SCIENTIFIC ISSUE PAPER ON STRONG VS. WEAK SENSITIZERS

This paper briefly refers to respiratory sensitization, but the bulk of the paper addresses skin sensitization.

Introduction

1. Currently, most countries/sectors regulate sensitization hazards, without additional differentiation into strong or weak sensitizers. However, in the U.S., consumer products falling under the Federal Hazardous Substances Act (FHS) are regulated only if they are deemed strong sensitizers

2. The GHS calls for classification of substances as respiratory or dermal sensitizers without differentiation regarding sensitization strength. Dermal sensitizers may be classified using animal or human data. However, there is no animal model for regulatory use to identify respiratory sensitizers. Classification for that endpoint is based primarily on human data. Positive results in animal studies are considered to provide additional indications of respiratory sensitization potential. In many cases, weight of evidence reasoning using expert judgment must be used to classify dermal or respiratory sensitizers (GHS Par. 1.3.2.4.8). "For classification purposes, reliable epidemiological data and experience on the effects of chemicals on humans (e.g. occupational data, data from accident data bases) should be taken into account in the evaluation of human health hazards of a chemical. Testing on humans solely for hazard identification purposes is generally not acceptable." (GHS Par. 1.3.2.4.7).

3. When harmonizing existing systems for hazard classification, the OECD Task Force for Classification and Labeling treated sensitization without additional differentiation, noting that there was no internationally accepted animal test at the time which could be used to determine sensitization strength (GHS Chapter 3.4). However, as documented in the OECD *Harmonized Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures* in the appendix to chapter 2.4, the terms strong/weak include consideration of severity of allergic manifestations in humans or animals, as well as frequency in exposed populations. The following text was provided as "Background Information":

> "118. Categorization of sensitizers accounting for differences in sensitizing capacity among substances would be a useful concept to develop. It may be appropriate to allocate both respiratory and dermal sensitizers to, for example, one of the following categories:
>
> Category 1, Strong Sensitizer:
>
> A strong sensitizer would be indicated by:
>
> - a high frequency of occurrence and/or severity of occurrence within an exposed population; or
>
> - a probability of occurrence of a high sensitization rate in humans based on animal or other tests.
>
> Category 2, Sensitizer:

A low to moderate sensitizer would be indicated by:

- a low or moderate frequency or severity of occurrence within an exposed population; or

- a probability of occurrence of a low to moderate sensitization rate in humans based on animal or other tests.

119. Some authorities currently categorize strong sensitizers. However, at present, animal or other test systems to subcategorize sensitizers as indicated above, have not been validated and accepted. Work is going on to develop such models for the potency evaluation of contact allergens."

4. In 2002, the IOMC Coordinating Group for the Harmonization of Classification and Labelling noted that the sensitization criteria for substances should be re-opened to consider the inclusion of new information and evolving testing approaches that address the question of distinguishing strong sensitizers from those that are weaker. Appropriate hazard communication should be considered along with the discussions on the criteria and the availability of an appropriate test method. The UN mandate (ST/SG/AC.10/C.4/2002/19, December 2002, UN Sub-Committee HCL) directed OECD to consider use of "strong vs. weak" sensitizers in the GHS. This mandate was extended for the biennium 2005 - 2006 (ST.SG/AC.10/C.4/16). The mandate is stated as follows:

"Sensitization - Strong versus weak

Objective: To examine the available information concerning strong vs. weak sensitizers and, if appropriate, propose revisions to the classification criteria for respiratory and/or dermal sensitization."

Background Information on Approaches for Determining Sensitization Strength

German Panel of Experts

5. The approach articulated in the paper by Schlede et al summarized work performed in the course of 15 years by a panel of German experts. In this publication, more than 200 contact sensitizers were ranked. (Schlede et al. Chemical substances and contact allergy - 244 substances ranked according allergenic potency. (2003), Toxicology 193, 219-259). Schlede et al include prevalence, strength of sensitization in animals and humans and severity of response and cross-reactivity to rank sensitizers. Weight-of-evidence determinations used human clinical data and patch test results as well as animal data when available.

1. Significant allergen: (1) proven strong allergenic effect in humans after short and/or almost negligible exposure taking into account existing animal data; (2) frequently proven contact allergenic effect in humans. Remarks: data on humans demonstrate that in larger collectives 1% or more of the patients react positive and that several independent case studies and experimental data on humans are available.

2. Solid-based indication for contact allergenic effects: (1) less frequently proven contact allergenic effect in humans taking into account existing positive animal data; (2) the capacity of substances to induce cross-reactions in humans without being a significant allergen itself. Remarks: data on humans demonstrate that in collectives less than 1% of the patients react positively and that independent case studies and/or experimental data on humans are available.

3. Insignificant contact allergen or questionable contact allergenic effect because of: (1) rarely proven contact allergenic effect in humans; (2) doubtful effect in humans; no or non-appropriate animal data; (3) no data on humans but positive animal data. Remarks: data on humans include isolated positive test results and isolated case studies and experimental data.

EU Expert Group on Sensitization

6. For induction of skin sensitization, the EU Expert Group proposed a 3-level potency scheme based primarily on potency scores from any of the Local Lymph Node Assay (LLNA), Guinea Pig Maximization Test (GPMT), and Buehler tests. According to the report of the EU Expert Group (4-6 November 2002):

> "The design of the LLNA makes it better suited than the guideline guinea pig assays to the assignment of skin sensitizers into specific potency categories. This is because the LLNA focuses on induction of sensitization only, incorporates a dose response assessment, and has an objective and quantitative endpoint."
>
> "The majority of skin sensitizing chemicals would then fall into the category corresponding to the current default value of 1% for labelling of preparation with R43. An additional 2 categories should be defined for substances with higher potency; these identify strong (>0.1%) and extreme (>0.001%) sensitizers, respectively. With regard to preparations, moderate and strong skin sensitizers would be listed on the label when present in a concentration of 10 ppm or greater, and extreme skin sensitizers when in a concentration of 1 ppm or greater."
>
> "Elicitation thresholds correlate only poorly with induction potency. Variation in elicitation thresholds between individuals is very large and depends on numerous factors of which the sensitizing potency of the substance is only one. Other factors affecting elicitation include the duration, extent and site of exposure, status of the skin and degree of specific sensitization. For this reason, the Expert Group considered that it would be inappropriate to define elicitation thresholds as a function of skin sensitizing potency.
>
> Human data should normally only be used to re-categorize a substance into a higher potency category. The EU Expert Group proposal does not consider questions of severity of response or cross-reactivity." (Report from the Expert Group on Sensitization, 18-19 April 2002 and 4-6 November 2002).

European Center for Ecotoxicology and Toxicology of Chemicals (ECETOC)

7. Report # 87 *Contact Sensitization: Classification according to Potency* includes the following definition of potency:

> "Potency in the context of allergic contact dermatitis is best defined as the amount of chemical required for the acquisition of skin sensitization in a previously naive individual (induction phase), or the amount of chemical necessary to elicit a clinically discernable cutaneous reaction in previously sensitized subjects."

U.S.: Consumer Products

8. In regulating consumer products that are strong sensitizers according to FHSA, the US uses severity of response, frequency of responses in exposed populations, and dose at which allergic reactions occur. In addition, Canada has expressed interest in such considerations.

9. The statutory definition in use by the Consumer Products Safety Commission (CPSC) for consumer products in the US combines these elements to determine strength of sensitization. The CPSC definition is as follows:

> A strong sensitizer means a substance which will cause on normal living tissue through an allergic or photodynamic process a hypersensitivity which becomes evident on reapplication of the same substance and which is designated as such by the Commission. Before designating any substance as a strong sensitizer, the Commission, upon consideration of the frequency of occurrence and severity of the reaction, shall find that the substance has significant potential for causing hypersensitivity.

10. Supplementary guidance issued by a Technical Advisory Panel on Allergic Sensitization follows:

> "a sensitizer is a substance that will induce an immunologically-mediated (allergic) response, including allergic photosensitivity. This allergic reaction will become evident upon re-exposure to the same substance. Occasionally, a sensitizer will induce and elicit an allergic response on first exposure by virtue of active sensitization".

11. In addition, the panel recommended that when determining that a substance is a "strong" sensitizer and a substance's sensitizing potential, available data is to be considered (i.e.; frequency of occurrence, severity of reactions in healthy or susceptible populations, human and animal experimental data with human taking precedence, bioavailability of sensitizers, data on cross-reacting substances, human threshold sensitivity). The severity of reaction was qualified as a "clinically important reaction", one producing substantial illness (i.e.; physical discomfort, distress, hardship, functional or structural impairment).

12. CPSC has assembled an expert panel, reflecting academia, industry and government regulators from Europe and North America to address the issue of need and criteria for identifying strong sensitizers for their statutorily mandated regulation. This panel is taking the GHS definition of sensitizers into account as part of its deliberations.

Testing for Sensitization

13. Human and animal testing and evaluation methods are described in Annex 1.

Animal data:

14. Traditional test methods used for regulation of sensitizers have focused on determining whether or not a substance is a sensitizer. In the Guinea Pig test methods, the determination is based on results in excess of a pre-determined percent of animals eliciting a response after repeated applications of the substance. In the LLNA test in mice, determination that a substance is a sensitizer is based on results exceeding a pre-determined ratio of effect in test animals versus controls.

15. Overall, the accuracy of the guinea pig tests as methods of predicting human sensitization is considered 88%. According to the ICCVAM LLNA peer review report [ref.: NIH(1999), NIH Publication No. 99-4494], the LLNA performed at least as well as currently accepted guinea pig methods (GPMT/BA) for the hazard identification of strong to moderate chemical sensitizing agents. The performance of the LLNA and the GPMT/BA was similar when each was compared to human data (HMT/HPTA). The accuracy of the LLNA vs. human data was 72% (N=74), GPMT/BA was 72% (N=57), and all guinea pig tests (GPT) vs. human was 73% (N=62).

16. In traditional guinea pig tests submitted for regulatory review, the dose at which responses occur is not generally recorded, although in some cases, such information is available. The guinea pig test are not designed for looking at potency; however, by a modified protocol using multiple induction doses, described by Andersen, potency can be assessed.

17. With the recent development of the mouse Local Lymph Node Assay, the dose at which the EC3 (discriminating level) is exceeded is normally available.

18. Many animal tests and human data/information actually observe elicitation, which is an indicator of both the induction and elicitation phases of sensitization. Some experts consider however, that animal tests are performed primarily to identify induction of sensitivity.

Human data:

19. Epidemiological evidence judges prevalence of effects based on frequency of response in humans, consideration of severity of response and its significance for regulatory purposes. (Prevalence in the general population is a reflection of intrinsic potency and the degree of exposure.) Human testing for epidemiological or diagnostic purposes normally measure elicitation responses in subjects who have been previously exposed (annex 1).

Issues to be addressed by the OECD

20. GHS paragraph 1.3.2.4.9.3 says that "generally data of good quality and reliability in humans takes precedence over other data." And the GHS paragraph 3.4.2.2.2.2 states "Positive effects seen in either humans or animals will normally justify classification. Evidence from animal studies is usually much more reliable than evidence from human exposure. However, in cases where evidence is available from both sources, and there is conflict between the results, the quality and reliability of the evidence from both sources must be assessed in order to resolve the question of classification on a case-by-case basis." Respiratory sensitization is classified primarily based on human data because there is no standardized animal model for this end point. Therefore, the scheme proposed by the EU Expert Group for induction of skin sensitization may not be applicable for respiratory sensitization.

21. Potentiation of skin sensitizers in mixtures can occur due to other ingredients which might enhance absorption or "toxicity" of the sensitizer. Solvents can have up to 20 fold influence on the measured potency of an allergen. Product matrices can also affect responses due to availability of the sensitizing ingredient. What are vehicle effects on dose-response? How would classification of untested mixtures with 3 or 4 potency levels address this? Are these factors playing roles different from their roles in the determination if the substance/mixture is a sensitizer or not a sensitizer?

22. Many animal tests and human data/information actually observe elicitation, which is an indicator of both the induction and elicitation phases of sensitization. Elicitation generally occurs at lower doses than induction and can occur at lower doses as exposure is repeated. The variable nature of such data would lead to special challenges for comparison among chemicals of sensitization potency if it were decided to link potency to induction and elicitation rather than to induction only. The GHS includes both induction and elicitation in defining a sensitizer: should potency be connected to induction only?

23. Given that prevalence is not necessarily an end by itself, but is a measurement used when evaluating human data, how can intrinsic potency be teased out from exposure when both factors contribute to prevalence? What considerations are needed regarding use of prevalence data in humans as part of the weight of evidence to discriminate between strong and weak sensitizers? How significant is consideration of the type of exposed population, i.e. general population, sensitive population, occurrence of

atopic individuals? Can intrinsic strength of sensitizers be adequately distinguished when there is likely to be a range of exposures?

24. The severity of allergic reactions varies. Allergic contact dermatitis can range from mild local reactions to erythroderma, which affects most of the body surface. Immediate hypersensitivity reactions can range from mild rhinitis to local hives (contact urticaria) to severe asthma and anaphylactic shock. Can such responses be related to strong vs. weak sensitizers? Can such manifestation be predicted from responses in animals? How can manifested responses in humans be distinguished from dose to which they are exposed? When can a severe response be determined to be caused by intrinsic properties of a chemical or high exposure or individual susceptibility? Can interspecies extrapolation be used to relate animal responses to severe responses in humans?

25. For some animal data, the sensitization response rate can be measured. How can such measurements be related to response rate in humans?

26. Can animal tests be correlated with probable human sensitization responses in order to distinguish strong from weak sensitizers?

Discussion

27.. The OECD Expert Group is exploring development of a scientifically defensible way to define strong versus weak sensitizers with sufficient clarity for classification purposes.

28. Harmonization must take into consideration generally hazard based existing systems. In the U.S., consumer products falling under the FHSA are regulated only if they are deemed strong sensitizers (taking into consideration frequency of sensitization in an exposed population, severity of response and dose at which the sensitization occurs).

29. The GHS defines sensitization hazards as including both induction and elicitation and advocates the use of animal and human data as and when available. "Generally, data of good quality in humans will have precedence over other data. However, even well designed and conducted epidemiological studies may lack sufficient numbers of subjects to detect relatively rare but still significant effects, or to assess potentially confounding effects. Positive results from well-conducted animal studies are not necessarily negated by the lack of positive human experience but require an assessment of the robustness and quality of both the human and animal data relative to the expected frequency of occurrence of effects and the impact of potentially confounding factors." (GHS Par. 1.3.2.4.9.3).

30. The approach under consideration in the EU is to define strong versus weak sensitizers based on the intrinsic capacity of the chemical to induce sensitization in animals. The current regulatory approach in the US is to consider, in defining strong versus weak sensitizers, data in animals and humans, with the latter based on the intrinsic capacity of the chemical to induce sensitivity, to elicit responses in sensitized individuals, and on the actual exposure encountered in the population.

31. An approach like that of the German panel (Schlede et al) merges human and animal data and takes into consideration severity, dose, and frequency of response.

32. At the second OECD Expert Group meeting (5-6 May 2004), "The Expert Group agreed that the LLNA is very important for ranking (contact) sensitizers but that additional factors (human data and animal data) also have to be taken into account".

33. The approach under consideration in Europe leads to categories based on potency for induction of sensitization. If dose leading to sensitization is to be used as a single parameter, in order to reflect the US

system in use currently, it must be shown to correlate with human response in a way which takes into account severity and frequency of response as well. For such a correlation to be validated, it is essential to have a reference list of chemicals characterized for sensitization in humans with consensus values for potency and other relevant parameters including frequency and severity of response. In addition, since many test or observational methods actually measure elicitation responses, agreement is needed on a consistent way to assess elicitation thresholds (noting that elicitation responses often are expressed at lower concentrations with repeated exposures).

34. The EU Expert Group advocates ranking chemicals primarily by means of LLNA. This is because the LLNA focuses on induction of sensitization only, incorporates a dose response assessment, and has an objective and quantitative endpoint. However, GPMT and Buehler evaluate both phases of sensitization – induction and elicitation, as included in the GHS, while LLNA evaluates primarily induction. The response rates of GPMT and Buehler have been used by some authorities for potency determination.

35. Classification should be able to be used for both existing and new chemicals. The approach based on animal test model (in particular the LLNA) is predictive (proactive) attempting to identify and rank the effect before human use and does not rely on human exposure information. When sensitizers are ranked based on human data, the approach is retroactive. This approach relies on human exposure information considering magnitude and frequency of exposure and severity of response among exposed individuals during the actual use of the marketed (commercially) available product(s). This is not an approach that is applicable for new substances before they are available on the market. It is necessary to decide on the advantages /disadvantages of these two approaches in relation to the requirements of the GHS which is based on hazard classification. Animal test data may be used to identify a sensitizer before human use. When human data becomes available it should be considered in evaluating the hazard.

36. Harmonized Categories proposed for classification must be able to be strongly differentiated such that substances can be classified consistently in the various UN nations.

37. The comprehensive examination of the current science by an international panel being undertaken for consumer products in the U.S. is expected to provide new insights into the question of strong vs. weak sensitizers.

ANNEX 1

Sensitization Evaluation Methods

Guinea Pig Maximization Test: Typically the highest concentration of a chemical causing mild to moderate irritation is multiply injected intradermally (with and with out adjuvant) on a shaven shoulder, 7 days later a patch containing the same highest to moderate irritating concentration of chemical is applied for 48hrs as a booster. At 14 days post induction, challenge (with a maximal non-irritating dose of the chemical) is carried out on the flank with occlusion for 24hr. The area of erythema and edema is evaluated at 24 and 48 hrs post challenge. A chemical is classified as a sensitizer if at least 30% of the animals have a positive response (grade 1 or higher). Past concerns with the GPMT regarded occurrences of false positive responses and interpretation of weak responses. Dose-responses are typically utilized with the challenge dose. The GPMT has not been formally validated for potency determination.

Buehler Test: The chemical is applied to a shaven flank at a minimal irritating dose and occluded for 6hrs. The procedure is repeated on day 7 and day 14. Chemical challenge is carried out 2 weeks later on the opposite shaven flank and occluded for 24hrs (though some may occlude for 6hrs) taking the highest non-irritating. Upon removal of the patch, the area is evaluated at 24hr and 48hrs for edema and erythema. A chemical is classified as a sensitizer if 15% of the animals demonstrate a positive response (grade 1 or higher). This protocol is considered less sensitive than the GPMT but is less prone to false positive results. The major differences from the GPMT lie in the induction phase, with the lack of utilization of both adjuvant and intradermal application. The Buehler test has not been formally validated for potency determination.

Local Lymph Node Assay: A 3 day repeated application of the chemical is applied to the ear dorsum. On day 5, tritiated-thymidium is injected (i.v.) and 5hrs later lymph nodes are excised and counted. A chemical is classified as a skin sensitizer if it induces at least a 3-fold increase in proliferative counts compared to vehicle-treated controls (the stimulation index – SI). The concentration of chemical which produces at least a SI of 3 is the EC3 value. A concern regarding the LLNA is that it is more appropriate for the class of chemicals considered Type IV sensitizers, which act through T-cell mediated mechanisms, and less so for Type I sensitizers whose mechanistic effects are antibody mediated. The LLNA has been validated for dermal sensitization hazard identification; the LLNA has not been formally validated for potency determination.

Epidemiological Data: Much of the population data is derived from diagnostic patch testing in dermatitis patients. However, this provides common exposure patterns such that a typical list of 25-35 compounds is maintained for universal patch test studies (standard series). Data from samples of the general population and exposed groups are also available, however more limited. Thus, the epidemiological data generated may not be representative of the general population. Nevertheless, the data permit recognition of allergens with high sensitizing potential.

Human Testing: Diagnostic patch testing is the procedure used for detection of contact allergy (skin sensitization) to substances in humans. Patch testing is performed in individuals with dermatitis, and in experimental and epidemiological studies. The test procedure is standardized, while different patch test systems are in use. The patches are applied on the back for 2 days and grading of reactions is recommended to be done 2, 3/4 and 5/7 days after application. A typical patch test system is aluminium

chambers (Finn chamber®) on adhesive tape with test substances at set concentrations (the standard series, other series etc.), generally in petrolatum. TRUE® test is comprised of chemical gel matrix patches put on with adhesive tape. Patch test dose-response studies have been carried out, which thereby demonstrate elicitation thresholds under certain circumstances (e.g.; the fragrance allergen isoeugenol).

Additional testing procedures may also be used, but they should not be used to replace patch testing. These include the open test, the semi-open test and use tests. Use tests with the products (e.g. the provocative use test, PUT) were originally intended to mimic the actual use situation without the goal to differentiate between allergic and irritant skin reactions. Nowadays they are most commonly used to evaluate the clinical relevance of a patch test reaction. The repeated open application test (ROAT) is a standardized method of use testing. The test substance, either a commercial product, as is, or a special test substance is applied twice daily for one week or longer. The value of ROAT has been verified in cases with positive, negative or questionable reactions at initial patch testing and in animal studies.

In addition to patch tests for contact allergy, diagnostic analysis for respiratory sensitization includes tests such as the Skin Prick test, intradermal tests, and serological immunological tests for the presence of specific antibodies (e.g; RAST test). Less commonly, challenge testing via oral, inhaled or other routes.

While human diagnostic tests only test for whether an individual has been pre-sensitized by prior exposure or to determine an elicitation threshold in a sensitized individual, the Human Repeat Insult Patch Test (HRIPT) is a test of allergenic potential and more comparable in function to the guinea pig test or the LLNA. It involves a chemical patch applied (to the same site) three times a week, occluded for 24 hours, for a three week period. Two weeks later, the same site is challenged with the chemical and responses noted. The amount of consecutive induction patches can vary. The HRIPT provides an exaggeration of product use and testing higher than use concentrations (typically a mild irritating dose).

Human predictive sensitization tests in volunteers is, in Europe, not considered ethical to perform due to the risk that patch test sensitization may elicit clinical disease in the subject. In addition, GHS Paragraph 1.3.2.4.7 says: "Testing on humans solely for hazard identification purposes is generally not acceptable."