Comment Number	Test Guideline Paragraph (¶)	Comments
1	General comment	The current draft Test Guideline (TG) states that these two test methods are only applicable to the EU hazard classification system, which would not justify development as OECD test guidelines. However, recent discussions with the European Centre for the Validation of Alternative Methods (ECVAM) suggest that these methods may also have some limited applicability to the United Nations Globally Harmonized System for the Classification and Labeling of Chemicals (GHS) (see below and **Table 1**). Data indicate that they may be useful for identifying GHS Category 2 irritants, but not GHS Category 3 irritants. Given the international applicability of the GHS, the test guideline should therefore discuss the potential usefulness of human skin cell models for identifying GHS Category 2 skin irritants. The usefulness for the EU system should not be the focus given that its applicability is limited to EU member countries. The ECVAM Skin Irritation Validation Study (SIVS) referenced in the draft TG, two human skin cell model test systems were evaluated for their reliability and relevance for predicting skin irritation hazard classification according to the European Union (EU) classification system. Depending on the skin model, accuracy ranged from 77% to 84% based on the EU classification system. Recent analyses by ECVAM of data from the SIVS indicate that sensitivity for EPISKIN™ is 94% (false negative rate = 6%) and specificity is 68% (false positive rate = 32%) when using the criteria for a GHS Category 2 skin irritant (i.e., average erythema or edema score > 2.3) as the threshold for a positive response instead of the EU R38 criteria (average erythema or edema score > 2.0). The false positive rate for EPISKIN™ in this analysis (32%) should be described and discussed in the TG. Use as a stand-alone assay for GHS classification could result in significant overlabelling of the actual skin irritation hazard due to the 32% false positive rate. Therefore, In order to avoid over-labeling that could result from such a high false positive rate, users should be encouraged to apply a weight-of-evidence approach to the interpretation of positive results in this assay that will aid in reducing the false positive rate. The accuracy of these human skin cell models for identifying GHS Category 3 (mild irritants) (i.e., average erythema or edema score between 1.5 and 2.3) is reduced to 53% to 58% when the 3-tiered Globally Harmonised System (GHS) classification system is used

Comment Number	Test Guideline Paragraph (¶)	Comments
		(under-prediction rates for the irritant category ranged from 54% to 57%; over- and under-prediction of the mild irritant category ranged from 62% to 65% and 8% to 15%, respectively; while over-prediction rates for the not labeled category ranged from 27% to 35%). The accuracy of EpiDerm™ was not considered adequate for the identification of skin irritants based on the original analysis according to the EU classification system. Therefore, it is unlikely to be considered adequate when compared to GHS Category 2 irritants either. For these reasons, there should either be no reference to EpiDerm™ in the TG, or the TG should not be finalized until the ongoing optimization/validation studies with EpiDerm™ are completed and determined appropriate for inclusion.
2	General comment	The TG as currently written reads more like a Guidance Document than a TG. From a practical perspective, it is unclear if a user could successfully conduct a test using this TG given the general nature of the procedural details regarding exposure duration, cytotoxicity decision criteria, and optional use of IL-1α.
3	General comment	The individual animal data for a subset of the 58 chemicals (57% [33/58]) included in the SIVS were not made publicly available. This prevented detailed analyses of these test methods using other national hazard classification systems to fully characterize their usefulness and limitations for specific regulatory requirements. This is an important step in the determination of test method acceptability by each member country. Consistent with OECD GD 34, all data used to validate an alternative test method should be made public.
4	2	OECD TG 430 (In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test), 431 (In Vitro Skin Corrosion: Human Skin Model Test), and 435 (In Vitro Membrane Barrier Test Method for Skin Corrosion) should be cited with the reference to *in vitro* alternatives to corrosivity testing.
5	3	The structural limitations of current EpiDerm™ and EPISKIN™ should be acknowledged. The human skin cell models are architecturally similar to human skin in many aspects and are greatly improved relative to monolayer cultures. However, these models are incomplete, since several cell types and organ systems are lacking, particularly immune tissues for assessment of the contribution of the inflammatory response. .

Comment Number	Test Guideline Paragraph (¶)	Comments
6	5	The phrase *reliably discriminate* is inappropriate in the context of an OECD Test Guideline (TG). Therefore, this sentence should be deleted. In order to adequately qualify the usefulness and limitations of these test methods, this paragraph should instead include detailed information that fully characterizes the extent to which these methods can detect irritants and non-irritants (e.g., accuracy statistics including false positive and false negative rates).
7	5 and 6	Paragraphs 5 and 6 should be combined into one paragraph to read: *Prevalidation and validation studies (2, 3, 4, 5, 6, 7, 8, 9, 10) have reported that in vitro tests employing human skin cell models are able to discriminate with an accuracy of 77% to 84% between known skin irritants and non-irritants according to the EU classification system; R38, no label (11). It does not provide adequate information on skin corrosion, nor does it allow the subcategorization of irritating substances as defined in the 3-tiered Globally Harmonized Classification System (GHS).*
8	6	It is not clear what is meant by *substances of high purity*. Furthermore, it is not clear why the applicability of the test methods has been restricted to such substances. Once clarified, this sentence should be moved to another section of the document (e.g., paragraph 18).
9	7	As noted in **comment 1**, the false positive rate for EPISKIN™ is 32% when using the GHS Category 2 as the threshold for a positive response. In order to avoid over-labeling that could result from using a method with such a high false positive rate, users should be encouraged to apply a weight-of-evidence approach to the interpretation of positive results in this assay. For this reason, the following text is recommended for inclusion in this paragraph: "Confirmatory testing *in vivo* or in appropriately validated *in vitro* tests should be considered for use if false positive results are suggested based on a weight-of-evidence evaluation of supplemental information (e.g., structure-activity relationships, other testing data, including animal tests).

Comment Number	Test Guideline Paragraph (¶)	Comments
10	7	No chemicals known to be false negative in an *in vitro* corrosivity assay were evaluated in the ECVAM SIVS. An estimated 12% to 21% of dermal corrosives are not identified as corrosives by at least one of the four *in vitro* corrosivity tests currently adopted for use (i.e., EPISKIN™, EpiDerm™, rat skin TER, and Corrositex™). In the United States, these corrosivity test methods are currently recommended for use in a weight-of-evidence tiered testing strategy in which substances testing negative for corrosivity would be tested *in vivo* for dermal irritation in up to three rabbits using a sequential testing approach. In this strategy, corrosive substances incorrectly identified as false negatives in an *in vitro* corrosivity test would be identified as corrosives during the *in vivo* dermal irritation test. As stated in OECD TG 431 (In Vitro Skin Corrosion: Human Skin Model Test), *the test described in this Guideline allows the identification of corrosive chemical substances and mixtures. It further enables the identification of non-corrosive substances and mixtures when supported by a weight of evidence determination using other existing information (e.g., pH, structure-activity relationships, human and/or animal data). It does not normally provide adequate information on skin irritation, nor does it allow the subcategorisation of corrosive substances as permitted in the Globally Harmonised Classification System (GHS).* If *in vitro* dermal corrosion and irritation methods are to be used as replacements for animals for dermal corrosion/irritation testing, then testing strategies must be capable of correctly identifying the false negative corrosive substances. ICCVAM has a study planned to evaluate how corrosive substances that produced false negative results in *in vitro* dermal corrosivity test methods act in *in vitro* dermal irritation test methods (e.g. EPISKIN™). **For this reason, the last sentence in paragraph 7 (line 45) should be deleted.**

Comment Number	Test Guideline Paragraph (¶)	Comments
11	10	As noted in the recent report of the Scientific Communities on Consumer Products (December 2007), the ECVAM SIVS did not provide *information with respect to colouration effects, potentially important in the case of colorants and hair dyes*. Therefore, *the SCCP is of the opinion that additional data are necessary to fully support the EPISKIN™ method for the safety assessment of cosmetic ingredients*. This important limitation should be mentioned in the context of the types of substances for which the human skin models are applicable. Also, the current paragraph should be concluded with, *which were not included in the validation study*. In addition, the complex multi-component structures of most of the 58 substances included in the ECVAM SIVS precludes any conclusions on the applicability of these test methods for specific chemical classes.
12	11	The paragraphs that are intended to describe the test method performance standards should be specifically identified. Also, since EpiDerm™was not considered adequately valid for the identification of skin irritants according to the EU classification system based on results from the SIVS, reference to EpiDerm™ as an example is not appropriate in the TG and should therefore be deleted unless there is other data available to substantiate its performance.
13	15	It is not clear what is meant by *on-going histological examination*.
14	16	It appears that this paragraph is intended as a performance standards requirement. If this is true, there should be criteria specified to indicate what would be considered adequate accuracy and reliability, including the required number of tests and laboratories. It also is not clear what is meant by the phrase *extended time period*.
15	17	EPISKIN™ and EpiDerm™ are both trademarked models and should be indicated as such throughout the TG. As noted in **comment 12**, EpiDerm™ should not be used as an example given that it was not considered adequately valid for the identification of skin irritants according to the EU classification system based on results from the ECVAM validation study. Therefore, there should either be no reference to EpiDerm™ in the TG, or the TG should not be finalized until the ongoing optimization/validation studies with EpiDerm™ are completed and determined appropriate for inclusion.
16	18	Since the human skin model system performance was originally validated using MTT, it is not clear why the guideline provides for a choice of dye indicators. If viability assays other than MTT are used, it may be necessary to optimize durations of exposure and incubation.

Comment Number	Test Guideline Paragraph (¶)	Comments
17	19	5% SLS should be identified as an example of a positive control that is a known skin irritant that is not severely irritating.
18	20	The 42-hour incubation post-exposure may not be optimal for viability assays other than MTT. For this reason, the Test Guideline should be restricted to using only the MTT assay for cell viability measurements until other viability assays are studied and appropriate post-exposure incubation times determined. Validation of other viability assays may necessitate optimization studies of incubation periods to achieve similar performance.
19	22	It is not clear if the statement that the MTT assay *has been shown to give accurate and reproducible results* is a general statement for all uses of the MTT assay, or if it is specific to the skin model systems.
20	23	The procedure for the MTT correction step should be included in the TG.
21	28	IL-1 α measurements did not prove useful for EpiDerm™ in the ECVAM SIVS. The usefulness of this endpoint for only one of the two methods included in the SIVS should be indicated in this paragraph.
22	30 and 31	Because the IL-1α endpoint has not been adequately standardized and validated for use in these test methods, specific decision criteria for use in hazard classification decisions should not be included in this TG. Therefore, these paragraphs should be deleted. However, based on the reduced false negative rate resulting from the inclusion IL-1α measurements (for EPISKIN™ only; it did not improve the performance of EpiDerm™) during investigations included in the validation study, ICCVAM considers further evaluation of this endpoint to be an important activity.
23	33	It is unclear what is meant by the phrase: *water solubility relevant to the conduct of the study*? Also, there is no mention of test method accuracy.
24	Table 2	Inaccurately cited as Table 1 in the accompanying text. There is no reference to which of these chemicals are to be tested for test method reproducibility and if they are intended for an accuracy evaluation (for which there is no mention of in the TG).

Comment Number	Test Guideline Paragraph (¶)	Comments
25	Table 2	The list of reference chemicals does not include physical (e.g., solids, liquids) and chemical (e.g., chemical class, pH, water solubility, molecular weight) properties for each chemical. The list also does not include individual animal data and mean edema and erythema scores for each chemical. In order to demonstrate that these chemicals are sufficiently representative of the anticipated range of responses and sufficiently diverse, this information should be added to Table 2. Additionally, the GHS hazard classification category for each reference chemical should also be added, based on the in vivo data. .