 157(6):1093-102.
21. Gerberick, G.F., Robinson, M.K. Ryan, C.A., Dearman, R.J., Kimber, I., Basketter, D.A., Wright Z. and Marks, J.G. (2001a). Contact allergenic potency: Correlation of human and local lymph node assay data. American Journal of Contact Dermatitis, 12(3), 156-161.
22. Griem, P., Goebel, C. and Scheffler, H. (2003). Proposal for a risk assessment methodology for skin sensitisation potency data. Regulatory Toxicology and Pharmacology, 38, 269-290.
23. Hall B, Tozer S, Safford B, Coroama M, Steiling W, Leneveu-Duchemin MC, McNamara C, Gibney M. (2007) European consumer exposure to cosmetic products, a framework for conducting population exposure assessments. *Food and Chemical Toxicology, 45(11), 2097-2108.*
24. Hall, B., Steiling, W., Safford, B., Coroama, M., Tozer, S., Firmani, C., McNamara, C., and Gibney, M. (2011) European consumer exposure to cosmetic products, a framework for conducting population exposure assessments Part 2. *Food and Chemical Toxicology*, 49:408-422.
25. Holsapple, M.P., Paustenbach, D.J., Charnley, G., West, L.J., Luster, M.I., Dietert, R.R., Burns-Naas, L.A., (2004). Symposium summary: children's health risk--what's so special about the developing immune system? Toxicology and applied pharmacology 199, 61-70
26. Hostynek, J.J. and Maibach, H.I. (2004). Thresholds of elicitation depend on induction conditions. Could low level exposure induce sub-clinical allergic states that are only elicited under the severe conditions of clincial diagnosis? Food and Chemical Toxicology, 42, 1859-1865
27. Kimber I, Basketter D, Butler M, Gamer A, Garrigue J-L, Gerberick GF, Newsome c, Steiling W, Vohr H-W. (2003). Classification of contact allergens according to potency: proposals. Food Chem. Toxicol. 41, 1799-1809.
28. Kimber I, Dearman RJ, Basketter DA, Ryan CA, Gerberick GF, McNamee PM, Lalko J, Api AM. (2008). Dose metrics in the acquisition of skin sensitisation: thresholds and importance of dose per unit area. *Regul Toxicol Pharmacol* 2008
29. McNamee, P.M., Api, A.M., Basketter, D.A., Frank Gerberick, G., Gilpin, D.A., Hall, B.M., Jowsey, I. and Robinson, M.K. (2008): A review of critical factors in the conduct and interpretation of the human repeat insult patch test. *Regul Toxicol Pharmacol*, 52:24-34
30. Loretz L.J., Api A.M., Barraj L.M., Burdick J., Dressler W.E., Gettings S.D., Hsu H.H., Pan Y.H.L., Re T.A., Renskers K.J., Rothenstein A., Scrafford C.G. and Sewall C. (2005). Exposure data for cosmetic products: Lipstick, body lotion, and face cream. Food and Chemical Toxicology, 43, 279-291
31. Loretz, L., Api, A.M., Barraj, L., Burdick. J., Davis de, A., Dressler, W., Gilberti, E., Jarrett, G., Mann, S., Laurie Pan, Y.H. et al. (2006) Exposure data for personal care products: hairspray, spray perfume, liquid foundation, shampoo, body wash, and solid antiperspirant. Food Chem Toxicol 2006, 44:2008- 2018. [52297]
32. Loretz, L.J., Api, A.M., Babcock, L., Barraj, L.M. and Burdick, J. (2008). Exposure data for cosmetic products: Facial cleanser, hair conditioner, and eye shadow. Food and Chemical Toxicology, 46(5), 1516-1524 2008.
33. Militello, G., Jacob, S.E., Crawford, G.H., (2006). Allergic contact dermatitis in children. Current opinion in pediatrics 18, 385-390.

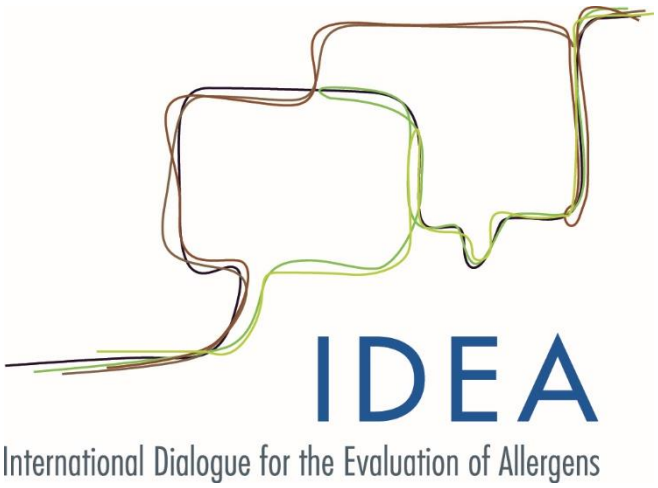
34. OECD (2002). OECD Guideline for the testing of chemicals No. 429. Skin Sensitisation: Local Lymph Node Assay. Available at http://dx.doi.org/10.1787/9789264071100-en.
35. Politano, V.T., Api, A.M. (2008) The Research Institute for Fragrance Materials' human repeated insult patch test protocol. Regul Toxicol Pharmacol, 52:35- 38
36. Robinson, M.K. (1999). Population differences skin structure and physiology and the susceptibility to irritant and allergic contact dermatitis: Implications for skin safety testing and risk assessment. Contact Dermatitis, 41(2), 65-79
37. Safford B, Api AM, Barratt C, Comiskey D, Daly EJ, Ellis G, et al. Use of an aggregate exposure model to estimate consumer exposure to fragrance ingredients in personal care and cosmetic products. Regul Toxicol Pharmacol. 2015;72(3):673-82
38. SCCP. Opinion on Dermal Sensitisation Quantitative Risk Assessment (Citral, Farnesol and Phenylacetaldehyde). Scientific Committee for on Consumer Protection, adopted 24 June 2008. (SCCP/1153/08)
39. SCCS (Scientific Committee on Consumer Safety), opinion on fragrance allergens in cosmetic products, adopted 26-27 June 2012. (SCCS/1459/11)
40. Ter Burg, W., Wijnhoven, S.W.P., Schuur, A.G., (2010), Observations on the methodology for quantitative risk assessment of dermal allergens, RIVM Rapport 320015003, Bilthoven, The Netherlands http://www.rivm.nl/Documenten_en_publicaties/Wetenschappelijk/Rapporten/2010/november/Observations_on_the_methodology_for_quantitative_risk_assessment_of_dermal_allergens.
41. Tozer, S.A., O'Keeffe, L., Cowan-Ellsberry, C.E. and Rich K. (2004). Use of probabilistic analysis in the refinement of exposure data for hydroalcoholic perfume products. Toxicology, 202(1-2), 123-124
42. WHO, (2004), IPCS Risk assessment terminology. Part 1: IPCS/OECD key generic terms used in chemical hazard/risk assessment. Part 2: IPCS glossary of key exposure assessment terminology http://www.inchem.org/documents/harmproj/harmproj/harmproj1.pdf

Other references:

1. Basketter DA, Jefferies D, Safford BJ, Gilmour NJ, Jowsey IR, McFadden J, Chansinghakul W, Duangdeeden I, Kullavanijaya P. The impact of exposure variables on the induction of skin sensitization.Contact Dermatitis 2006 55(3):178-85.
2. Basketter DA, Natsch A, Ellis G, Api, AM, Irizar A, Safford B, Ryan C, Kern P. Interspecies assessment factors and skin sensitization risk assessment. Regulatory Toxicology and Pharmacology 2018 97 186-188.
3. Bil W.B, Schuur A.G., Ezendam J. and Bokkers B.G.H. Probabilistic derivation of the interspecies assessment factor for skin sensitization. Regul Toxicol Pharmacol 2017 88, 35-44.
4. European Chemicals Agency (ECHA). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7A: Endpoint specific guidance. Version 5.0. December 2016. https://echa.europa.eu/information_requirements_r7a_en.pdf
5. Gefeller O, Pfahlberg AB and Uter W. What can be learnt from nothing? – A statistical perspective. Contact Dermatitis 2013;69:350-354.
6. ICCVAM (2011). ICCVAM test method evaluation report: usefulness and limitations of the LLNA for potency categorization of chemicals causing allergic contact dermatitis in humans. National Institute of Environmental Health Sciences, Research Triangle Parc, NC.
7. Kligman AM (1966). The identification of contact allergens by human assay. II. Factors influencing the induction and measurement of allergic contact dermatitis. Journal of Investigative Dermatology, 47,375–392.
8. Roberts DW & Api AM (2018). Chemical Applicability domain of the local lymph node assay (LLNA) for skin sensitisation potency. Part 4. Quantitative correlation of LLNA potency with human potency.

9. SCCNFP, Final opinion concerning the predictive testing of potentially cutaneous sensitising cosmetic ingredients or mixtures of ingredients, adopted by the SCCNFP during the 11th plenary session of 17 February 2000.
10. SCCS (Scientific Committee on Consumer Safety), Memorandum on use of Human Data in risk assessment of skin sensitisation, SCCS/1567/15, 15 December 2015.
11. WHO/IPCS (2005) Chemical-specific adjustment factors for interspecies differences and human variability : guidance document for use of data in dose/concentration-response assessment. IPCS harmonization project document ; no. 2.

Appendix 1: IDEA Comments on the SCCS Preliminary Opinion on Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2) Submission I



IDEA
Comments on the SCCS Preliminary Opinion on Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2) Submission I

January 23, 2018

INTRODUCTION AND GENERAL COMMENTS

On behalf of the IDEA (International Dialogue for the Evaluation of Allergens) project, the IDEA Supervisory Group and its management team, we would like to express our gratitude for the detailed and insightful preliminary Opinion on Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2) Submission I, adopted by the SCCS at the plenary meeting on 24-25 October 2017.

We are pleased to note the Scientific Committee's identification of significant progress in the methodology since the initial publication of the QRA in 2008, and its support for the numerous additional considerations and refinements that have been incorporated in the proposed revised methodology.

We also recognise and welcome the additional suggestions for improvements made by the Scientific Committee. Like the SCCS, we firmly believe that these improvements to the QRA2, which is work in progress, will lead to the development of a useful methodology not only for risk assessment of fragrance allergens, but potentially also for other cosmetic ingredients.

This response seeks to acknowledge areas where the Opinion has identified a need for clarity, provide additional information where it is felt to be beneficial and, most importantly, incorporate our learnings from the Opinion into the evolution of the QRA process. We will also consider the need to review the previously submitted Final Report on QRA2 dossier in terms of its overall clarity in any future publications or communications. The Opinion provides insights that will also be important input to an enhanced surveillance study currently being planned which is designed to support monitoring of the operation of the QRA. This study will be presented at the next IDEA Annual Review Meeting. Also, QRA2 will be applied to IFRA Standards with the next Amendment, which is scheduled to be published in 2018.

In that overall context, it is our hope that the members of the SCCS will join the next IDEA Annual Review Meeting (likely planned for June of this year), where they can observe how their comments and insights have been incorporated into the process of continual improvement of the QRA and hear details of the planed study.

This document is also intended to supplement the original submission of September 2016 "IDEA Project. Final Report on the QRA2", which, it is fully acknowledged, is work in progress. It is trusted that the current format will facilitate the SCCS review of the submission.

Each section of these comments intends to provide further clarification on points raised by the SCCS under section 3.4 (Discussion) of the Opinion and additional questions raised in the body of the Opinion. Each response has been grouped under a general heading (or theme) related to QRA2 to allow several points from the opinion to be addressed together.

These comments have been prepared under the IDEA Project by the following people under the supervision of and taking into account the comments of the IDEA Supervisory Group: Anne Marie Api (RIFM); David Basketter (Consultant); Peter Cadby (Consultant); Graham Ellis (Givaudan); Bob Safford (Consultant); Boris Müller (Symrise); Carsten Goebel (Coty); Cécile Gonzalez (IDEA Management Team); Peter Griem (Symrise); Joseph Huggard (Consultant); Amaia Irizar (IFRA); Petra Kern (P&G); Charles Laroche (IDEA Management Team); John O'Brien (CREME Global); Cindy Ryan (P&G) and Matthias Vey (IDEA Management Team). The section 12 "Attribution of skin sites and aggregation" was prepared with support from Prof. Peter Friedmann (University of Southampton) and Prof. Ian Kimber (University of Manchester).

1. ORDER OF APPLYING PRODUCT SPECIFIC SAFs

SCCS COMMENTS

Page 10: 8-9: SCCS suggests modifying Figure 1 so that SAFs are divided from the start into exposure-related (product-specific), location-related (site-specific) and general (hazard-specific).
Page 10: 19 Since product-specific SAFs are defined, this would pose a problem in an aggregate exposure assessment if different SAFs would apply for the different products. It would seem easier to apply the product-specific SAFs to the single exposures that are later aggregated.
Page 21: 31-35: the SCCS does not agree with a determination of the total SAF at the level of the hazard assessment. Since product-specific SAFs are defined, this poses a problem in an aggregate exposure assessment in the case that different SAFs apply for the different products. It would be more appropriate to apply the product-specific SAFs to the single exposures that are later aggregated.
Page 32-33: 48-54 and 1-2: The flow of elements in Figures 1 and 2, and the explanatory text should be made clear and consistent. For example, Figure 1 may be adjusted so that from the start SAFs are divided into exposure-related (product-specific), location-related (site-specific) and general (hazard-specific) SAFs, and then be fed into the process at the respective places. Instead of calculating "overall SAFs", the SAFs should be used as weighting factors at each step in the process of aggregation. Also, it would be more useful to apply the product-specific SAFs to the single exposures that are later aggregated. In SCCS view, the "Upper use level" may be changed to "Upper concentration levels" because "use" is often related to use patterns of consumers.
Page 34: 7-12: The SCCS does not agree with a determination of the total SAF, which is done at the level of the hazard assessment. Since product-specific SAFs are defined, this poses a problem in an aggregate exposure assessment in the case that different SAFs apply for the different products. In the view of the SCCS, it would be easier to apply the product-specific SAFs to the single exposures that are later aggregated.

RESPONSE

We agree with the SCCS that the term "upper concentration level" would be much more appropriate to replace the current terminology of "upper use level" for the reasons stated in the Opinion.
Regarding the application of product specific SAFs and where in the QRA2 methodology these are applied, we would like to provide additional explanation and clarity on this point below, including a revision of Figures 1 and 2, which represent two steps within the process. We trust that this will clarify that the approach does, in fact, follow the sequence outlined in the SCCS Opinion.
Figure 1 below summarises the process followed under QRA2.
The first part of the process follows traditional toxicology risk assessment practice to determine upper concentration limits (safe use levels) of a specific ingredient at the individual product level (UCLPRODUCT). An acceptable exposure level (AEL) is determined by division of the NESIL by the appropriate Sensitisation Assessment Factors (SAFs). The SAFs include assessment of inter-individual variability, product considerations, frequency/duration of use and skin condition (related to the skin site(s) where a product would be used). Hence, the product consideration SAF is accounted for in the determination of the Acceptable Exposure Level (AEL) of the individual ingredient for a specific product type. This is then compared with the Consumer Exposure Level (CEL) for the fragrance ingredient within the product and an Upper Concentration Level for that ingredient in an individual product (UCLPRODUCT) can be derived.
This value is then used in the next step of the process which includes consideration of aggregated exposure to a designated skin site from an ingredient potentially present across all product types at its Upper Concentration Level in product (UCLPRODUCT). This allows for derivation of an Aggregate Upper Concentration Level (UCLagg) accounting for potential combined exposure from multiple products on the same skin site. Here it is necessary to calculate the aggregate exposure to consumer products for each of the 18 designated body

sites using the probabilistic model and the previously determined UCLPRODUCT. This provides an aggregate exposure to the fragrance ingredient per body site (CELagg). This can then be compared to the Acceptable Exposure Level (AELagg) for each of the body sites which are derived by dividing the NESIL over the SAF for that body site.
Figure 2 in the original submission showed the product SAF as being applied to the individual product CEL rather than as part of the SAFs applied to the NESIL. The Opinion has drawn our attention to this being incorrect and it has therefore been removed from this version.
The final step is to adjust the UCLPRODUCT so that the AELagg : CELagg ratio is ≥1 for each body site (i.e., aggregated exposure does not go above the calculated acceptable exposure per body site). This is done by applying weighting factors to calculate the revised UCLPRODUCT as described. This provides the final upper concentration level for the specific ingredient for each product type for the purpose of risk management of induction of sensitization.
The process for the calculation of the weighted adjustment factors is summarized in Figures 2 and 3 below to provide, what we trust, is greater clarity. In addition, a detailed working example has been provided for products applied to the lip area later in this document (Section 16).

Figure 1: Summary of QRA2 process

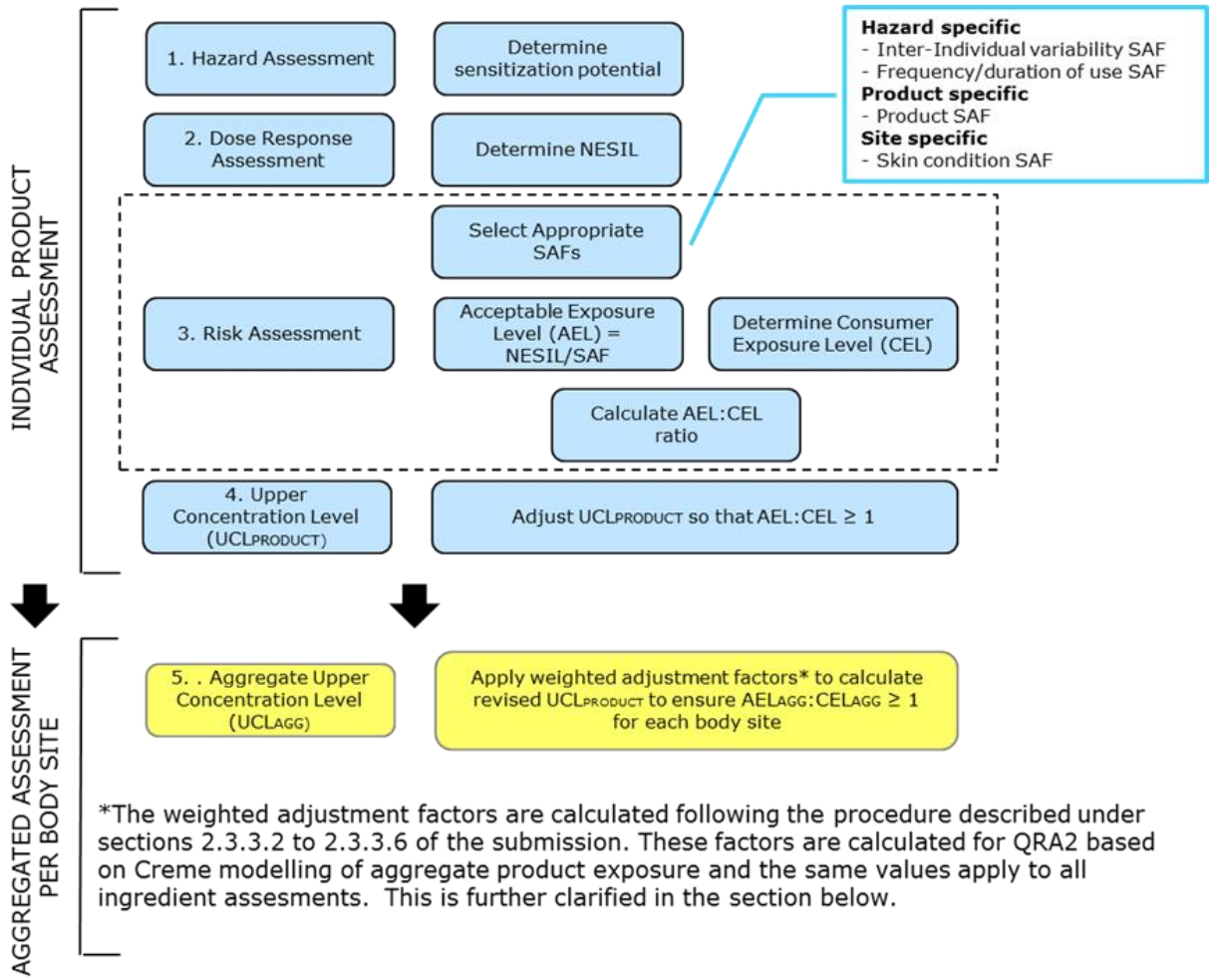
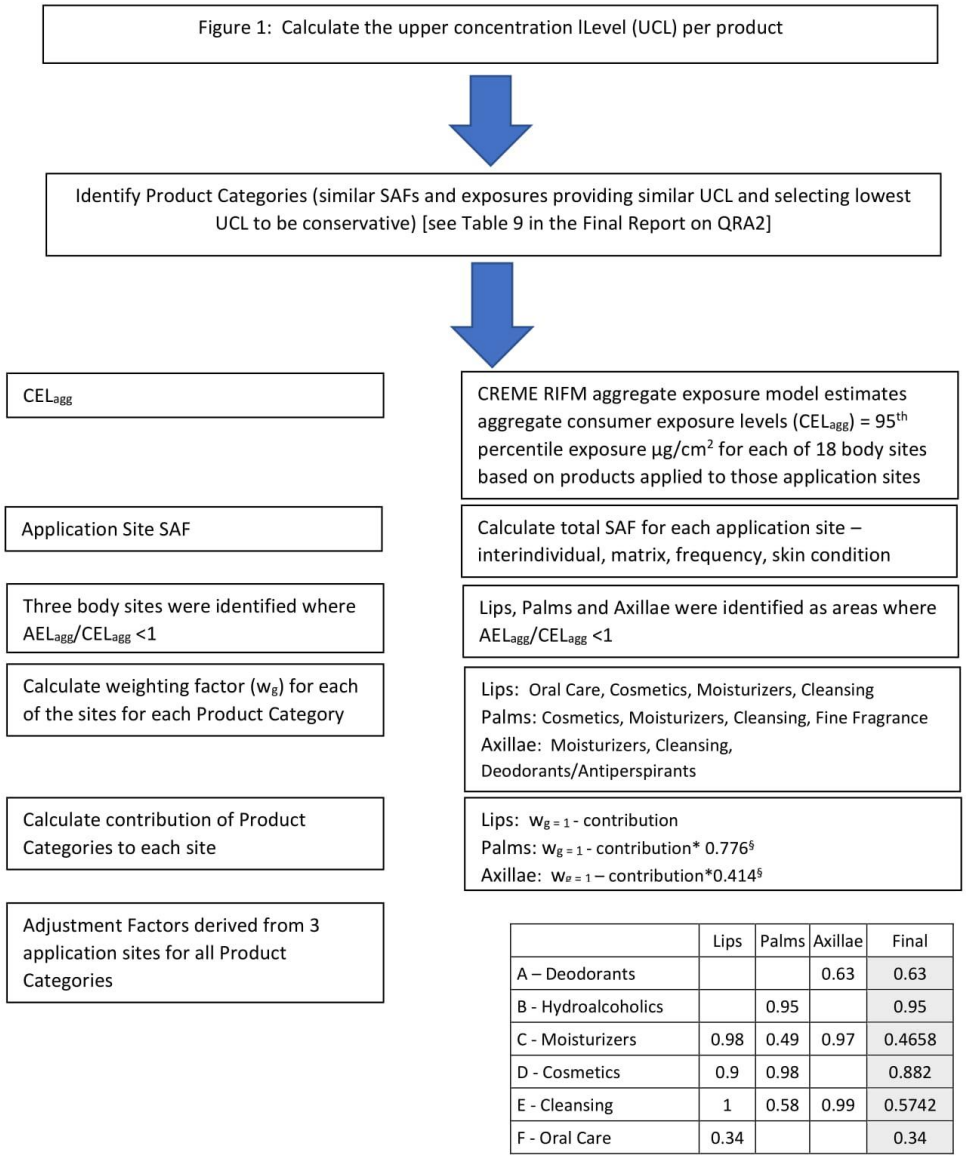


Figure 2: Determination of the Adjustment Factors

(this figure replaces Figure 2 in the dossier).



	Lips	Palms	Axillae	Final
A – Deodorants			0.63	0.63
B - Hydroalcoholics		0.95		0.95
C - Moisturizers	0.98	0.49	0.97	0.4658
D - Cosmetics	0.9	0.98		0.882
E - Cleansing	1	0.58	0.99	0.5742
F - Oral Care	0.34			0.34

§A suitable multiplication factor obtained through an iterative process with an initial value of 1 in order to obtain an $AEL_{agg}/CEL_{agg} \leq 1$

As noted in the original submission dossier, product categorization was introduced in the implementation of the IFRA Standards based on QRA1. The categorization provided in Table 9 in the final report on QRA2 was for illustrative purposes and was subject to change, to

allow addition of other products and categories to reflect the full range of products covered by IFRA Standards. Since the time that the final report on QRA2 was provided in September 30, 2016, work has been done to introduce product categorization in QRA2 for implementation into the IFRA standards. Table 1 provides all the categories for implementation of QRA2 into the IFRA Standards. Figure 3 details how the adjustment factors were derived after all the products were introduced.

Table 1: Product Categories for introduction of QRAs into the IFRA Standards.

Product Type
Category 1 - Products applied to the lips
Category 2 - Products applied to the axillae
Category 3 - Products applied to the face/body using finger tips
Category 4 - Products related to fine fragrances
Category 5A - Body Lotion Products applied to the face and body using the hands (palms), primarily leave-on
Category 5B - Face Moisturizer Products applied to the face and body using the hands (palms), primarily leave-on
Category 5C - Hand Cream Products applied to the face and body using the hands (palms), primarily leave-on
Category 5D - Baby Cream, Oil, Talc
Category 6 - Products with oral and lip exposure
Category 7 - Products applied to the hair with some hand contact
Category 8 - Products with significant ano-genital exposure
Category 9 - Products with body and hand exposure, primarily rinse off
Category 10A - Household Care
Category 10B - Aerosol Air Freshener
Category 11 - Products with intended skin contact but minimal transfer of fragrance to skin from inert substrate
Category 12 - Other Products not intended for direct skin contact, minimal or insignificant transfer to skin

§A suitable multiplication factor obtained through an iterative process with an initial value of 1 in order to obtain an $AEL_{agg}/CEL_{agg} \leq 1$

Figure 3: Determination of Adjustment Factors for final IFRA Categories

Figure 1: Calculate the upper concentration level (UCL) per product

↓

Developed Product Categories (similar SAFs and exposures providing similar UCL and selecting lowest UCL to be conservative) [see Table above]

↓

CEL_{agg}	CREME RIFM aggregate exposure model estimates aggregate consumer exposure levels (CEL_{agg}) = 95th percentile exposure µg/cm^2 for each of 18 body sites based on products applied to those application sites
Application Site SAF	Calculate total SAF for each application site – interindividual, matrix, frequency, skin condition
Five body sites were identified where AEL_{agg}/CEL_{agg} <1	Lips, Palms, Intra-oral, Back of Hand and Axillae were identified as areas where AEL_{agg}/CEL_{agg} <1
Calculate contribution of Product Categories to each site	Lips: Categories 6, 1, 9, 5 Palms: Categories 5, 9, 7, 10, 4 Axillae: Categories 2, 5, 9 Intra-Oral: Back of Hand:
Calculate weighting Factor (w_g) for each of the sites for each Product Category	Palms: $w_{g=1}$ - contribution*2[§] Lips: $w_{g=1}$ – contribution*0.776[§] Axillae: $w_{g=1}$ – contribution Note: The AEL_{agg}/CEL_{agg} ratio was >1 for the Back of Hand site after the adjustment to the palms was made. The AEL_{agg}/CEL_{agg} ratio was >1 for the Intra-Oral site after the adjustment to the lips was made.

Adjustment Factors derived from 3 application sites for all Product Categories

Product Category	Palms	Lips	Axillae	Final
Category 1		0.9077		0.9077
Category 2			0.6256	0.6256
Category 3				1.0000
Category 4	0.9492			0.9492
Category 5	0.3510	0.9985	0.9378	0.3286
Category 6		0.3207		0.3207
Category 7	0.5776			0.5776
Category 9	0.5210	0.9971	0.9667	0.5022
Category 10	0.6010			0.6010

*A suitable multiplication factor obtained through an iterative process with an initial value of 1 in order to obtain an $AEL_{agg}/CEL_{agg} \leq 1$

2. LIMITATIONS IN THE COVERAGE OF QRA2

SCCS COMMENTS
Page 7: 38-42: Importantly, the use of fragrances not covered by IFRA standards (10% by volume) is not included in the aggregate exposure assessment and subsequent risk characterisation. Hence, the methodology cannot be considered adequate to ensure full primary prevention of fragrance contact allergy.
Page 32: 45-46: However, the methodology cannot be considered fully comprehensive for primary prevention of fragrance contact allergy since fragrances not covered by IFRA standards will not be taken into account.

RESPONSE
This comment is correct because, as pointed out in the Final Report on QRA2, "IFRA membership accommodates about 90% (by volume) of the fragrances compounds produced globally and used in consumer products". This 90% is a global figure relating to the coverage of IFRA membership, meaning the coverage of IFRA risk management. This is further supported by IFRA Standards, which are included in several regulatory frameworks. As indicated, there are some uses of fragrance ingredients that are outside the area that IFRA's risk management standards tool (IFRA Standards) typically address (e.g. aromatherapy, drugs and topical treatments, massage and spa therapies, foods, etc.). Application to these sectors should be a target for the future through the appropriate industry bodies and it would be hoped that the QRA methodology might be beneficially applied here. Also, the CREME RIFM aggregate exposure model has considerable flexiblibility with regard to input data and new product types and thus exposures can be added as this information becomes available.
The method as presented has been designed to cover the vast majority of consumer products that, in general, the consumer is most frequently exposed to and in this regard, aligns well with those products considered in the SCCS Notes of Guidance (SCCS/1564/15).

3. COMPARISON OF HUMAN NOAELs WITH MOUSE LLNA EC3

SCCS COMMENTS
Page 11: 24-38: When comparing classifications of fragrances based on EC3 values and human data according to GHS potency categories, the fragrances included in Api et al. (2015) and ICCVAM (2011) show that 6 (11%) and 7 (24%) fragrances, respectively, are classified as other sensitizers based on EC3 values (>2%) but as strong sensitizers based on human data (≤500 µg/cm2). Likewise, when comparing classification of EC3 values based on the CLP classification, the fragrances included in Api et al. and ICCVAM show that 7 (13%) and 8 (28%) of the fragrances, respectively, are classified as weak-moderate sensitizers based on EC3 values (≥1-≤100%) but as strong sensitizers based on human data (≤500 µg/cm2). Thus, for a significant proportion of the fragrances, the LLNA EC3 value is higher than the human threshold.
A recent publication has shown that applying an interspecies factor is required to ensure that the sensitisation threshold determined in the LLNA does not underestimate the human threshold (Bil et al., 2017). The SCCS would therefore suggest the inclusion of an interspecies SAF in the absence of human data that could overrule the LLNA EC3.
Page 33: 15-20: In the view of the SCCS, although the LLNA EC3 values correlate relatively well with human NOAELs derived from HRIPT and HMT studies, different publications show that for some chemicals this correlation is not so perfect and, for a significant proportion of the fragrances, the LLNA EC3 value may be higher than the human threshold. In this regard, application of interspecies SAF can be suggested in the absence of human data to overrule the LLNA EC3.

RESPONSE
The publication by Bil et al. (2017) has been reviewed and comments are in preparation for submission as a short communication to Regulatory Toxicology and Pharmacology for review

in the open literature. A preliminary draft on which that submission will be based is provided in Annex 1.
The Final Report on QRA2, 2016 states that "the true maximum HRIPT NOAEL is generally somewhere well above the dose levels chosen for this confirmatory test and for ethical reasons, is not determined in the QRA process". For fragrance ingredients, as described in the Final Report on QRA2, NESILs are historically based on LLNA and human data. When, as described in our explanation of how dose levels are chosen for running HRIPT (see below), negative HRIPT data are used to set a NESIL, the true maximum HRIPT NOAEL may be somewhere well above the dose levels chosen for this test. This occurs when the HRIPT is run principally for the purpose of reassurance. Hence, for ethical reasons determination of a true NOAEL, is not the goal of the HRIPT. Further details and examples are provided in our response to the question of how dose levels are chosen for this test.
Annex 1 contains a number of arguments that we believe would make the application of an interspecies SAF unnecessary. In cases where there are wide divergences between the EC3 and the dose level used in a negative HRIPT, this should be taken into account in the weighing of evidence. As indicated below in our response to the question of how dose levels are chosen for the HRIPT, the dose levels of more recently run HRIPT studies have been determined by prior run LLNA studies.
A good example of the need for a weight of the evidence approach is seen with the following fragrance ingredient: Hexyl salicylate (CAS number 6259-76-3). It shows perhaps the greatest discrepancy between the LLNA EC3 (45 µg/cm2) and the negative HRIPT (35433 µg/cm2). In this case, consideration was given to the discrepancy between the results of the LLNA, three negative guinea pig studies, the absence of structure alerts of protein reactivity (OECD Toolbox V3.1; Toxtree 2.5.0) and, as reported in the SCCS Opinion of Fragrance Allergens (SCCS/1459/11), the absence of reactions in any reported clinical patch tests including one in 218 patients with known contact allergy to fragrance ingredients (tested at 5% in petrolatum). This last item of data relates to an ingredient that is a "top 100" fragrance substance with significant consumer exposure.
In the absence of human data, the risk assessor needs to ensure that a necessary adjustment of the NESIL is made so that it reflects a threshold value (in µg/cm2) in humans. It is recognized that data from the LLNA is subject to inter-assay variability, as is the case with most in vivo assays. As such, this variability should be considered in situations where data from a single LLNA is available to be considered in the derivation of a NESIL. In the future, the determination of a NESIL will include data from alternative assays because LLNAs or other in vivo data will not be available. The sources of the data must be taken into consideration when deriving the NESIL itself, rather than making adjustment for the data source via use of an additional SAF. It is believed to be essential that a consistent definition of the NESIL is maintained as the quantitative threshold exposure level that is considered not to induce skin sensitisation in humans.

4. ETHICS OF RUNNING HRIPT

SCCS COMMENTS
Page 13: 10-13: Although the current experience with the risk on sensitisation due to this 'confirmatory' HRIPT indicates that it is low (0.3%), it is not absent. More importantly, one fragrance material was responsible for 12 of the 24 cases, showing that for this specific material the selected HRIPT concentration was not safe.
Page 33:13-14: the SCCS strongly discourages new human testing for HRIPT data, but good quality existing data can be used.

RESPONSE
The SCCS concerns related to the use of a Human Repeated Insult Patch Test (HRIPT), even if only for 'confirmatory' purposes, are shared.
The safety assessment process uses a tiered approach to evaluation of the safety of fragrance ingredients, including use of historical data. Existing data from HRIPTs have provided information for the quantitative risk assessment and risk management of over 100 allergenic fragrance ingredients. A number of these are listed in Table 1 of the original QRA

publication (Api et al., 2008) and others are currently available in the IFRA Standards (http://www.ifraorg.org/en-us/standards-library/snew#.WlxqK6hKtPY).
Notwithstanding this contribution, we fully understand the concerns of the SCCS regarding testing on human subjects as provided in the SCCS Memorandum on use of Human Data in risk assessment of skin sensitisation (SCCS/1567/15). The replacement of this test at the same time as the replacement of animal tests presents an acknowleded challenge. A number of activities are underway to develop and evaluate the application of non-animal data in the hazard characterisation and risk assessment of contact allergens. IDEA is partnering with the JRC and others to facilitate the dialogue around the risk assessment of fragrance allergens without the generation of new animal data. A recent draft report (Zuang et al., 2017) provides some reason for optimism in regard to the use of alternatives to the LLNA in sensitisation risk assessment, even though validation of the different methodologies may be still some time away.
The Opinion cites a total of 24 positive reactions (12 from one substance) in HRIPTs performed for RIFM. These are clearly unacceptable and many occurred before reliable indications of potency could be gauged from tests on animals or in vitro systems. Testing trans-2-Hexenal gave 12 positive reactions in an HRIPT run at 236 µg/cm2. This dose level had been chosen based on two LLNA studies which gave a vehicle weighted EC3 value of 4.0% (1012 µg/cm2) (i.e. more than five times higher than the dose eventually chosen for the HRIPT). The volatility of the test material may explain the reason for the discrepancy between the results of the open LLNA and the occluded HRIPT. Subsequent HRIPT studies with similarly volatile test substances have taken account of this factor.
This particular material notwithstanding, the data on a diverse set of 57 fragrance ingredients detailed in a paper by Api et al. (2015), confirm that, in general, there is a consistent relationship between the LLNA EC3 and HRIPTs. Murine and human thresholds for the 57 fragrance chemicals in the Api et al. (2015) publication span approximately four orders of magnitude variation in potency. Good correlation (with half an order of magnitude) was seen with three-quarters of the dataset. In an extended series of studies, the accuracy of this murine induction threshold as the predictor of the absence of a sensitizing effect was verified by conduct of an HRIPT. The paper provides details on the LLNA data and all the human data that are available on these 57 fragrance ingredients (see Table 1 in Api et al., 2015).
RIFM is currently working on a manuscript that will provide information on additional HRIPTs conducted from 2005 to the present time according to the methodology published in Politano and Api (2008). This future publication will provide detail on the study protocols, results and observations, which have not been covered by Api et al. (2015).

5. CHOICE OF DOSE LEVELS IN HRIPT

SCCS COMMENT
Page 13: 14-18: Both the final report and the accompanying literature do not describe the procedure of arriving at the concentration that has been considered safe to test in the confirmatory HRIPT. According to Api (2008), this concentration is based on hazard assessment data from animal tests, but guidance on how the animal data are extrapolated to the concentration used in the induction phase of the HRIPT is not provided. This should be clarified.

RESPONSE
We acknowledge the need to provide detail of the procedure. The concentration that has been considered safe to test in the confirmatory HRIPT should be determined as follows. All available data (animal test data, particularly the LLNA, in silico, in chimico and in vitro predictions and available test results from close structural analogues or likely metabolites) are firstly collected and evaluated. The maximum use levels of the substance in these categories is obtained from industry surveys. This plays an important role as studies may have been conducted at levels below the expected NESIL if this is sufficient to support

reported use levels. How these different elements interplay is exemplified below using two real cases.
EXAMPLE 1: A sterically-hindered α,β-unsaturated ketone is predicted to be a moderately strong skin sensitizer. Structural analogues also indicate the same. A guinea pig maximization test carried out at high doses (topical induction 100%, intradermal 50 and 100%, challenge at 5%) gave mild erythematous reactions in 9/10 animals. Two LLNAs were run (one with and the other without antioxidant). Both gave the same EC3 of 550 μg/cm^2.
Maximum upper concentration levels in cosmetics were obtained by industry-wide surveys including fragrance companies and consumer product companies. The highest concentration of this substance in a finished product was found to be 0.02% (in fine fragrances). In this category, an upper concentration level of 0.02% would, through the calculation of QRA, result from a NESIL of 50 μg/cm2. From this reverse QRA calculation, it was concluded that a skin area dose of 50 μg/cm² needed to be tested in the HRIPT, in order to support the current marketed use of the fragrance ingredient.
In this example, we see that the dose level chosen for the HRIPT is based only on levels of real use. The NESIL thus obtained is well below the NOAEL that is indicated by the LLNA but it is large enough to confirm the safety of current use without putting the subjects of the HRIPT at unnecessary risk.
EXAMPLE 2: An allylic ester is predicted to be a moderate to weak skin sensitizer. Structural analogues also indicate the same. An LLNA gave an EC3 of 3.1% (775 μg/cm2).
Maximum use levels in cosmetics were obtained from surveys and the highest concentrations in some categories were such that they would, according to QRA2 exposure estimation, deliver up to 1750 μg/cm2. However, in view of the LLNA EC3 an HRIPT was run at only 0.6% (709 μg/cm2) which, in the absence of any reactions, confirmed a NESIL of 700 μg/cm2.
In this example, we see that although it might be feasible to test the safety of current use levels by running an HRIPT at 1750 μg/cm2, it was decided instead to run the HRIPT at a dose level that is 2½ time lower (709 μg/cm2) as indicated by the LLNA.
Historically, of course, a large number of HRIPTs were run prior to the LLNA EC3 values becoming available. Indeed, many were carried out prior to the LLNA being developed or validated. This has led to some cases where there are large discrepancies between the LLNA EC3 and the dose level in the HRIPT. In many of these cases, the HRIPT was carried out at dose levels that were considered on the basis of other tests, read-across or other reasons involving expert judgement, as likely to give negative results in the HRIPT.
In the future, the goal is to be able to determine NESILs on the basis of alternative methods within an AOP/IATA framework. Fortunately, there is now a quite extensive collection of HRIPT results from which to read-across to untested fragrance ingredients.

6. LIMITED CONFIDENCE LEVEL FROM RUNNING HRIPT ON ONLY 100 SUBJECTS

SCCS COMMENTS
Page 13: 22-27: when a confirmatory HRIPT in 100 subjects yields the (expected) result of no sensitised individual (i.e. 0%), there is, based on statistical considerations, a confidence interval to be considered. This implies that for a sample of 100, a confidence interval of 95% would include up to 3 individuals (i.e. 3%) who still could have been sensitised (Gefeller, 2013).
Page 33: 11-13: Also at a confidence interval of 95%, even the negative test results may still include up to 3 individuals (i.e. 3%) who could have been sensitised.

RESPONSE
It is well accepted that for toxicological studies where it is necessary to limit the number of test subjects, statistical considerations challenge the interpretation of the implications of the confidence interval, in terms of either false positive or false negatives, for the consumer. In the case of tests such as the HRIPT, a statistical analysis (from Hendersen and Riley, 1945) is addressed by Politano and Api (2008).

The HRIPT has a number of conditions that make for a conservative outcome by enhancing the likelihood of induction compared to exposure from cosmetics or household products. These are detailed elsewhere (Politano and Api, 2008), but include the use of complete occlusion for 24-hour periods, whereas use of cosmetic products occurs at most only under very partial occlusion (e.g. in axillæ or under clothing). Basketter and Safford (2016) have cited published data that show that occlusion appears to maximize induction by an approximate 3-fold factor. This was shown with substances of low volatility (DCNB, NDMA and PPD) but the effect of occlusion in maintaining dose may be significantly higher for volatile fragrance ingredients.
As mentioned elsewhere, the standard vehicle used in the HRIPTs for fragrance ingredients (Politano and Api, 2008) is a mixture of diethyl phthalate and ethanol. This mixture of solvents was shown to be optimal for the induction of sensitisation in the LLNA (Lalko et al., 2004). On the other hand, water, a common ingredient of cosmetic products, has been shown to be a sub-optimal vehicle in predictive tests in animals.
These maximizing conditions will reduce, to a degree that it is recognised cannot be readily quantified, the maximum number of individuals that could be sensitized but go unobserved at the 95% confidence level.

7. CLARIFICATION NEEDED ON WEIGHT OF EVIDENCE GUIDELINES

SCCS COMMENTS
Page 14: 51 – Page 15: 29: Requests for clarification and guidance in exceptional circumstances.
Page 33: 21-24: Clarifications are needed in relation to the criteria used in the Guidelines for the weight of evidence approach for determining NESIL. These have been detailed in the SCCS comments under the relevant section. For example, all the Guidelines suggest that data from a well-run HRIPT always takes precedence over other available (animal or human) data. It is however not clear how this will work where HRIPT NOAEL exceeds LLNA EC3.

RESPONSE
We would like to provide additional clarification on the guiding principles followed in the derivation of the NESIL. These principles, presented below, reflect the current best practice in this area of toxicology following a balanced approach to determination of NESIL values on a case by case basis. This should be seen as an update of and replacement for the original guidance on NESIL derivation issued a decade ago (Api et al., 2008).
From experimental investigations and on the grounds of basic immunological considerations, the quantity of chemical per unit area of the skin (e.g. µg/cm²), is considered as the most appropriate dose metric for skin sensitization. This approach is considered to reflect the current state of science and is in line with the overwhelming majority of available data in both, humans and experimental animals. Therefore, quantitative descriptors of skin sensitization, such as NOAELs, LOAELs and LLNA EC3 values, will be expressed as dose per unit area. (While the term EC3 is used here for simplicity, it should be noted that the corresponding threshold values from non-radioactive variants of the LLNA may also be used).

The NESIL is the skin area dose of a sensitizer per day that is considered not to induce skin sensitization in naive humans (i.e., those not yet sensitized against the investigated chemical). The weight of evidence evaluation must decide if, e.g., the NESIL should be set at the level of a human NOAEL or of a LLNA EC3 or if it should be derived from in vitro testing data, QSAR or be based on a read-across approach, or on any combination of these.

In the derivation of a NESIL all available data should be taken into consideration in a weight of evidence approach. In addition to experimental data from studies performed with the substance under evaluation, these data may also include results from in vitro tests (for which approaches providing quantitative sensitization strength information are only currently being developed), in silico quantitative structure-activity predictions, other

alternative tests, as well as read-across to data on similar ingredients (Figure 4). Adjustments of thresholds derived from any source other than human to derive a NESIL should be made in the process of derivation of the NESIL, i.e. on the hazard side in the QRA approach and not by application of a generic interspecies adjustment factor to derive the AEL.

Since fragrance ingredients are used in many consumer products intended to be used on the skin, human skin test data, especially HRIPTs, are considered relevant data in characterizing sensitizing potency and area concentration-sensitization frequency relationships. It is important to evaluate the robustness and reliability of these studies. A well run HRIPT employs a published methodology, is well documented and involves approximately 100 or more subjects. Results from other historical human sensitization induction tests (e.g. human maximization tests (HMT)) will be integrated into the evaluation, but will usually carry a lower weight. Historical HRIPTs employed test concentration ranges that sometimes resulted in cases of skin sensitization in the high dose groups and, thus, allowed derivation of NOAEL and LOAEL values. HRIPTs with fragrance ingredients are done for confirmatory purposes, i.e. they will employ a single area dose level that was carefully selected to support the safety assessment of concentrations used in marketed products, but at the same time the dose tested will be considered not sensitizing based on the (limited) information available. As a consequence, most HRIPTs will support the dose area tested as a non-sensitizing level while the real NOAEL level may be considerable higher. Consequently, NESILs based on HRIPTs will have an additional, though not quantifiable, built-in safety margin.

Data from diagnostic patch test studies cannot be used directly for the determination of NESILs. These studies can be useful to help determine the need for additional data. For example, they may indicate where current exposures to a fragrance ingredient may be a source of clinically relevant allergic contact dermatitis. The absence of confirmed positive reactions following testing in dermatology clinics may provide support current exposure levels for that fragrance ingredient.

LLNAs are valuable data in the derivation of a NESIL since they use topical application of the test material and deliver dose-response information. From a large body of data comparing human NOAEL and LOAEL dose area values for sensitization induction with LLNA EC3 values, it has been concluded that in most instances the EC3 value can be used as a surrogate value for the human NOAEL provided that any known differences have been taken into account.

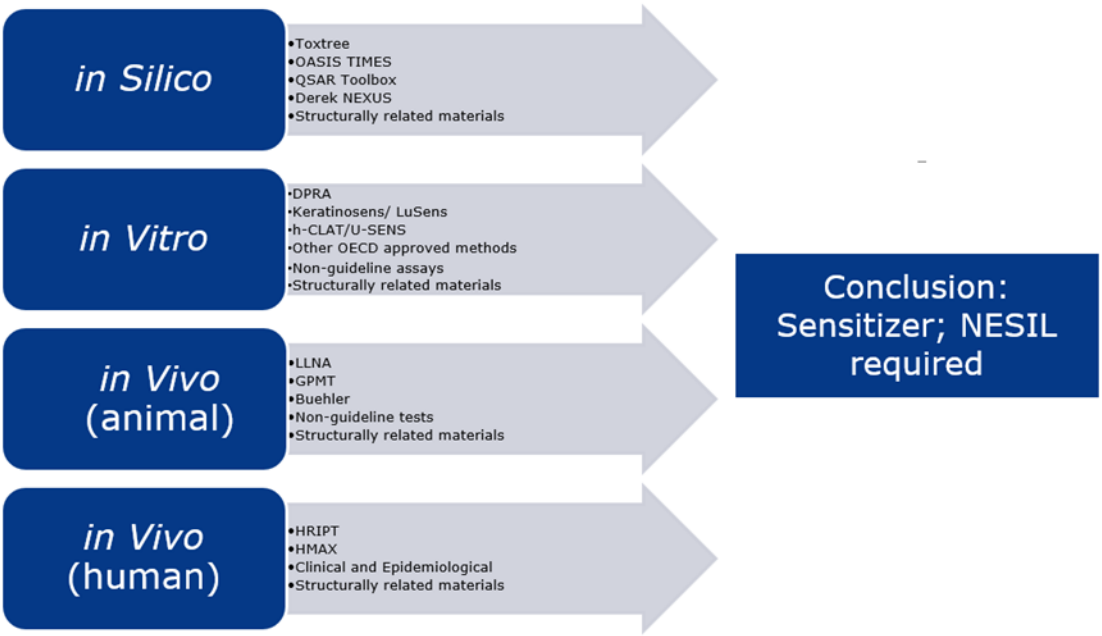
Adjuvant tests in animals (Guinea Pig Maximization Test (GPMT), Freund's Complete Adjuvant Test (FCAT), Mouse Ear Swelling Test (MEST), etc.) and non-adjuvant tests in guinea pigs (e.g. Buehler Test, Open Epicutaneous Test (OET), Closed Epicutaneous Test (CET)) will usually carry a lower weight in the derivation of NESILs because derivation of a skin area dose from such studies is generally unreliable. They may be used to contribute information to determine the potency classification (See ECETOC, 2003 for further reference). Also, these data may help elucidating species differences, e.g., with regard to metabolism in the skin, skin penetration, and vehicle effects.

When the available data basis is considered insufficient for characterizing the sensitization dose-response and does not allow a high degree of confidence in NESIL derivation, a HRIPT may be considered for fragrance ingredients because these will be marketed in consumer products for skin applications and the HRIPT uses the same application route (dermal). Unquestionably, a cautious approach is required for selection of the dose level of fragrance ingredient in the conduct of any such confirmatory HRIPT in order to minimize the likelihood of sensitizing exposed study volunteers.

The current progress in establishing in vitro testing methods for skin sensitization, the increasing regulatory acceptance of such tests and the first attempts towards characterizing

the sensitization potency of chemicals from in vitro testing means that in the future NESILs will need to be derived from in vitro testing results without a need to resort to human or experimental animal data.

Figure 4 : Data considerations in establishing a NESIL



8. PRODUCT SAF AND INFLUENCE OF PENETRATION ON SENSITISATION

SCCS COMMENTS

Page 17:3- 5: The products SAF is applied to take into consideration the influence of the product matrix on the degree of sensitisation. It is, however, not fully clear to the SCCS, why and when this SAF is applied.
Page 17: 6-11: According to the final report, and the Basketter and Safford (2016) paper, the SAF is used to account for the ingredients that enhance skin penetration. However, the final report mentions that "enhancement of penetration does not necessarily enhance sensitisation". This seems to be contradicting the need for the products SAF. In addition, since the CEL is based on the external dose and not on the epidermal dose, the enhancement of penetration does not seem to be relevant.
Page 17:11- 12: In the final report on QRA2, this SAF is 1 for all product categories, so it is unclear for which products a SAF of 0.3 or 3 is applied.
Page 17:13- 15: The SCCS has not been able to evaluate the relevance of this SAF and needs further clarification on the scientific substantiation in this regard. Examples of product categories for which a SAF of 0.3 or 3 is applied would also be useful.
Page 33: 34-38: The product SAF is applied to take into consideration the influence of the product matrix on the degree of sensitisation but it not clear why and when this is applied. It is described to account for the ingredients that enhance skin penetration but the report also mentions that "enhancement of penetration does not necessarily enhance sensitisation". Also, because the CEL is based on the external dose and not on the epidermal dose, the enhancement of penetration does not seem to be relevant.

RESPONSE

What we trust is an improved description of the rationale for the SAFs is provided in Table 2. The product SAF is applied to the NESIL in order to account for changes that the product matrix may have on the sensitising potential of a chemical. It has been shown that the matrix in which a sensitizer is applied may influence the degree to which sensitisation occurs, both in the LLNA and in human studies (HMT, HRIPT) (see Basketter and Safford,

2016). In setting a SAF, it must be remembered that the factor should represent the likely difference in degree of sensitisation between the consumer product and the derived NESIL as described above. The NESIL is benchmarked against the HRIPT, in which the sensitizer is applied in a vehicle considered to be optimal for the induction of sensitisation (DEP/ethanol or petrolatum). Thus, for products based on these or similar solvents, a factor of 1 is considered appropriate to account for the matrix. Although it is possible that sensitisation potential will be reduced in aqueous based products based on observations in the LLNA, it is proposed to maintain a factor of 1 for these products since they are rarely purely aqueous, and will contain other ingredients such as surfactants which help the product wet the skin. Since all the products considered in this dossier are applied directly to skin, and are solvent or aqueous based, a product SAF of 1 applies to all. In the case of some product types, particularly those in a solid matrix, such as dry facial tissues, napkins and incense sticks, a product SAF of 0.3 is considered appropriate since the sensitisation risk is considered to be lower.
For skin sensitisation, the induction and elicitation processes are a consequence of a covalent reaction between the skin sensitizer and self-protein and which occurs locally in the epidermis/dermis. As was noted by Roberts and Williams in their seminal paper in 1982, the essential factors which govern induction are the reactivity of a substance (the electrophile), its concentration and, importantly, its duration at the target site in skin. Increased penetration of the skin by a substance may impact concentration and duration, but the direction of that impact could be positive or negative; for example, increased penetration might result in a more rapid transit through the skin and more rapid removal into the bloodstream. The systemic load may therefore be higher, but the opportunity for reaction in skin may be reduced due to the shorter duration at the target site. This may be summarised as; penetration "measures what goes through the skin" whereas induction "depends on what remains (and reacts) within the skin".
Basketter and Safford (2016) provide examples. Whereas "barrier disruption was demonstrated to have a profound effect on the skin penetration of salicylic acid (>100-fold enhancement) (Benfeldt et al, 1999)" it has been demonstrated that "even removal of the stratum corneum down to the glistening layer generated only a very small increase in the frequency of the induction of contact allergy (Kligman, 1966b). The impact on the nickel allergy elicitation dose response occurred in the absence of an effect on the barrier (Agner et al, 2002)". The authors then provide evidence that the effect of vehicles in different tests on animals and humans provides no clear relationship. Those that might be considered to enhance penetration do not appear to systematically enhance induction. It must be concluded that frequently there is very limited information about the impact of matrix on the disposition in the skin of a sensitising substance compared to the vehicle system in which the sensitising activity was originally described. In addition, we know from a series of studies (summarised in Jowsey et al. 2008) that different vehicles often have only a modest, albeit unpredictable, impact on sensitising activity. Accordingly, it is being suggested that unless a risk assessor has a reason to apply a different safety assessment factor, a value of 1 should be chosen for the Product SAF. Conversely, a factor of 3 should be applied where there is evidence that the product matrix may enhance sensitisation over and above that from standard solvents. Although products containing penetration enhancers may be considered here, evidence does not suggest that enhancement of skin penetration increases sensitisation risk. However, it may be prudent to include a factor of 3 for products that included penetration enhancing ingredients and this would need to be assessed on a case by case basis. It was therefore considered important to maintain the Product SAF (also referred to as "Matrix" SAF) in the overall SAF assessment for application in cases where this may be useful.

9. FREQUENCY AND DURATION SAF

SCCS COMMENTS
Page 17: 48 – page 18:2: The impact of frequency of exposure on the induction of skin sensitisation is not fully clear due to a lack of relevant studies. Basketter et al. (2006) showed in a human study that frequent exposures to a low concentration of the strong skin

sensitizer PPD resulted in a higher rate of sensitisation than less frequent high dose exposure. This study supports the need for an SAF to account for frequent exposure. Ubiquitous use of fragrance materials in a broad range of consumer products makes frequent exposures likely, and the SCCS agrees that a SAF to account for this needs to be applied. No scientific rationale is provided supporting the factor of 3 that is assigned to account for this uncertainty. The SCCS is aware that other regulatory frameworks use higher factors for covering this type of duration extrapolation.
Page 33: 40-44: The impact of frequency of exposure on the induction of skin sensitisation is not fully clear. For example, frequent exposure to a low concentration of a strong skin sensitizer (PPD) has been reported to result in a higher rate of sensitisation than less frequent high dose exposure. Ubiquitous use of fragrance materials in a broad range of consumer products makes frequent exposures likely, and an SAF to account for frequent exposures needs to be applied.

RESPONSE
The QRA2 model is based on the assumption of continuous daily exposure to a given fragrance ingredient.
The rationale for factor of 3 is related to the SAF applied for use considerations specific for dermal sensitization extrapolating from the experimental situation to use considerations in real life scenarios including extrapolation for the duration and frequency of exposure. For fragrance ingredients in cosmetic products the lowest use consideration SAF is 3, with the exception of foot care products, due to limited skin permeability of foot skin (Feldmann and Maibach, 1967), and may be increased to a value of 10 depending on limitations related to the suitability of the experimental data. This leads to a minimal SAF of 30, when multiplied with the standard SAF for inter-individual variability. The values chosen are consistent with the approach used by EPA for general risk assessment to extrapolate from e.g. LOAEL to NOAEL and from less frequent to chronic exposures (Dourson et al., 1996; Dourson, 1996).
The rationale to apply at least 3 as a SAF for use frequency considerations is also supported by information available on a repeated exposure slightly below the EC3 level generated for formaldehyde releasers (De Jong et al., 2007). Increased LLNA exposure frequencies studied at dose levels calculated as EC2 (i.e. approximately 63% of the EC3) indicate that an LLNA design with 13 open applications over 57 days was found to be more effective than the standard LLNA (3 applications over 3 days) producing an average of a 2.65-fold increase in stimulation indices in 8 separate studies.
The SCCS comments also refer to a study with PPD in hair dyes with applied doses above the concentration that may be considered the AEL (Basketter et al., 2006). According to Ezendam et al. (2013) the NESIL for PPD is 17.5 µg/cm² corresponding to an AEL of 0.58 µg/cm² (considering a SAF for inter-individual variability of 10 and for inter-species differences of 3).
In this study by Basketter et al. (2006), subjects were instructed to use an oxidative hair dye containing approx. 1.5 % PPD once per month for 30 to 40 min or to use an oxidative hair dye containing approx. 0.5% PPD 5 min/day for the first 4 days and then approximately once per week thereafter (still with 5 min/exposure). Sensitization rates were 1.3% and 7.2%, respectively. Considering a dose per unit area of approximately 4 µg/cm² and 12 µg/cm² for 0.5 % and 1.5% PPD, respectively (based on Goebel et al., 2012 assuming linear extrapolation from 16 µg/cm² for 2% PPD) each single exposure is approx. 7-fold and 21-fold above the AEL of 0.58 µg/cm². Consequently, the reference does not include information as to whether daily application with exposure doses below the AEL would lead to an increased risk of induction.
A comprehensive description of the respective safety factors is available in the most recent publication by Basketter and Safford (Basketter and Safford, 2016) and is in line with the conclusion from several expert meetings under IDEA (http://ideaproject.info/uploads/Modules/Documents/idea-qra-workshop-%28may-13-15-2014%29---final-progress-report-4-9-14.pdf).

10. SKIN CONDITION SAF

SCCS COMMENTS

Page 20: 4 - Page 21: 18: The rationale provided for the skin condition SAF is not clear. Table 2 (Table 4 in the Final report on QRA2, 2016) is based on the product types used. In Table 2, the sentence "No additional contribution to skin condition is expected from product irritation" has been used for each product. This needs further explanation, because some matrices may contain ingredients that have irritant properties. In such cases, it would be more logical to cover these matrix effects in the product SAF, which according to the final report, only covers penetration enhancers. It would also be more logical to apply the skin condition SAF only for those skin sites that have more susceptibility to inflammation, irrespective of the products used.

The final report provides no scientific justification as to why certain body parts are considered more susceptible to inflammation, e.g. for axillae it is only mentioned because it is an intimate region without any further reasoning. Also, no explanation is provided why an SAF of 3 is assigned to a large proportion of the skin sites, which implies that many body sites are susceptible to inflammation. Furthermore, an SAF of 10 is used for product types that are used all over the body, without any justification why such products need such a high SAF. It is unclear why different SAFs have been used for bar and liquid soap, which both could be used all over the body. SCCS doesn't understand why face scrubs have the same SAF as face gels and face washes, whereas body scrub has been missed out altogether. A better description of the rationale for the SAFs needs to be provided in Table 17 2.

Page 33: 46-50: The rationale provided in Table 2 for the skin condition SAF is not clear and the statement "No additional contribution to skin condition is expected from product irritation" needs further explanation, because some matrices may contain ingredients with irritant properties. It would also be more logical to apply the skin condition SAF only for those skin sites that have more susceptibility to inflammation, irrespective of the products used.

Page 34:1-5: No scientific justification has been provided as to why certain body parts are considered more susceptible to inflammation than others. Also, no explanation is provided why an SAF of 3 is assigned to a large proportion of the skin sites, which implies that many body sites are susceptible to inflammation. Furthermore, an SAF of 10 is used for product types that are used all over the body, without any justification as to why such products need a high SAF.

RESPONSE

We agree that the sentence "No additional contribution to skin condition is expected from product irritation" which has been used for each product should have been explained more fully. It is detailed in the paper by Basketter and Safford (2016) where the contribution to the overall SAFs from product irritation is discussed under point 10.4. The authors state, "In considering the effect that the product matrix may have on skin sensitisation it is also important to consider the irritation potential of the product. It has already been mentioned under skin condition that inflammation of the skin may increase susceptibility to skin sensitisation. A SAF of 3 is already proposed for areas of skin that may be prone to irritation from product use, and therefore a further SAF is not considered necessary."

The same paper also provides more information on the rationale behind the choice of 1, 3, or 10 as a SAF for skin site; "One key parameter for lowering the threshold for the induction of skin sensitisation is that of compromised/inflamed skin. The HRIPT is conducted on uninflamed and intact skin, whilst consumers in the population at large may have compromised/inflamed skin due to a number of factors. There is little evidence to suggest that subjects with diseased skin (e.g. atopic eczema, psoriasis) are more sensitive to skin sensitizers. In addition, there is little evidence that compromising the skin barrier by physical or chemical means increases the potential for the induction of sensitisation. However, the generation of inflammation in skin, particularly from contact with irritant chemicals (such as SLS), may increase sensitivity to skin sensitizers. In determining a SAF to account for skin condition, skin sites that are more prone to inflammation, for example due to a chemical stimulus (irritant contact dermatitis), are considered to be more susceptible to the induction of skin sensitisation. The available data suggest that the

magnitude of this increased susceptibility is less than 10-fold, and it is therefore proposed to include a SAF value of 3 to account for this on susceptible sites such as hands. For most sensitive skin areas, such as axillae and ano-genital regions it may be appropriate to use a SAF of 10. For other skin sites that are less prone to irritation a SAF value of 1 is proposed." Apart from hair spray products all products have been assigned a SAF of 3 (to assume some inflammation may be present) or 10 (most sensitive sites). Hair sprays are intended for application to the hair with limited scalp exposure.
The justification for the choice of SAF values per product type is provided in table 2. Bar soaps and shower gels are considered for use all over the body, including most sensitive sites, and therefore have a SAF of 10 applied. Body scrubs would be considered similar to shower gels. Liquid soaps (i.e. hand liquid soaps) here refers to products that would be applied to face and hands and therefore a SAF of 3 is applied.

Table 2 : Rationale for Skin Condition SAF for Various Product Types

Product Type	**Rationale for Skin Condition SAF**
Deodorants and antiperspirants of all types including fragranced body sprays	The SAF is 10 as these products are applied to the axillae where the skin is easily irritated due to a combination of factors including the unique environment of the axillae (humid, oil rich sebum production and site for perspiration). There may also be acute transient irritation due to product application or mechanical irritation. Shaving may produce a transient irritation response.
Fine fragrance products (eau de toilette, parfum, etc.)	The SAF is 3* because the area is the neck, wrists, antecubital fossa. Irritation from shaving may produce an acute transient response. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Eye products (e.g. eye shadow, mascara, eyeliner, eye make-up)	The SAF is 3* because product is applied to the peri-ocular site and face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Body creams, lotions	The SAF is 10 because the area is the entire body which may include areas of inflamed skin, i.e.: intimate regions and axillae. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Hand cream	The SAF is 3* because the product is applied to the hands. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation
Facial cream (moisturizing) / Facial balm	The SAF is 3* because the product is applied to the face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Women's make-up (foundation)	The SAF is 3* because the product is applied to the face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Make-up remover	The SAF is 3* because the product may be applied to eyelids (peri-ocular region) and face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Lip products	A SAF of 3* is applied because the site is applied to the lips (highly vascular and there is exposure to mucous membranes and possible exposure to dry or chapped lips). Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation

Product Type	Rationale for Skin Condition SAF
Hair styling aids (mousse, gels, leave-on conditioners)	The SAF is 3* because when the product is applied to the hair there will also be exposure to the scalp and the palms of the hands. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Hair sprays	The SAF is 1 because it is applied to the scalp. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Shampoo	The SAF is 10 because the product is applied to the head (hair) and scalp with the hands and may also be used over the entire body as a shower gel. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Body wash / Shower gels	The SAF is 10 because product may be used all over the body including intimate regions and axillae. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Conditioner (rinse-off)	The SAF is 3*because the product is applied to the head (hair) and scalp with the hands. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Bar soap	The SAF is 10 because product may be used all over the body including the axillae and intimate regions. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Liquid soap	The SAF is 3* because product may be used on the hands and face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Face washes, gels, scrubs	The SAF of 3* has been attributed because the product is applied to the face. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Bath gels, foams, mousses	The SAF is 10 because product may be used all over the body including intimate body regions and the axillae. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.
Toothpaste and mouthwash	The SAF is a 3*. The sites are the lips and mouth. Products are not expected to be irritant and no additional contribution to skin condition is expected from product irritation.

3* approximates 3.16 or the half log of 10

11. SAF FOR INTER-INDIVIDUAL VARIABILITY

SCCS COMMENTS

Page 16: 7-21: The scientific reasoning for a SAF of 10 to account for inter-individual variability has not been made transparent in the QRA2 final report, but more discussion is provided in Basketter & Safford (2016). The available information from human studies (HMT and HRIPT) suggests that human variability of susceptibility to induction of skin sensitisation is likely to span 3-4 orders of magnitude. However, the underlying database seems to be rather weak. The authors further state that by the range of 3–4 orders of magnitude, the majority of variability would be covered (i.e. this would not represent the total variability). Although the SCCS acknowledges that there are limitations in the available human data, these data point to a quite high inter-individual variability. It is argued that some portion of variability might already be covered in the HRIPT (which, however, is performed in only 100 healthy volunteers and which cannot be quantified). Despite the indications that inter-individual variability might be very high, a considerably lower factor of 10 is suggested as SAF to account for this variability, using the reasoning that most regulatory frameworks use a default factor of 10 for this. The SCCS is of the opinion that the step from an indication of a variability spanning several orders of magnitude to the proposed SAF of 10 deserves a better substantiated justification.
Page 33: 26-32: the scientific reasoning for an SAF of 10 to account for inter-individual variability has not been given in the report. The information from human studies suggests that human variability for susceptibility to induction of skin sensitisation is likely to span 3-4 orders of magnitude, but the underlying database seems to be rather weak. Despite the limited availability, the data point to a quite high inter-individual variability. Therefore, the view of the SCCS, the suggested use of a considerably lower factor of 10 on the grounds that most regulatory frameworks use such a default factor requires a more substantiated justification.

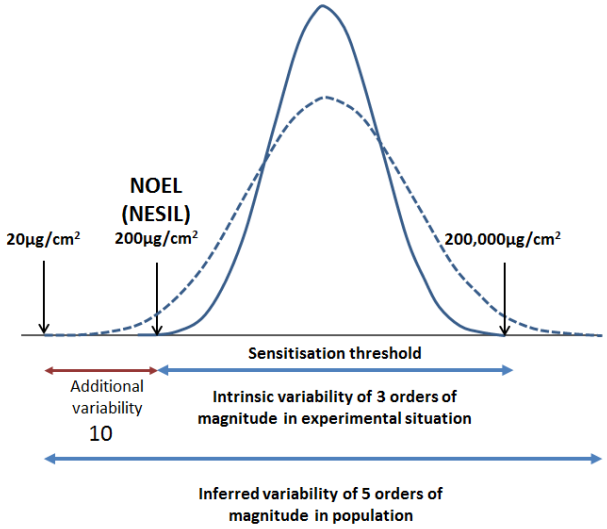
RESPONSE
We understand the point made by the opinion as to failure to fully clarify the use of a SAF of 10 as being sufficient to cover inter-individual variability. In seeking to address this we considered the SCCS Notes of Guidance (SCCS/1564/15), which state that "The default assessment factor of 10 is usually sufficient to protect the larger part of the population, including, e.g., children", particularly in the context of skin exposure.
Applying this concept to the induction of skin sensitization, the factor of 10 is considered supportive to protect the larger part of the population for the following reasons:

- Derived NESIL is in line with NO(A)EL concept
- Available human induction data can be considered a conservative scenario in terms of exposure due to fully occlusive conditions (HRIPT) or local trauma-like conditions (e.g. SLS pre-treatment) in cases where HMT data are available
- Preclinical induction data (e.g. LLNA) provide dose response information in vehicles selected for maximized skin exposure

Consequently, applying the inter-individual SAF of 10 to the concentration where no experimental induction has been observed under the above described conditions is generally considered to cover subjects in the population that may have a threshold lower than the derived NESIL. The current basis is that no data exist that indicate toxicokinetic and/or dynamic parameters relevant for skin sensitization induction (reviewed in Basketter and Safford, 2016) that are significantly different from those relevant for other toxicological endpoints.
In addition, adjustment of the default factor of 10 may be considered on a case-by-case basis taking into account, for example, exposure scenarios indicating that the dose extrapolation from the experimental data underlying the derived NESIL is not relevant, because of enhanced predisposition to sensitization caused by lesions to which fragranced medication will be applied.
The quoted 3-4 orders of magnitude from the publication of Basketter and Safford (2016) refers to the data presented in Tables 2 and 3 of the paper which are derived from induction dose response studies conducted with HMTs and HRIPTs, respectively. The data presented for p-phenylenediamine in which a 100-fold increase in induction concentration increase the proportion of positives from 7% to 53% could be taken to indicate that a 3 to 4 orders of

magnitude reduction in dose would be necessary to reduce the proportion of positively reacting subjects from 100% to 0%. However, it is important to note that these doses represent LOAELs for induction and that from those data it can be taken that the LOAEL may span 3 orders of magnitude.
A single positive reaction in 100+ subjects in an HRIPT is sufficient to establish a LOAEL whereas 0/100+ provides the NOAEL. This NOAEL, taken within a weight of evidence approach, may be used to establish a NESIL. The key question then is to what degree it is necessary to reduce the dose level producing a 0/100+ score in an HRIPT to ensure that no induction will occur in a much larger and more varied population of consumers. The use of the 10x factor for inter-individual variability applied to the NESIL provides an additional order of magnitude of protection below an already established NOAEL. This has been pointed out by Basketter and Safford ("it is important to understand that the inter-individual SAF is not intended to represent the total variability of sensitization threshold values for the entire population"). Further arguments have been provided by Basketter and Safford (2016) to show that a factor of 10 is sufficient.
This can be illustrated as below (Figure 5). Overall it is important to note that the 10-fold inter-individual variability is applied to the NOAEL (NESIL) which sits at the low end of the dose-response curve or total human variability and therefore effectively provides additional safety beyond the normal distribution.

Figure 5 : Impact of including a 10-fold factor for intrinsic variability



12. ATTRIBUTION OF SKIN SITES AND AGGREGATION

SCCS COMMENTS
Page 21: 19-20: QRA2 uses 18 body sites, which is very specific and detailed. A rationale as to why so many different body sites are used in the methodology needs to be provided.
Page 27: 41-43: In this approach aggregation is done for each of the 18 body sites separately. As mentioned earlier, the Final Report on QRA2 needs to explain the rationale for designating these different body sites.
Page 27: 48 - page 28 7: It is well understood that priming of the immune system takes place at the level of the lymph nodes. After dermal absorption, skin sensitisers will induce a local immune response in the skin followed by dislocation of Langerhans cells to the draining lymph nodes and finally followed by T cell priming in the lymph nodes draining the exposed area. It is possible that exposure of different body sites used in QRA2 will target the same lymph node, e.g. the hands are divided in three different body sites but presumably target the same lymph node if exposure is on one side of the body. Evidence for this is provided in a human study from Kligman 1966), who showed that four sequential exposures on different but adjacent sites on one extremity (arm or leg) was far more effective to induce sensitisation than four sequential exposures to each of the four extremities. Kligman

concluded that "bombardment of the same lymph node is superior to stimulation of four different nodal systems".
Mirroring this process within the aggregated exposure assessment is not trivial, since this process is not fully understood. It may impact the risk assessment and may lead to an underestimation of the risk. SCCS finds it of great importance to address and discuss this uncertainty in the Final Report on QRA2 as well.
Page 34: 33-39: The use of 18 body sites in QRA2 needs to be explained and a scientific reasoning for dividing and considering them separately should be provided.
In QRA2 aggregation is done for each of the 18 body sites separately, while aggregation over multiple body sites is not foreseen. There is uncertainty that this may impact the risk assessment, since exposure to certain adjacent body sites may lead to trafficking of Langerhans cells to the same lymph node, leading to a more effective priming of the immune system. This uncertainty needs to be better ad-dressed and discussed.
Page 34: 35-38: In QRA2 aggregation is done for each of the 18 body sites separately, while aggregation over multiple body sites is not foreseen. There is uncertainty that this may impact the risk assessment, since exposure to certain adjacent body sites may lead to trafficking of Langerhans cells to the same lymph node, leading to a more effective priming of the immune system. This uncertainty needs to be better addressed and discussed.

RESPONSE
We would like to put forward the following rationale for selecting 18 application sites:
The following 3 guiding principles were used in deciding on a suitable set of application sites for the CREME RIFM aggregate exposure model.

(1) The set of application sites should cover all areas in which the products populating the model are typically used.

(2) The set of application sites should be mutually exclusive, i.e., non-overlapping.

(3) The set of application sites should be as specific as possible.

The reasoning behind the first of these principles is quite straightforward – if a particular application site is not included, then exposure at this site cannot be calculated. The second principle is based on considering that if the application site definitions overlap, then it could be possible that exposure to the overlapping areas would be underestimated. The third principle is designed to help prevent underestimation of the per unit surface area exposure – which could occur if dermal exposures are averaged over a larger than suitable surface area.

In order to estimate aggregate exposure, it is necessary to have data on quantities, frequency and body-site of use of different products. For this exercise, the data has come mainly from the Kantar World Panel Usage Toiletries and Cosmetics Database (http://www.kantarworldpanel.com/global). Subjects in these surveys, recorded application sites that were assigned according to the 18 skin sites that were used in the calculations of aggregate exposure. This is the origin of the 18 body sites used in our calculations.
With regard to the concern of adequately reflecting aggregate exposure consideration with regard to draining lymph nodes, we agreed with the Opinion as to the role played by those in the induction of contact allergy. In devising the methodology for aggregating exposure, consideration was given to the need to avoid where possible, aggregation of exposures to sites served by completely different draining lymph nodes. At the same time, it was recognised that, for instance, exposure to the palms and back of the hands as well as to the wrist and arms could all implicate the participation of the same families of (supratrochlear) lymph nodes.
It is true that the events in regional (draining) lymph nodes [recognition by responsive T lymphocytes of processed antigen, and T lymphocyte activation and proliferation] are

critical elements for the acquisition of skin sensitization, and also determine the level of sensitization that is induced.
The exception to this rule is where the area of exposure to the inducing chemical allergen is less than 1cm2, in which circumstance the area becomes important. It is well established that - under most conditions of exposure - the important metric in determining the extent of sensitisation that will develop is the dose per unit area of chemical allergen for single exposures, at a given Dose Per Unit Area (DPUA) above 1 cm2, size of area is not a factor – so at that DPUA, applying something once, whether on a small or large area, will have (does have) the same effect.
The implication of the point above is that it will not make any difference if a certain dose per unit area of chemical allergen is applied to one site, or if the same dose per unit area is applied to multiple areas drained by different lymph nodes. In other words, multiple exposure to different sites - at the same DPUA - will have no effect, whereas multiple exposures to the same site will in effect increase the dose of chemical per unit area of skin and enhance sensitization, which is accounted for in the aggregate exposure model.

In consequence, in the Final Report on QRA2, aggregation over multiple body sites is not mentioned because increasing the area of application/exposure has very little or no effect once one is above the small area threshold of 1 cm2. Exposure of multiple sites on the same limb or on multiple limbs has no significant effect.

With regard to the experiment by Kligman (1966), who showed that four sequential exposures on different, but adjacent, sites on one extremity (arm or leg) was far more effective in inducing sensitisation than four sequential exposures to each of the four extremities, here two aspects are mixed. 4 doses repeated on 1 limb is repeated doses (even though not on the same site); 4 single doses scattered around is single doses which is less potent than repeated doses.
Given this, it is less important to get an accurate assessment of the aggregated area. Once it is more than 1 cm2, there are no differences between one or several areas receiving a single exposure – at a constant DPUA. That applies whether the single exposure is applied to a number of areas on the same limb or on different limbs.
But what makes for different effects is when there are repeated exposures. There are two components to this. Firstly, if they are repeated on the same site, then there is an accumulation of chemical so the DPUA is increased. The second point applies both to repeated exposures on one limb and repeated exposures on the same site – there are repeated stimulations of the lymph node system. In the repeats on the same site, this is accompanied by an increase in DPUA. When it is to different sites on the same limb, the DPUA remains constant but the repeated stimulations of the lymph node become a factor. This is what Kligman showed and others (e.g. Friedmann, 2007) confirmed. Repeated exposures to multiple limbs is really about the fact that repeated exposures on one limb gives repeated stimulus to that lymph node system. If it is happening on more than one limb, it is no more powerful than the effect of repeated stimulation on a single limb.
In conclusion,

- If the area of exposure is greater than 1 cm2 of skin then simultaneous aggregate exposure to several different sites will not induce any greater level of sensitisation compared with exposure to a single site. This is because a threshold of immune activation has been reached.
- The situation is different with repeated exposures over time. Repeat exposure to the same site in effect increases the accumulated dose per unit area so the level of sensitisation may be increased.
- If there is repeated exposure to a different site on the same limb then the level of sensitisation may be increased because there is repeated stimulation of the draining lymph node.

Repeated exposure leading to a certain dose per unit area should be adequately addressed by modelling aggregate exposure, leading to a certain CELagg/unit area.

Uncertainty related to the influence of duration and frequency of exposure should be addressed by the respective SAF, as described in section 9.

13. SUSCEPTIBILITY OF CHILDREN

SCCS COMMENTS
Page 23: 20-21: Susceptibility to skin sensitisation of children should not be discussed under the exposure chapter of the final report on QRA2, because this relates to hazard.
Page 34: 11-13: Also, susceptibility to skin sensitisation of children should not be discussed under the exposure chapter of the final report on QRA2, because this relates to hazard.

RESPONSE
We fully agree that the questions of susceptibility to skin sensitisation of children should be discussed under hazard considerations.
In this context, it is worth noting that data suggests that children are not more susceptible to skin sensitisation than adults (Cassimos et al., 1980; Epstein, 1971). The experimental evidence appears to show that young children are less easy to sensitise, such that a risk assessment which is protective for adults should be sufficiently protective for children. A review on developmental immunotoxicology and risk assessment by Holsapple et al. (2004) concluded that current risk practices have been generally shown to be sufficient in protecting children (> 6 months old) and an additional safety factor is not needed to provide additional protection from that which is already achieved.

14. CLARITY IN EXPLAINING CREME/RIFM METHODOLOGY

SCCS COMMENTS
Page 23: 22-36: Furthermore, it is unclear to the SCCS what "full distribution of exposure data" entails. Specifications are needed on which parameters are meant, and what use patterns are considered. It is also not clear whether this section only describes single-product exposures (since aggregate exposure is considered in the next chapter), and what the general purpose is of this chapter.
The Table in Appendix 1 (Table 7 in the final report) has the title "Summary of available habits and practices…product types". It is unclear whether this Table lists the parameters recommended by QRA2 for single-product use, or whether these data are also used for the "Creme RIFM Aggregate Exposure Model".
From the beginning of the chapter on exposure, one gets the impression that the Creme RIFM aggregate exposure model, which according to SCCS knowledge is a probabilistic model based on distributions for selected input data, represents an integral part of the QRA2 approach. However, Appendix 1 only lists point values for the input data and not distributions and only the next chapter lists different data on which the Creme RIFM Aggregate Exposure Model seems to be based. The methodology used for exposure assessment under QRA2 therefore needs to be clarified.
Page 34: 14-16: In regard to aggregate exposure assessment, clarifications are needed on "full distribution of exposure data" in terms of what parameters are meant, and what use patterns are considered, and whether this only refers to single-product exposures.
Page 34: 17-23: An integral part of the QRA2 approach is the use of the Creme RIFM aggregate exposure model, which is a probabilistic model based on distributions for selected input data. The description of the methodology in relation to the use of the model for exposure assessment under QRA2 needs to be clarified, especially in regard to which parameters are treated probabilistically and which as point values. It is also important to point out that the Creme RIFM Aggregate Exposure Model uses the common methodology of probabilistic exposure assessment, which is currently not used or recommended for cosmetics by the SCCS and needs further evaluation by the SCCS.

RESPONSE
CREME RIFM aggregate exposure model

We would like to provide the following enhanced explanation of the CREME RIFM aggregate exposure model in order to address the points above:
As part of the ongoing improvements to the QRA, methodology was developed to account for aggregate exposure of consumers to ingredients in personal care products. Traditional deterministic calculations were considered to provide inaccurate estimates since:

- Consumers are unlikely to use all products under consideration, and certainly not on a daily basis.
- Consumers do not use the same amounts of each of the products.
- The ingredient will not be included in products at the same concentration, and some products will not include the ingredient at all.

Probabilistic modelling overcomes these issues since it uses consumer habits data and manufacturers' product data, and was therefore considered to be the most accurate method of estimating aggregate exposure.
CREME Global (www.cremeglobal.com) and RIFM have developed a model to estimate aggregate exposure to fragrance ingredients, which are used in a range of common consumer products (Comiskey et al., 2015; 2017; Safford et al., 2015; 2017). The model uses probabilistic (Monte Carlo) simulation, sampling from distributions of measured variables, to provide a realistic estimate of aggregate exposure to individuals across a population.
The CREME RIFM aggregate exposure model estimates population aggregate exposure to fragrance ingredients in personal care products, in cosmetics and in a small number of air care products. See Table 3 below for the full list of products. The model can estimate exposure via three routes of exposure - ingestion, inhalation and dermal and, in the latter case, can produce application site specific estimates. The three routes can be combined to estimate systemic exposure. Exposure estimates can be reported in absolute terms or relative to body weight or skin surface area according to the particular exposure in question. The 25 products covered by the model are grouped into nine categories as shown in Table 3 below. The model can produce exposure estimates for each individual product, for each category or for all products. Exposure is estimated for normal use in adult consumers, both male and female. Professional use is currently not covered, nor is exposure in young children or adolescents. The model draws on data obtained from the USA and from Western Europe and so is most reliable in those geographical regions.
Work is ongoing to expand the model to:

- include household cleaning products;
- add several cosmetic and personal care products;
- to add habits and practices data from other European countries which will increase the representativeness of the European population;
- update existing habits and practices survey data to those conducted by Kantar Worldpanel 2013-2014;
- include habits and practices data for 13-17 age group for Europe and the USA and
- investigate whether presence probabilities for fragrance ingredients can be assessed.

Table 3 : CREME RIFM aggregate exposure model products and categories.

Category	Product
Body lotion	Mass market body lotions
	Prestige body lotions
	Other body lotions
Cosmetic styling	Hair spray
	Hair styling

Category	Product
	Lipstick
	Liquid make-up foundation
Deodorant	Body spray
	Roll-on deodorants
	Spray deodorants
Hydroalcoholics	Aftershave
	Eau de Parfum
	Eau de Toilette
Moisturizers	Face moisturizer
	Hand cream
Oral care	Mouthwash
	Toothpaste
Shower products	Rinse-off conditioner
	Shampoo
	Shower gel
Soaps	Bar soap
	Liquid hand soap
Air care	Scented candles
	Air freshner aerosol
	Air freshner plugin

Equation (1) shows the formula for calculation of a subject's dermal exposure, at a particular application site, to a fragrance from a particular product on a particular day.

$$\text{Daily Dermal Exposure (DDE)} = \frac{\text{Frequency} \times \text{Amount} \times \text{Concentration} \times \text{Retention}}{\text{Surface Area}} \quad \textbf{(1)}$$

Where:

Frequency refers to the number of usage occasions of the product;
Amount is the amount of product applied in each application;
Retention is the proportion of the product staying on the body after use;
Concentration is the proportion of the fragrance in the product by mass, and
Surface area is the area of the site of application.

The essence of the CREME RIFM aggregate exposure model is that, in place of using single constant values in equation (1) above, it draws on a database of real-world measurements to calculate tens of thousands of individual consumer exposures to a fragrance from multiple sources. From these individual aggregate exposures, statistical estimates of population exposure are made.

Retention factors are taken from SCCS Notes of Guidance (SCCS/1564/15). The following sections briefly describe the data sources drawn on for the other variables, and how the population estimates are calculated.

Data Sources – Habits and Practices

The Kantar World Panel Usage Toiletries and Cosmetics Database (http://www.kantarworldpanel.com/global) (See Comiskey et al., 2015; 2017 and Safford et

al., 2015; 2017 for details) contains data on the habits and practices of subjects from the United States of America, France, Germany, Great Britain and Spain who use cosmetics and personal care products. Subjects in the surveys were documented concerning their age, gender, sample representativeness of the region, and whether they were habitual users of all personal care cosmetic products, brands and categories. For the purposes of the present study, only subjects from the age of 18 upwards were analysed; yielding a total of 36446 subjects.
These data supply frequency of use and application sites for each product of interest.
Data Sources – Amounts
Estimates of the amounts of each product used were taken from recent consumer studies in the USA and the UK. These studies are described in recent publications (Hall et al., 2007, 2011; Loretz et al., 2005, 2006, 2008; Tozer et al., 2004). Further details are given in Comiskey et al., 2015 and Comiskey et al., 2017.
From these, it was possible to derive probability distribution expressions that describe the variability and range of amounts of each product consumed by the subjects, according to age, gender and country.

Data Sources – Concentrations
Regularly, industry wide surveys are conducted on the concentrations of fragrance ingredients that are added to fragrance mixtures intended for each of the consumer products covered by the CREME RIFM aggregate exposure model. These data are aggregated into product-specific probability expressions that describe the range and variability of fragrance ingredient concentrations in fragrance mixtures.
Similar surveys are conducted among consumer product manufacturers to collect data on concentrations of fragrance mixtures in consumer products. The data are similarly aggregated and the combination of both surveys provides a statistical description of final fragrance ingredient concentrations in consumer products.
Data Sources – Surface Areas
The surface areas of major bodily application sites can be derived from a subject's height (H) and weight (W) using equation (2) below, where a, b and c are age- and gender-specific constants obtained from the Exposure Factors Handbook (EPA, 1997).

$$SA = a \times H^b \times W^c \qquad \textbf{(2)}$$

Paired height and weight data for US subjects were obtained from the NHANES database (https://www.cdc.gov/nchs/nhanes) and these data were scaled using national averages to obtain heights and weights for European subjects.
Equation (2) above provides surface areas for major body parts (head, trunk, legs, and others), finer detail was obtained by deriving surface areas from those parts, for example scalp surface area is taken as half the head surface area, or by using constant values, for example the area around the eyes was taken to be 24 cm2 for all adults. More detail can be obtained from Comiskey, et al., 2015.

Monte Carlo Simulation
The variables above are combined by working through each Kantar subject and assigning values from each source to the subject. The assignment is random, but the probability of assigning any particular value is the same as that value's relative frequency in the population of interest.
As an example, take the height variable. Height is normally distributed, so the data are described by the mean and the standard deviation. Heights at or close to the mean are most likely to be assigned, and the probability of assigning shorter or taller heights drops with increasing distance from the mean, and according to the shape of the normal distribution. The outcome of this is that the distribution of heights among all subjects in the model will match the distribution of heights among real people as measured by the NHANES survey.
Similar considerations apply to the concentration, amount and surface area variables in equation (1). Retention factors are fixed and are a property of the products.

Aggregate Exposure
We have already seen, with equation (1), how daily dermal exposure (DDE) to a fragrance from a single product at a single site is calculated. Aggregate exposure at that site then, as per equation (3) below, is the sum of exposures to all products (1…N) used by the subject in that day.

$$Aggregate\ DDE = \sum\left(DDE_{Product\_1},\ DDE_{Product\_2},\ \ldots,\ DDE_{Product\_N}\right) \qquad \textbf{(3)}$$

This calculation, when applied to the seven-day Kantar diaries, provides us with seven aggregate DDEs for each of the subjects. The mean of these seven values provides us the chronic exposure, while the maximum provides the acute exposure.

Population Exposure Estimate
Repeating the aggregate exposure calculations for all Kantar subjects for all subjects yields 36446 individual exposures (both chronic and acute). A weighting for each subject is also supplied from the Kantar diaries. From this point, it is a straightforward statistical exercise to produce aggregate exposure estimates, along with standard errors or confidence intervals, for the general population. The 95th percentile acute exposure provides the CEL_{agg} used in this report.

15. DERIVING UPPER USE LEVELS

SCCS COMMENTS
Page 26: 12- page 27: 13: In addition to the example of Citral case, a general description of the approach to calculate aggregate exposure is needed. The methodology for deriving the upper use levels needs to be better described. For example, a careful recalculation has shown that Table 3 (Table 8 in the Final report on QRA2, 2016) has listed the upper use levels derived for single products, whereas it is mentioned later in the text that "QRA2 upper use levels assumptions" encompass that "e.g. shower gel is used on palms and face only." It is not clear why this is an important factor for the differences between the product-specific calculation and the aggregate exposure calculation if no aggregation for body parts is done in Table 3. Furthermore, Table 3 should contain all assumptions used and the references for the exposure value, and also that shower gel is considered to be applied only on palms and face. There is also a need to better explain how the product categories A-F were derived. For example, it is not clear what "same exposure" means, and what is common between exposure to make-up remover and lip balm (apart from the obvious application to the face). It is stated that aggregation has been done for each product category separately (last paragraph). From the following sections, it seems that aggregation was also done for different categories (Figure 3; Figure 4 in final report on QRA2, 2016) and this needs to be clarified. Explanation is also needed on why SAFs are different for the same application site for conditioner (100), hair spray (30) and shampoo (300) when in the following chapter, the matrix is always SAF=1.
Many text passages are difficult to understand and do not seem logical; e.g. "in many cases the upper use levels far exceeded realistic industry use levels (e.g. body wash/shower gel, 31.10%; Table 3) due to the assumption that some products are used evenly all over the body leading to a reduced exposure per unit surface area which affords them a greater QRA2 upper use level". It is not clear why the assumption of an even spread of products on the skin is not realistic, and if it is not realistic, why it was not adjusted accordingly. Further down in the paragraph it is described that "subjects in the habits and practices survey applied products in a way that is contrary to the QRA2 upper use levels assumptions". It is not clear then why the QRA2 assumptions were not revisited after this reality check, rather than remediating in a somewhat arbitrary way with the "categorized upper use levels".
From a methodology point of view the SCCS notes that upper use levels for single products are of limited value for use in QRA2, since fragrances will always be applied in multiple products. Therefore, the product-specific "upper use levels" have little relevance, except to derive maximum levels for risk management. In the text, it should be clarified that they

only serve as a starting point for assessing upper use levels based on an aggregate assessment. Also, in the QRA2 report, the upper use level approach has been described as part of the exposure assessment. Instead, they are a means for risk management, and the header or structure of the chapter needs to be changed accordingly.
Page 33: 3-7: Explanation is also needed for the concept of categorisation of the products in broader product categories as well as the use of the product with the lowest upper use level in the aggregate exposure model.
Page 34: 25-32: The use of the methodology for deriving upper use levels needs to be better described along with a general description of the approach to calculate aggregate exposure (see specific comments under 3.4.1.3). From a methodology point of view, the SCCS notes that upper use levels for single products are of limited value for use in QRA2, since fragrances will always be applied in multiple products. Therefore, the product-specific "upper use levels" have little relevance, except to derive maximum levels for risk management. It would be useful to clarify that they only serve as a starting point for assessing upper use levels based on an aggregate assessment.

RESPONSE
We accept that there is a need to improve the descriptions related to aggregate exposure. We also agree that, in the case where aggregate exposure needs to be considered for an ingredient (i.e. a fragrance ingredient used in multiple product types) that, the product-specific "upper concentration levels" are only used as a starting point for assessing upper use levels based on an aggregate assessment. This allows us to derive maximum levels for risk management considering aggregate exposure.
Additional elements describing the methodology for deriving the upper concentration levels have been provided in sections 1 and 16 of this document.
Table 8 in the Final report on QRA2 lists the calculated upper concentration levels as derived for single products. This is based on the NESIL for the specific fragrance ingredients and the high (90th or 95th) percentile product exposure to each application site (Api et al, 2008). For example, the exposure to shower gels is taken from 90th percentile exposures as described by Loretz et al. (2005, 2006, 2008) assuming a full body exposure (16900 cm2) of 25.5 g product per day with an 0.01 retention factor. This is more conservative than the values within the SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation (SCCS/1564/15) where it is assumed 18.67 g are used per day over a body area of 17500 cm2 with a 0.01 retention factor.
The CREME RIFM aggregate exposure model for the calculation of aggregate exposure uses actual reported consumer habits and practices data. Hence, where consumers have reported use of shower gels on hands and face only then this is included in the consideration of exposure to the hands and face body sites respectively during the aggregate exposure calculation. This then provides for additional conservatism relative to the Api et al. (2008) and SCCS Notes of Guidance (SCCS/1564/15) assumptions for shower gel use through modification of the upper use levels for shower gels. Lower upper use level values for product categories are therefore obtained following the aggregate exposure and consumer habits and practices considerations (Table 4).
The references to the exposure values are provided in table 7 of the Final Report on QRA2.
Product categories in the submission were proposed to be based on 6 general categories as described in Section 1 of this document. It is also noted in the submission that "The categorisation shown here is for illustrative purposes and is subject to change, where other products and categories may need to be introduced".

Table 4 : Examples of Kantar survey products per product category as used in the IFRA Standards

General Product Category	**19 Categories for aggregate exposure**	**Kantar survey products**

General Product Category	**19 Categories for aggregate exposure**	**Kantar survey products**
Products applied to the face and body using the hands (palms), primarily leave-on	Body lotion (mass market) Body lotion (prestige) Body lotion (other) Hand cream	Body lotion, body milk, body cream, body butter, body firming/toning moisturiser, other body moisturiser, general purpose Body lotion moisturiser Hand moisturiser Hand and nail moisturiser
Products applied to the axillae	Deodorant spray Deodorant roll-on Body spray	Deodorant spray (i.e. antiperspirant) Roll-on, stick, cream, gel Body spray (not antiperspirant)
Products with oral and lip exposure	Toothpaste Mouthwash	Toothpaste Mouthwash
Products applied to the lips	Lipstick	Lipstick
Products applied to the face using finger tips	Liquid makeup foundation Face moisturizer	Liquid make up foundation Daily face moisturiser, SPF moisturiser, tinted face moisturiser, night face moisturiser, anti-ageing face moisturiser, other face moisturiser
Fine fragrance products	Eau de toilette Eau de parfum After shave	Eau de toilette Eau de parfum Splash-on, aftershave, cologne
Products applied to the hair with some hand contact	Hair styling	Leave in conditioner, mousse, total gel, gel, gel spray, wax, cream, putty, setting lotion, gloss/serum
Products with body and hand exposure, primarily rinse off	Shower gel Shampoo Rinse-off conditioner	Shower gel Shampoo Rinse-off conditioner

16. ADJUSTMENT OF UPPER CONCENTRATION LEVELS – EXAMPLE OF LIPS

SCCS COMMENTS

Page 32:7-10: From a methodology perspective, it is not clear why a contribution of a product category of 84% should result in a reduction of 84% of the upper use level. This needs to be better explained as it implies that all product categories contribute equally to the joint upper use level. If this assumption is correct, then this would need a different algorithm.

RESPONSE

We understand the rationale for this query and offer a further explanation of the methodology used to adjust upper concentration levels to ensure that aggregate exposure remains below an acceptable exposure level below. The inclusion of the consideration of the overall contribution of a product to exposure to a body site is used to ensure, when necessary, a proportional reduction of upper concentration limits where several products are contributing to exposure on the same body site.

Initial upper concentration levels are set for each product type as per the procedure illustrated in Figure 1. The next step is to adjust these upper concentration levels so that AEL_{agg}/CEL_{agg} ratios at all application sites are greater than or equal to 1. We use an iterative approach (beginning with the site with the lowest ratio) whereby products that the application site is exposed to are identified and their upper concentration levels adjusted until the site's AEL_{agg}/CEL_{agg} ratio is increased to at least 1.

When adjusting upper concentration levels for a given application site, we consider the relative contribution to exposure from each product and reduce the products with highest relative contributions the most and products with the lowest relative contributions the least. Note that, since products are generally applied to more than one application site, the process of adjusting upper concentration levels for one site will often increase the AEL_{agg}/CEL_{agg} ratio for other application sites too.

We establish the CEL_{agg} values for each application site by using the CREME RIFM aggregate exposure model (see Section 15 for a description of the model). The table 5 shows the calculation of AEL_{agg}/CEL_{agg} for each application site, ordered from lowest to highest. The lowest ratio is produced from the lips and so this application site will be used as an example to illustrate the procedure described above.

Table 5: Calculation of the ratio AELagg/CELagg for application sites

Application site	Inter-individual SAF	Matrix SAF	Frequency SAF	Skin Condition SAF	Total SAF	NESIL (µg/cm²)	AEL (NESIL/ Total SAF)	CEL_{agg}	AEL_{agg}/CEL_{agg}
Lips	10	1	3	3	100	1400	14	31.1	0.45
Intra-oral	10	1	3	3	100	1400	14	29	0.48
Palms	10	1	3	3	100	1400	14	22.3	0.63
Axillae	10	1	3	10	300	1400	4.7	7.22	0.65
Back of Hand	10	1	3	3	100	1400	14	8.93	1.57
Face	10	1	3	3	100	1400	14	8.37	1.67

Application site	Inter-individual SAF	Matrix SAF	Frequency SAF	Skin Condition SAF	Total SAF	NESIL (µg/cm^2)	AEL (NESIL/ Total SAF)	CEL_{agg}	AEL_{agg}/CEL_{agg}
Neck	10	1	3	3	100	1400	14	6.35	2.2
Ano-genital	10	1	3	10	300	1400	4.7	1.61	2.9
Scalp	10	1	3	1	30	1400	46.7	9.77	4.78
Wrists	10	1	3	3	100	1400	14	2.8	5
Feet	10	1	3	3	100	1400	14	2.65	5.28
Peri-ocular	10	1	3	3	100	1400	14	2.36	5.93
Behind ears	10	1	3	1	30	1400	46.7	4.16	11.22
Legs	10	1	3	1	30	1400	46.7	2.15	21.72
Arms	10	1	3	1	30	1400	46.7	1.71	27.29
Chest	10	1	3	1	30	1400	46.7	1.52	30.7
Abdomen	10	1	3	1	30	1400	46.7	1.52	30.7
Back	10	1	3	1	30	1400	46.7	1.51	30.91

The table 6 below shows that lips have exposure to four product categories (F, D, C, and E) and lists the 95th percentile dermal exposure for each. It can readily be seen that reducing exposure to category F will be much more effective in reducing the lip's AEL_{agg}/CEL_{agg} ratio than reducing the exposure to category E. So, to spread the burden, we link the reduction in upper concentration to the relative contribution to exposure for each category.

To calculate the relative contribution to exposure for each category we sum all four individual exposures (which is not otherwise a meaningful operation) and divide each one by the total. These relative contributions are then used to produce weighting factors for each contribution that are equal to one minus the weighting factor. Finally, the new upper concentration limits are produced by multiplying the current upper limits by the weighting factors.
The AELagg/CELagg ratios are then checked by running a new CREME RIFM exposure assessment using the new upper concentration limits (results not shown). It was found that the ratio for lips was now 1.9, which means that the reduction in upper concentration limits was too severe and so the weighting factors needed to be increased. This was achieved by applying a multiplication factor to each category's relative contribution term and then proceeding as before. Table 7 shows this correction. The value of 0.776 was arrived at by iterative experimentation and produced an AELagg/CELagg ratio for lips of 1.13 which was deemed acceptable.

Table 6: Calculation of approximate relative contribution to aggregate exposure from individual product categories applied to lips to produce upper concentration limit weighting factors

Product Category	95th Percentile Dermal Exposure (µg/cm²)	Relative Contribution	Percentage Relative Contribution	Upper Concentration Limit Weighting Factor
F	28.2	28.2/33.3 = 0.847	84.7	1 - 0.847 = 0.15
D	4.2	4.2/33.3 = 0.126	12.6	1 - 0.126 = 0.87
C	0.7	0.7/33.3 = 0.021	2.1	1 - 0.021 = 0.98
E	0.2	0.2/33.3 = 0.006	0.6	1 - 0.006 = 0.99
Total	**33.3**	**1**	**100%**	**-**

Table 7: Calculation of upper concentration limit weighting factors based on product category contribution and adjustment factor.

Product Category	Relative Contribution	Multiplication Factor	Upper Conc. Limit Weighting Factor
F	0.847	0.776	1 – (0.847*0.776) =0.34
D	0.126	0.776	1 – (0.126*0.776) = 0.9
C	0.021	0.776	1 – (0.021*0.776) = 0.98
E	0.006	0.776	1 – (0.006*0.776) = 1

After these adjustments, the application site with the lowest AEL_{agg}/CEL_{agg} (and still below 1) was palms, so the same process was repeated with palms as the focus.

GLOSSARY OF TERMS

Term	Definition
AEL	Acceptable Exposure Level
AEL_{agg}	Aggregate Acceptable Exposure Level
AOP	Adverse Outcome Pathway
CEL	Consumer Exposure Level
CEL_{agg}	Aggregate Consumer Exposure Level
CET	Closed Epicutaneous Test
DDE	Daily Dermal Exposure

Term	Definition
DEP	Diethyl phtalate
DNCB	Dinitrochlorbenzene
DPUA	Dose per unit area
EC3	Estimated Concentration required to result in a threshold positive response; i.e. a Stimulation Index = 3
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EPA	(US) Environmental Protection Agency
EURL ECVAM	European Union Reference Laboratory for Alternatives To Animal Testing
FCAT	Freund's Complete Adjuvant Test
GPMT	Guinea Pig Maximisation Test
HMT	Human Maximisation Test
HRIPT	Human Repeat Insult Patch Test
IATA	Integrated Approaches to Testing and Assessment
ICD	Irritant Contact Dermatitis
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IDEA	International Dialogue for the Evaluation of Allergens
IFRA	International Fragrance Association
LLNA	Local Lymph Node Assay
LOAEL	Lowest Observed Adverse Effect Level
MEST	Mouse Ear Swelling Test
NESIL	No Expected Sensitisation Induction Level
NHANES	National Health and Nutrition Examination Survey
NDMA	N-Nitrosodimethylamine
NOAEL	No Observed Adverse Effect Level
OET	Open Epicutaneous Test
PPD	*para*-Phenylenediamine
QRA	Quantitative Risk Assessment
QSAR	Quantitative Structure-Activity Relationship
RIFM	Research Institute for Fragrance Materials, Inc.

Term	Definition
SAF	Sensitisation Assessment Factor
SCCS	Scientific Committee on Consumer Safety
SLS	Sodium lauryl sulfate
UCL	Upper Concentration Level
$UCL_{product}$	Upper Concentration Level of a specific ingredient at the individual product level
UCL_{agg}	Aggregate Upper Concentration Level
w_g	Weighting factor
WoE	Weight of Evidence

REFERENCES:

1. Api, A.M., Basketter, D.A., Cadby, P.A., Cano, M.-F., Ellis, G., Gerberick, F., Griem, P., McNamee, P.M., Ryan, C.A., Safford, B. (2008) Dermal Sensitization Quantitative Risk Assessment (QRA) for fragrance ingredients. Regulatory Toxicology and Pharmacology. 52, 3-23.
2. Api, A.M., Basketter, D.A., Lalko, J. (2015) Correlation between experimental human and murine skin sensitization induction thresholds. Cutaneous and Ocular Toxicology. 34(4), 298-302.
3. Basketter D.A., Jefferies D., Safford B.J., Gilmour N.J., Jowsey I.R., McFadden J., Chansinghakul W., Duangdeeden I., Kullavanijaya P. (2006) The impact of exposure variables on the induction of skin sensitization. Contact Dermatitis. 2006 Sep;55(3):178-85.
4. Basketter D., Safford B. (2016) Skin sensitization quantitative risk assessment: A review of underlying assumptions. Regul Toxicol Pharmacol. 2016 Feb;74:105-16. doi: 10.1016/j.yrtph.2015.11.013.
5. Bil W., Schuur A.G., Ezendam J., Bokkers B.G.H. (2017). Probabilistic derivation of the interspecies assessment factor for skin sensitization. Regulatory Toxicology and Pharmacology 88, 34-44.
6. Cassimos C., Kanakoudi-Tsakalidis F., Spyroglou K., Ladianos M., Tzaphi R. (1980) Skin sensitization to 2,4- dinitrochlorobenzene (DNCB) in the first months of life. J Clin Lab Immunol Mar;3(2):111-3.
7. Comiskey, D., Api, A.M., Barratt, C., Daly, E.J., Ellis, G., McNamara, C., O'Mahony, C., Robison, S.H., Safford, B., Smith, B., Tozer, S. (2015). Novel database for exposure to fragrance ingredients in cosmetics and personal care products. Regulatory Toxicology and Pharmacology. 72(3):660-72. doi: 10.1016/j.yrtph.2015.05.012
8. Comiskey, D., Api, A.M., Barratt, C., Ellis, G., McNamara, C., O'Mahony, C., Robison, S.H., Rose, J., Safford, B., Smith, B., Tozer, S. (2017). Integrating habits and practices data for soaps, cosmetics and air care products into an existing aggregate exposure model. Regulatory Toxicology and Pharmacology. 88: 144-156.
9. De Jong W.H., De Klerk A., Ter Beek M., Veenman C., Van Loveren H. (2007) Effect of Prolonged Repeated Exposure to Formaldehyde Donors with Doses Below the EC3 Value on Draining Lymph Node Responses. Journal of Immunotoxicology, Volume 4, 2007 - Issue 3, Pages 239-246.
10. Dourson M.L, Felter S.P., Robinson D. (1996) Evolution of Science-Based Uncertainty Factors in Noncancer Risk Assessment. Regulatory toxicology and Pharmacology 24(2), 108-120.
11. Dourson (1996) Editorial: Uncertainty Factors in Noncancer Risk Assessment. Regulatory toxicology and Pharmacology 24(2), 107.
12. ECETOC (2003). Contact Sensitisation: Classification According to Potency. Technical Report No. 87. European Centre for Ecotoxicology and Toxicology of Chemicals. ISSN-0773-8072-87. Brussels, April 2003.
13. EPA, (1997). Exposure Factors Handbook. Doc EPA/600/P-95/002Fa. Office for Research and Development, U.S. Environmental Protection Agency, Washington, DC.
14. Epstein E. (1971) Contact dermatitis in children. Pediatr Clin North Am. Aug;18(3):839-52.

15. Ezendam J., Park M., Salverda N., Nijhof J.G.W. (2013). A quantitative approach to assess the risk of skin sensitisation from hair dye ingredients. A case study using p-phenylenediamine (PPD). RIVM Letter report 050012001/2013. Dutch National Institute for Public Health and the Environment. http://www.rivm.nl/bibliotheek/rapporten/050012001.pdf
16. Feldmann R.J., Maibach H.I. (1967) Regional variation in percutaneous penetration of 14C cortisol in man. J Invest Dermatol. Feb;48(2):181-3.
17. Friedmann, P.S. (2007) The relationships between exposure dose and response in induction and elicitation of contact hypersensitivity in humans. Br J Dermatol., Dec;157(6):1093-102.
18. Goebel C., Diepgen T.L., Krasteva M., Schlatter H., Nicolas J.F., Blömeke B., Coenraads P.J., Schnuch A, Taylor JS, Pungier J, Fautz R, Fuchs A, Schuh W, Gerberick GF, Kimber I. Quantitative risk assessment for skin sensitisation: consideration of a simplified approach for hair dye ingredients. Regul Toxicol Pharmacol. 2012; 64:459-65.
19. Hall B., Tozer S., Safford B., Coroama M., Steiling W., Leneveu-Duchemin M.C., McNamara C., Gibney M. (2007) European consumer exposure to cosmetic products, a framework for conducting population exposure assessments. Food and Chemical Toxicology, 45(11), 2097-2108.
20. Hall, B., Steiling, W., Safford, B., Coroama, M., Tozer, S., Firmani, C., McNamara, C., Gibney, M. (2011) European consumer exposure to cosmetic products, a framework for conducting population exposure assessments Part 2. Food and Chemical Toxicology, 49:408-422.
21. Holsapple, M.P., Paustenbach, D.J., Charnley, G., West, L.J., Luster, M.I., Dietert, R.R., Burns-Naas, L.A., (2004). Symposium summary: children's health risk--what's so special about the developing immune system? Toxicology and applied pharmacology 199, 61-70.
22. International Dialogue for the Evaluation of Allergens (IDEA) (2014). Report on the IDEA Workshop on Validity of the QRA Methodology and Possibilities of Further Refinement, May 13-15, 2014. http://www.ideaproject.info/uploads/Modules/Documents/idea-qra-workshop-(may-13-15-2014)---final-progress-report-4-9-14.pdf .
23. International Dialogue for the Evaluation of Allergens (IDEA) (2016). Final report on QRA2. http://www.ideaproject.info/uploads/Modules/Documents/qra2-dossier-final--september-2016.pdf
24. Jowsey I.R., Clapp C.J., Safford B., Gibbons B.T., Basketter D (2008) The impact of vehicle on the relative potency of skin-sensitizing chemicals in the local lymph node assay. Cutaneous and Ocular Toxicology Volume 27 - Issue 2.
25. Kligman A.M. (1966) The SLS provocative patch test in allergic contact sensitization. J Invest Dermatol. Jun;46(6):573-83.
26. Lalko J., Isola D., Api AM. (2004) Ethanol and diethyl phthalate: vehicle effects in the local lymph node assay. Int J Toxicol. 2004 May-Jun;23(3):171-7.
27. Loretz, L.J., Api, A.M., Barraj, L.M., Burdick, J., Dressler, W.E., Gettings, S.D., Han Hsu H., Pan, Y.H., Re, T.A., Renskers, K.J., Rothenstein, A., Scrafford, C.G., Sewall, C. (2005). Exposure data for cosmetic products: lipstick, body lotion, and face cream. Food Chem. Toxicol. 43 (2), 279–291.
28. Loretz, L., Api, A.M., Barraj, L., Burdick. J., Davis de, A., Dressler, W., Gilberti, E., Jarrett, G., Mann, S., Laurie Pan, Y.H. et al. (2006) Exposure data for personal care products: hairspray, spray perfume, liquid foundation, shampoo, body wash, and solid antiperspirant. Food Chem Toxicol 2006, 44:2008- 2018.
29. Loretz, L.J., Api, A.M., Babcock, L., Barraj, L.M. and Burdick, J. (2008). Exposure data for cosmetic products: Facial cleanser, hair conditioner, and eye shadow. Food and Chemical Toxicology, 46(5), 1516-1524.
30. Politano, V.T., Api, A.M., 2008. The Research Institute for Fragrance Materials' human repeated insult patch test protocol. Regulatory Toxicology and Pharmacology. 52, 35-38.
31. Roberts D.W., Williams D.L. (1982) The derivation of quantitative correlations between skin sensitisation and physico-chemical parameters for alkylating agents, and their application to experimental data for sultones. J theor Biol. Dec 21;99(4):807-25.
32. Safford, B., Api, A.M., Barratt, C., Comiskey, D., Daly, E.J., Ellis, G., McNamara, C., O'Mahony, C., Robison, S., Smith, B., Thomas, R., Tozer, S., (2015). Use of an aggregate exposure model to estimate consumer exposure to fragrance ingredients in personal care and cosmetic products. Regulatory toxicology and pharmacology, 72: 673-682.
33. Safford, B., Api, A.M., Barratt, C., Comiskey, D., Ellis, G., McNamara, C., O'Mahony, C., Robison, S., Rose, J., Smith, B., Tozer, S., 2017. Application of the expanded Creme RIFM consumer exposure model to fragrance ingredients in cosmetic, personal care and air care products. Regulatory Toxicology and Pharmacology 86 (2017) 148-156.
34. SCCS (Scientific Committee on Consumer Safety), opinion on fragrance allergens in cosmetic products, 26-27 June 2012, SCCS/1459/11.
35. SCCS (Scientific Committee on Consumer Safety), Memorandum on use of Human Data in risk assessment of skin sensitisation, 15 December 2015, SCCS/1567/15.

36. SCCS (Scientific Committee on Consumer Safety), SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation 9 th revision, 29 September 2015, SCCS/1564/15 (revision of 25 April 2016).
37. Tozer, S.A., O'Keeffe, L., Cowan-Ellsberry, C.E. and Rich K. (2004). Use of probabilistic analysis in the refinement of exposure data for hydroalcoholic perfume products. Toxicology, 202(1-2), 123-124.
38. Zuang V., Barroso J., Belz S., Berggren E., Bernasconi C., Bopp S., Bouhifd M., Bowe G., Campia I., Casati S., Coecke S., Corvi R., Dura A., Gribaldo L., Grignard E., Halder M., Holley T., Janusch Roi A., Kienzler A., Lostia A., Madia F., Milcamps A., Morath S., Munn S., Paini A., Pistollato F., Price A., Prieto-Peraita P., Richarz A., Triebe J., van der Linden S., Wittwehr C., Worth A., Whelan M., EURL ECVAM Status Report on the Development, Validation and Regulatory Acceptance of Alternative Methods and Approaches (2017), EUR 28823, Publications Office of the European Union, Luxembourg, 2017, Available at http://publications.jrc.ec.europa.eu/repository/handle/JRC108831.

ANNEX
Interspecies factor. Comments on recent publication of Bil et al. (2017)

The following text will be redrafted and submitted as a short communication to Regulatory Toxicology and Pharmacology for review in the open literature. However, we believe that it covers all elements to be included in our submission.
We have read with interest the recent publication by Bil et al. (2017)1. The paper provides information on the application of an additional safety assessment factor (SAF) to address interspecies variability in the absence of human data where only an LLNA is available for skin sensitization risk assessment. A key consideration for the risk assessment of sensitizers, is the derivation, through a weight of evidence (WoE) approach, of a NESIL or the human threshold for induction of skin sensitization as the point of departure for the risk assessment. Currently, except in exceptional cases, the correlation between human data and the LLNA do not suggest the need for additional assessment factors (Basketter et al (1999)2; (2000)3; (2005)4 (2008)5; (2011)6; (2012)7 (2018)8; Api et al (2015)9; Gerberick et al (2001)10; (2004)11; Griem et al (2003)12; Schneider and Akkan, (2004)13. However, we do recognize that during the derivation of the WoE NESIL, if there are only limited data (e.g. a single LLNA) factors such as assay variability should be considered in deriving the WoE NESIL. We would like to challenge the authors conclusion that "It can be concluded that a murine-based NESIL requires the use of an interspecies SAF (of 15) in the Quantitative Risk Assessment (QRA) for skin sensitization, to correct for the differences between mice and humans" which we do not believe is justified for the following reasons:
The NESIL is defined as the quantitative threshold exposure level that is considered not to induce skin sensitisation in humans. It is the toxicological threshold that is used as the point of departure for the QRA. A Weight of Evidence (WoE) approach is utilized in the determination of a NESIL enabling use of all available data and a more scientifically valid to estimate the allergenic potency of a substance for its risk assessment. The NESIL can be established using "NOEL" data from animal studies, especially the murine LLNA, and taking existing (historical) human studies into account. Whenever available reliable human data would take precedence over animal data, a reflection that the NESIL is defined as a human threshold. Adjustments of thresholds derived from any source other than human to derive a NESIL should be made in the process of derivation of the NESIL, i.e. on the hazard side in the QRA approach and not by application of a generic interspecies adjustment factor to derive the AEL.
In the absence of human data, the risk assessor needs to ensure that the necessary adjustment of a threshold in µg/cm2 value is done to reflect the situation in humans and to determine a NESIL. It is recognized that the LLNA is subject to variability and such variability should be considered in situations where only a single LLNA is available to derive a NESIL. In the future, the basis of a NESIL will be data from alternative assays and not the LLNA; the appropriate adjustment factors reflecting the uncertainty and the translation to a human threshold will need to be applied. Depending which data are the basis for the NESIL, an adjustment could be done in various ways. It is therefore essential to maintain a consistent definition of the NESIL as the quantitative threshold exposure level that is considered not to induce skin sensitization in humans.
The ratio of the geometric means of the distributions is effectively 1. This indicates that there is not a systematic species difference, i.e. mice are not systematically less (or more) sensitive than humans. Use of an interspecies factor thus means that the authors suggest an additional safety factor of 15 is applied, because there is a variation for different molecules regarding how well the LLNA predicts the human value. The proposed SAF is based on this variability and it is strongly influenced by the probabilistic assessment made, by the 95th percentile chosen as threshold and by outliers in the dataset. Such a SAF would only be required on a case by case basis where, for example, knowledge of the physical or chemical properties of the substance provide a reason to expect a poor prediction, i.e. an underestimation of potency, or if the data basis is very small (e.g. only one LLNA and the chemical belongs to new class of chemicals with no supporting SAR comparisons). Accounting for this uncertainty would, as stated above, be better considered at the NESIL setting stage on a case-by-case basis rather than applying an additional SAF.

For fragrance chemicals, the HRIPT data may deliver the final confirmation that an important difference between mouse and man is not being missed.
Attribution of the variation between the human data and the LLNA data observed by Bil et al. to an interspecies effect is premature and without biological evidence. Whilst we must accept that there may be an interspecies effect in certain cases, the observed variation is due, at least partly, to inter-test effects / test variability, both for the LLNA and for the human data (Hoffmann (2015)14; Roberts et al (2016)15). Vehicle effects are also known to play a role in inter-test variability (Anzai et al (2010)16; Jowsey (2008)17). A statistical analysis of the LLNA variability in the Hoffmann database reveals this. When asking, "What is the 95th -percentile that one LLNA gives the correct NESIL derived from another LLNA on the same molecule?" one comes to exactly the same conclusion, simply based on the 95th percentile for inter-LLNA variability of EC3 values. To be 95% sure that the one LLNA value is within the 95% confidence interval of another LLNA result, one would to have to add a SAF of 14 to account for LLNA variability. Repeated studies on humans are scarce and this analysis cannot be done on the human data, but we would expect at least this amount of inter-test variability in human data when looking at the 95th percentile, given the fact that human studies are done at fewer doses and with less standardized protocols.
Over 50% of substances reviewed in the paper are fragrance materials. The decisions for choosing test concentrations in the human test for these, and possibly many of the other substances, has been based primarily on levels of use (as a confirmation of safety) and not on potency. Of course where LLNA data indicate that use-level based test concentrations could lead to reactions in the human tests, these concentrations have been lowered as a precaution, but when this is not the case, the quoted "NOELs" are potentially well below the true maximum NOELs. Obviously, these could only have been confirmed by the conduct of human tests at higher dose levels.
The assessment is heavily influenced by one value (benzosiothiazolinone; LLNA:human ratio 49 fold). Using a 95th percentile on a relatively small dataset containing such an extreme outlier heavily influences the analysis.
For the 13 fragrance materials in the dose-response dataset (i.e. the dataset mostly used by Bil et al.), the maximal LLNA:Human ratio is 2.6. Thus, the data on fragrance molecules in this data set strongly argue against the proposed SAF for QRA on fragrance materials.
Similarly, in the full interpolation dataset, which is used as a confirmatory dataset by Bil et al., 34 fragrance molecules are included. The maximal fold-difference is 19 for trans-2-hexenal, but this is due to the high volatility of 2-Hexenal, which in an openly applied test such as the LLNA is indicated as a weak sensitizer due to rapid evaporation but is a strong sensitizer in human tests using occlusion. This substance is also rated strong and highly reactive in occluded in vitro and reactivity tests (Roberts and Natsch (2009) 18. The second highest value is for methyl-2-nonynonate (7.9-fold lower human vs. LLNA) – but this chemical was tested in one LLNA test, while the homologue methyl-2-octynoate is rated 5-times stronger in the LLNA. These differences are clearly not intrinsic to the small molecular difference between the two molecules but rather to LLNA variability. Thus, both cases illustrate the effect of variability in EC3 values from single LLNA and the effects of test conditions, rather than true interspecies effects. Indeed, for the 34 fragrance molecules in this data-set, excluding these two cases, which can easily be explained by physico-chemistry and data variability, the geometric mean of the ratio is 0.9 and the average 1.7, again showing no systematic under-prediction of human potency by mouse data.
The paper refers to a "switch to probabilistic assessment instead of deterministic SAF's". However, the approach uses a probabilistic assessment at one stage (interspecies) to come to a new deterministic SAF, which you would then sum up with other deterministic SAFs. A true probabilistic approach would require a probabilistic approach at all SAF stages and also combine the SAFs by probabilistic calculations to obtain an overall probabilistic value for safe use levels.
Often the confidence intervals based on small datasets are high and they give us little information about the true shape of the dose-response curve one is trying to fit. The paper, when modelling the dose-response curves for human data, states that, "When the lower and upper confidence bound varied more than a factor 100 (e.g. due to a limited number of tested doses situated away from the dose of interest, the DSA05), the compound was

excluded from further analysis." This gives an indication of how big the confidence intervals are in this study.
It would have been helpful if the authors had quantified how the final SAF that they have derived would have been impacted by the various assumptions that they have made (e.g. the core assumption of the log-normal distribution) and the accuracy of the estimated parameters of the distribution, on the final SAF derived. For example, the confidence range for the GSD (full interpolation, N=63) is 4.0-7.4. It's not clear how this range is reflected in the probabilistic framework.

(1) Bil, W., Schuur, A. G., Ezendam, J. and Bokkers, B. G. H. (2017) Probabilistic derivation of the interspecies assessment factor for skin sensitization. Regulatory Toxicology and Pharmacology 88, 34-44.
(2) Basketter, D.A., Lea, L.J., Dickens, A., Briggs, D., Pate, I., Dearman, R.J., Kimber, I., 1999. A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. Journal of Applied Toxicology 39 (4), 621–627.
(3) Basketter, D.A., Blaikie, L., Dearman, R.J., Kimber, I., Ryan, C.A., Gerberick, G.F., Harvey, P., Evans, P., White, I.R., Rycroft, R.J.G., 2000. Use of the local lymph node assay for the estimation of relative contact allergenic potency. Contact Dermatitis 42 (6), 344–348.
(4) Basketter DA, Clapp C, Jefferies D, Safford RJ, Ryan CA, Gerberick GF, Dearman RJ and Kimber I. (2005) Predictive identification of human skin sensitisation thresholds. Contact Dermatitis, 53, 260 - 267.
(5) Basketter DA, Darlenski R and Fluhr J. (2008) Skin irritation and sensitization: mechanisms and new approaches for risk assessment. Part II: Skin sensitization. Skin Pharmacology and Physiology, 21, 191 - 202.
(6) Basketter DA and Kimber I (2011) Predictive tests for irritants and allergens and their use in quantitative risk assessment. In "Contact Dermatitis", 5th edition, chapter 13. Springer, Berlin, Eds Johansen JD, Frosch PF and Lepoittevin J-P, pp 229-240.
(7) Basketter DA and McFadden JP (2012) Cutaneous allergies. In "Immunotoxicity, Immune Dysfunction and Chronic Disease. Eds Dietert RR and Luebke RW., Humana Press, New York, pp 103-126.
(8) Basketter DA, Kimber I and Kolle SNE (2018) Contact hypersensitivity. In: Comprehensive Toxicology, Ed: McQueen, CA, 3rd edition, vol. 11, pp 582-598.
(9) Api AM, Basketter DA and Lalko J (2015) Correlation between experimental human and murine skin sensitization induction thresholds. Cut Ocul Toxicol, 34, 298-302.
(10) Gerberick, G.F., Robinson, M.K., Ryan, C.A., Dearman, R.J., Kimber, I., Basketter, D.A., Wright, Z., Marks, J.G., 2001a. Contact allergenic potency: correlation of human and local lymph node assay data. American Journal of Contact Dermatitis 12 (3), 156–161.
(11) Gerberick, G.F., Ryan, C.A., Kern, P.S., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2004. A chemical dataset for evaluation of alternative approaches to skin-sensitization testing. Contact Dermatitis 50 (5), 274–288.
(12) Griem, P., Goebel, C., Scheffler, H., 2003. Proposal for a risk assessment methodology for skin sensitization potency data. Regulatory Toxicology and Pharmacology 38, 269–290.
(13) Schneider, K., Akkan, Z., 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. Regulatory Toxicology and Pharmacology, 245–255.
(14) Hoffmann, S. (2015) LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. ALTEX 32, 379-383.
(15) Roberts DW., Api AM., Aptula AO., (2016) Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitisation potency. Part 2. The biological variability of the murine Local Lymph Node Assay (LLNA) for skin sensitisation. Regulatory Toxicology and Pharmacolog 80:255-9.
16) Anzai T., Ullmann LG., Hayashi D., Satoh T., Kumazawa T., Sato K. (2010) Effects of strain differences and vehicles on results of local lymph node assays. Exp Anim. 2010;59(2):245-9.
(17) Jowsey IR., Clapp CJ., Safford B., Gibbons BT., Basketter DA. (2008) The impact of vehicle on the relative potency of skin-sensitizing chemicals in the local lymph node assay. .Cutan Ocul Toxicol. 27(2):67-75.
(18) Roberts, D. W. and Natsch, A. (2009) High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of Michael acceptor electrophiles. Chem. Res. Toxicol. 22, 592-603.

Appendix 2: Summary of Available Habits and Practices and Human Parameters Data Used in the Calculation of Consumer Exposure to Different Product Types (Table 7 in the Final report on QRA2)

(Exposures used in the QRA methodology are shown in bold-face and highlighted)

Product Type	Surface Area cm^2	Surface Area Reference	Retention Factor [1]	SCCS Notes of Guidance, 8th Revision, 2012		Loretz *et al*., 2005; 2006; 2008		Tozer *et al*. 2004; Cano, 2006	Hall *et al*., 2007; 2011; Steiling *et al*., 2012		HERA [1]	Api *et al*., 2007	Cowan-Ellsberry *et al*., 2008	RIFM [2]
				mg/d	$mg/cm^2/d$	90th Percentile		95th Percentile	90th Percentile		$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$
						mg/d	$mg/cm^2/d$	$mg/cm^2/d$	mg/d	$mg/cm^2/d$				
Deo non-spray	100	Bremmer, 2003, per axillae	1	1500	7.5				1500	7.5				
Deo aerosol Spray	100	Bremmer, 2003, per axillae	1	1430	7.2				1430	7.2				
Deo Spray (not ethanol based)	100	Bremmer, 2003, per axillae	1	690	3.5				6910	3.5				
Solid AP	96.8	Cowan- Ellsberry *et al.,* 2008, per axillae	1			1700	8.50						9.1**	
Shaving Cream/ Depilatory [3]	305	Bremmer, 2003 (1/4 area head, male)	0.01	2000	0.07									
Lip Products	4.8	Ferrario *et al.*,2000	1	57	11.9	55	11.46		56.53	11.8				
Eye Products [5]	24	Bremmer, 2003	1	20	0.83	52	2.17							
Body Cream/Lotion [6]	12895	EPA, 1997 (area body - head and ½ trunk, female) [10]	1	7820	0.6	14400	1.12		7800	0.60				
Men's Facial Cream	775	Bremmer, 2003 (1/4 area head + 1/2 area hands, male)	1	1540	2.0									

Product Type	Surface Area cm^2	Surface Area Reference	Retention Factor[1]	SCCS Notes of Guidance, 8th Revision, 2012		Loretz *et al*., 2005; 2006; 2008		Tozer *et al*. 2004; Cano, 2006	Hall *et al*., 2007; 2011; Steiling *et al*., 2012		HERA[1]	Api *et al*., 2007	Cowan-Ellsberry *et al*., 2008	RIFM[2]
				mg/d	$mg/cm^2/d$	90th Percentile		95th Percentile	90th Percentile		$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$
						mg/d	$mg/cm^2/d$	$mg/cm^2/d$	mg/d	$mg/cm^2/d$				
Toothpaste	216.8	Collins et al., 1987; Ferrario et al., 2000 (buccal + lips)	0.1[11]	2750	1.27				2750	1.27				
Mouthwash	216.8	Collins *et al*., 1987; Ferrario *et al*., 2000 (buccal + lips)	0.01[11]	21600	1.0				21620	1.0				
Hydroalcoholic Products for Shaved Skin	775	Bremmer, 2003 (1/4 area head + 1/2 area hands, male)	1					2.21						
Hydroalcoholic Products for Unshaved Skin	100	Bremmer, 2003, perfume spray	1			1770	17.70	2.21						
Women's Facial Cream	555	EPA, 1997 (1/2 area head, female)	1	1540	2.8	3500	6.31		1540	2.8				
Women's Facial Liquid Make-up	555	EPA[3] (1/2 area head, female)	1	510	0.92	1760	3.17		513	0.92				
Hair Sprays – Aerosol[8]	555	EPA, 1997(1/2 area head, female)	0.1			7730	1.39							
Hair Sprays - Pump Spray[8]	555	EPA, 1997(1/2 area head, female)	0.1			12220	2.20***							
Hair Styling Aids	1010	Bremmer, 2003 & EPA, 1997 (1/2 area hands +1/2 head)	0.1	4000	0.4				4000	0.4				

Product Type	Surface Area cm^2	Surface Area Reference	Retention Factor [1]	SCCS Notes of Guidance, 8th Revision, 2012		Loretz *et al*., 2005; 2006; 2008		Tozer *et al*. 2004; Cano, 2006	Hall *et al*., 2007; 2011; Steiling *et al*., 2012		HERA [1]	Api *et al*., 2007	Cowan-Ellsberry *et al*., 2008	RIFM [2]
				mg/d	$mg/cm^2/d$	90th Percentile		95th Percentile	90th Percentile		$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$
						mg/d	$mg/cm^2/d$	$mg/cm^2/d$	mg/d	$mg/cm^2/d$				
Shampoo	1430	EPA, 1997 (area hands + 1/2 head)	0.01	10460	0.07	23630	0.17		10460	0.07				
Conditioners, Rinse-off	1430	EPA,1997 (area hands +1/2 head)	0.01	3920	0.03	28200	0.20							
Make-up Remover	555	EPA, 1997 (1/2 area head, female)	0.1	5000	0.90									
Nail care	11	RIVM [2]	0.1	107.5	0.97									
Bar Soaps	840	EPA, 1997 (area hands)	0.01	20000	0.2									
Liquid Soap	840	EPA, 1997 (area hands)	0.01	20000	0.2									
Hand Cream	840	EPA, 1997 (area hands	1	2160	2.6				2160	2.6				
Face Washes, Gels, Scrubs	555	EPA, 1997 (1/2 area head, female)	0.01			8300	0.15							
Body WashGels, Foams, Mousses	16900	EPA, 1997 (body area, female)	0.01			25500	0.015							
Bath Foams, Gels, Mousses	16900	EPA, 1997 (body area, female)	0.01	18670	0.010				18670	0.010				
Feminine Hygiene - Tampons														2.9

Product Type	Sur-face Area cm^2	Surface Area Reference	Reten-tion Factor[1]	SCCS Notes of Guidance, 8th Revision, 2012		Loretz *et al*., 2005; 2006; 2008		Tozer *et al*. 2004; Cano, 2006	Hall *et al*., 2007; 2011; Steiling *et al*., 2012		HERA[1]	Api *et al*., 2007	Cowan-Ellsberry *et al*., 2008	RIFM[2]
				mg/d	$mg/cm^2/d$	90th Percentile		95th Percentile	90th Percentile		$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$	$mg/cm^2/d$
						mg/d	$mg/cm^2/d$	$mg/cm^2/d$	mg/d	$mg/cm^2/d$				
Feminine Hygiene - Pads														0.14
Feminine Hygiene - Liners														0.14
Baby Diapers														0.0006
Baby Wipes														4.0
Intimate Wipes														4.4
Aerosol Air Freshener	3425	EPA, 1997 (1/2 area head + upper extremities, female)	1											0.025
Hand wash Laundry											0.1			
Laundry Tablets & Powder												Insigni-ficant		
Hand Dishwashing												0.01		
Fabric Clothing												Insigni-ficant		
Hard Surface Cleaner												0.12		
Candles												0.00033		

1 **This exposure value is used in the QRA for fragrance ingredients for all types of deodorants and antiperspirants.
2 ***This exposure value is used in the QRA for fragrance ingredients for all types of hair sprays.
3 Note: Products that contain sunscreen are not addressed separately but are included in the major product type (e.g. lip creams with sunscreen are included in lip
4 product category).

1) HERA, Technical Guidance Document, 2003.
2) RIFM, 2005, AM Api, Internal memo December 12, 2005, on dermal exposure to pressurised aerosol air fresheners. RIFM, 2006, Memo to AM Api from RIFM Member Company, May 2006 on exposure to feminine hygiene products and baby wipes.
3) Shaving cream/depilatory cream products – the amount used was derived from the EC, 1996 Technical Guidance Document in support of Commission Directive 93/67/EEC on risk assessment for new notified substances and Commission Regulation (EC) No 1488/94 on risk assessment for existing substances. This reference did not distinguish between shaving the face or shaving the leg. As such, the dose/unit area for shaving the face was calculated and the same value was applied to shaving or depilating the legs. In the absence of more robust data, this was assumed to be a reasonable and conservative approach.
4) For frequency of use less than once per day, the default of once per day was used with the exception of nail care products where a frequency of 0.43 was used.
5) Eye products – This is based on the Loretz *et al*. 2008 measured data for all types of eye shadows from a specifically designed exposure study for eye products. The SCCS, 2012 exposure data on mascara product types were not used for the eye product category because there is little if any skin contact from this product type.
6) Body cream/lotion – The surface area comprises the total body surface area for a female minus the area of the head and half the trunk. This is based on habits and practices data for adults that indicate that body lotion is not applied to the head or the back.
7) These are product dilution factors. Different dilution factors are used for mouthwashes and toothpastes. The dilution factor used for mouthwashes is 1% or 0.01 and that used for toothpastes is 10% or 0.1. These values are different from the values used in the SCCS 2012 Guidelines, but considered to be more relevant since it takes into account the amount remaining in the oral cavity and perioral area rather than that ingested. It also takes into account salivation and distribution across the oral cavity surface (Muhlemann and Rudolf, 1975; Zero *et al*., 1988; Issa and Toumba, 2004). The difference in the dilution factors used for mouthwashes and toothpastes is based on the fact that while very different volumes of each product are applied (i.e. 30 g/day of mouthwash vs. 2.7 g of toothpaste), it is reasonable to expect that similar amounts of product would be in contact with the mouth (buccal cavity and lips) at any one time since the same surface area is involved. The exposure to oral care products (toothpastes and mouthwashes) is impacted by salivation, product dilution and distribution across the oral surfaces and the focus for sensitisation reactions is the perioral area. As such, in order to benchmark against the exposure approach used here, a worst case exposure scenario was evaluated using the principles of HERA. In HERA, it was assumed that a 0.01 cm film thickness was left on the skin (Vermeire *et al.*, 1993) from a 10% aqueous product solution. This would result in a worst case exposure of 1mg/cm^2, assuming 100% retention of the fragrance ingredient from the product solution. This is consistent with the value identified by the primary exposure approach.

Hair Spray – exposure for the pump spray is recommended for all hair sprays since this figure was the most conservative (e.g. highest) value