Regulatory Toxicology and Pharmacology 88 (2017) 34–44

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Probabilistic derivation of the interspecies assessment factor for skin sensitization

W. Bil [a], A.G. Schuur [b], J. Ezendam [c], B.G.H. Bokkers [b, *]

[a] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
[b] National Institute for Public Health and the Environment, Centre for Safety of Substances and Products, Bilthoven, The Netherlands
[c] National Institute for Public Health and the Environment, Centre for Health Protection, Bilthoven, The Netherlands

ARTICLE INFO

Article history:
Received 18 November 2016
Received in revised form
9 May 2017
Accepted 18 May 2017
Available online 22 May 2017

Keywords:
Allergic contact dermatitis
Dose-response modeling
HMT
HRIPT
Interspecies sensitization assessment factor
LLNA
Probability distribution
Quantitative risk assessment
Skin sensitization

ABSTRACT

An interspecies sensitization assessment factor (SAF) is used in the quantitative risk assessment (QRA) for skin sensitization when a murine-based NESIL (No Expected Sensitization Induction Level) is taken as point of departure. Several studies showed that, on average, the murine sensitization threshold is in good correspondence with that determined in humans. However, on an individual level, the murine and human sensitization thresholds may differ considerably. In this study, the interspecies SAF was quantified by using a probabilistic approach, to be able to take these cases into account. As expected, the geometric means of the probability distributions of murine and human sensitization threshold ratios were close to one, but taking the 95 th percentile of these distributions resulted in an interspecies SAF of 15. By using this value, one is sure that with 95% probability, the sensitization threshold determined in mice does not underestimate the human threshold. It can be concluded that a murine-based NESIL requires the use of an interspecies SAF (of 15) in the QRA for skin sensitization, to correct for the differences between mice and humans. This empirically derived interspecies SAF contributes to refinement of the risk assessment methodology.

1. Introduction

Allergic contact dermatitis (ACD) is a delayed type IV allergic response of the skin, caused by contact with low weight chemical compounds (WHO/IPCS, 2012). ACD consists of two phases: an initial induction phase (also referred to as the sensitization phase) in which the immune system is primed, and a subsequent elicitation phase, in which the clinical response is initiated. During the induction phase, no signs or symptoms of allergy are observed and hence this phase occurs unnoticed. Therefore, detection of sensitization among the general public is usually performed retrospectively, by epidemiological or clinical studies on the incidence of ACD (SCCS, 2015). One of the major causes of ACD in the general public is exposure to chemicals in consumer products, such as fragrances or preservatives in cosmetics and cleaning agents, or metals in jewelry (Wijnhoven et al., 2008). Ideally, an individual should be protected against induction of skin sensitization caused by chemical substances.

To achieve this protection, the fragrance industry (International Fragrance Organization/Research Institute for Fragrance Materials (IFRA/RIFM)) proposed in 2008 to develop a risk assessment for skin sensitizing (fragrance) substances (Api et al., 2008). In this quantitative risk assessment (QRA) methodology, an Acceptable Exposure Level (AEL) is estimated, which aims to prevent individuals from becoming sensitized to allergenic fragrances. For derivation of the AEL, a No Expected Sensitization Induction Level (NESIL) is modified by using various Sensitization Assessment Factors (SAFs). Generally, the NESIL is obtained from an Estimated Concentration that induces 3-fold lymphocyte proliferation (EC3) determined in the Local Lymph Node Assay (LLNA), or the No Observed Effect Level (NOEL) obtained in a human sensitization test (e.g. the Human Repeated Insult Patch Test (HRIPT) or Human Maximization Test (HMT)). SAFs are applied to extrapolate the NESIL obtained in an experimental setting to the level that is safe in the real-life situation. In a recent publication of Basketter and

* Corresponding author. National Institute for Public Health and the Environment (RIVM), Centre for Safety of Substances and Products (VSP), PO Box 1, 3720 BA Bilthoven, The Netherlands.
E-mail addresses: w.bil@students.uu.nl (W. Bil), gerlienke.schuur@rivm.nl (A.G. Schuur), janine.ezendam@rivm.nl (J. Ezendam), bas.bokkers@rivm.nl (B.G.H. Bokkers).

http://dx.doi.org/10.1016/j.yrtph.2017.05.015
0273-2300/

Safford (2016), various SAFs were proposed for this extrapolation, i.e. a SAF for inter-individual variability, for matrix differences, for differences in frequency/duration, for variability in occlusion conditions, and for differences in skin condition and skin site. When the AEL is lower than the Consumer Exposure Level (CEL) determined in the exposure assessment, there is a potential risk of skin sensitization in consumers.

Experience with this QRA is mainly available within industry. This quantitative approach is not yet used by regulatory authorities, e.g. for the restriction of chemical compounds, or in the risk assessment of cosmetic ingredients. One of the reasons is that the Scientific Committee for Consumer Safety (SCCS, formerly SCCP) considered that the approach needed further refinement and recommended obtaining scientific consensus on the approach, especially on the selection of SAFs (SCCP, 2008). In this article, we aim to further elaborate on the SAFs by describing a probabilistic approach to derive an assessment factor, and providing an empirically derived interspecies SAF based on available human and murine sensitization data.

The interspecies SAF is currently ignored in the QRA, as the NESIL is preferred to be derived from human (HRIPT) data, either using generated data or data from historical origin (Api et al., 2008). However, human sensitization assays are not performed in Europe anymore, as they are considered unethical (SCCS, 2015). Under European regulations for the safety of chemicals, such as Registration, Evaluation, and Authorization of Chemicals (REACH, Regulation [EC] No 1907/2006) and the Biocidal Products Regulation (BPR, Regulation [EC] No 528/2012), the LLNA is the preferred *in vivo* assay to determine whether a compound has a sensitizing capacity. Thus, the NESIL can only be based on the EC3 determined in the LLNA. It is therefore necessary to know whether the EC3 is a reliable predictor for the human sensitization induction level.

The interspecies SAF is based on the relationship between the EC3, and the Dose per Skin Area that shows a 5% incidence of sensitization (DSA_{05}) in the HRIPT or HMT assay (Schneider and Akkan, 2004). By comparing these two indicators of sensitizing potency, one is able to derive the factor needed to extrapolate the murine derived EC3 to the DSA_{05} obtained from the test population under HRIPT or HMT testing conditions. In several simple linear regression analyses LLNA data was compared to human data (Griem et al., 2003; Schneider and Akkan, 2004; Basketter et al., 2005; ICCVAM, 2011). These studies showed that the murine sensitization threshold is in good correspondence with that determined in humans and hence inclusion of an interspecies SAF in the QRA would be superfluous (i.e. interspecies SAF of 1).

It is questionable whether simple linear regression alone is the best approach to determine an interspecies SAF. Only when the slope of the regression line is one and the intercept is zero (i.e. the unity line), one can conclude that there is a direct, one-to-one correspondence between the EC3 and the DSA_{05}. The coefficient of determination (R^2) does provide information on the deviation of points (i.e. substances) from the regression line, but does not indicate whether the slope and y-intercept of the regression line deviate from unity. It is particularly important to take into account the possible cases for which the EC3 value, for whatever reason, deviates from the DSA_{05}. Regression analysis without any analysis of the residuals does not put enough emphasis on these cases. By using a probability distribution to derive a default deterministic assessment factor instead, one is able to take the deviations from unity into account (Kramer et al., 1996).

Whenever in future the QRA will be used for the assessment of new chemicals, the NESIL will most certainly be based on data other than that originating from a HRIPT or HMT. Although in future the performance of animal tests (LLNA) will be reduced due to recent changes in legislation (e.g. REACH, Cosmetics Regulation [EC] No 1233/2009), chemicals will still be assessed on their sensitization potency in Europe using historical LLNA data. It is therefore essential to investigate if there are potential differences between mice and humans in terms of the level at which sensitization is induced. In this study, the interspecies SAF is quantified using a probabilistic approach, to empirically establish the value needed for this SAF. The obtained factor will indicate whether the EC3 is a good predictor for human sensitization induction, or that extrapolation may be required to adjust an EC3 into a DSA_{05}.

2. Materials and methods

2.1. Data sources

2.1.1. LLNA data

Data of LLNA studies were retrieved using the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) LLNA potency database, version December 23, 2013 (NICEATM, 2013). The data used are considered sufficiently biologically valid, and have been used by various authorities (e.g. Joint Research Centre (Dumont et al., 2016),) to assess the general performance of the LLNA. The same applies to the HRIPT and HMT data used in this study.

The NICEATM database is comprised of both published and unpublished data of 669 different chemicals in 1060 LLNA studies in total. Data presented in this database were verified against the primary sources. The NICEATM database was taken as a starting point, but when discordant values were encountered, the values of the original source were considered more reliable and therefore were used. Additionally, LLNA data were retrieved from public databases. LLNA data were added for farnesol (Lapczynski et al., 2008), 4-phenylenediamine (Warbrick et al., 1999b), isoeugenol (Basketter and Cadby, 2004; Bertrand et al., 1997), 2,4-dinitrochlorobenzene (Loveless et al., 1996), glutaraldehyde (Basketter et al., 2003), hydroxycitronellal (Lalko et al., 2004), and formaldehyde (Basketter et al., 2001). A description of the LLNA protocol is provided in Fig. 1.

2.1.2. Human data

The main source for the human data was the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) LLNA Test Evaluation Report, appendix C annex II-2 (ICCVAM, 2011). This document contains 298 test results of both HMT and HRIPT studies, for 136 chemicals. Data were not validated against the primary sources but were directly taken from the ICCVAM database, because the main part of the original sources was not publicly available. Furthermore, in several of the accessible sources, crucial information such as patch size or participant number was lacking. It was assumed that taking values directly from the database increased transparency, because of the harmonized implementation of assumptions by ICCVAM (e.g. the same patch size in HMT studies). Moreover, additional data were retrieved from public databases, i.e. data for citral and farnesol (Lalko and Api, 2008; Lapczynski et al., 2008). A description of the HRIPT and HMT protocols is provided in Fig. 1.

2.2. Composition of the datasets

2.2.1. Dose-response dataset

The EC3 and DSA_{05} values presented in literature are generally calculated by means of linear inter- or extrapolation (ICCVAM, 2011). In principle, a straight line is drawn between two data points. This method leads to imprecise values, because a straight line generally does not describe the underlying dose-response curve. Therefore, in the current paper, the threshold values as