

Development of an Open-Source Integrated Test Strategy for Skin Sensitization Potency

J Pirone¹, M Smith¹, N Kleinstreuer², T Burns², J Strickland², Y Dancik³, R Morris¹, L Rinckel², W Casey⁴, J Jaworska³

¹SSS, Inc., Durham, NC, USA; ²ILS, RTP, NC, USA; ³P&G NV, Strombeek-Bever, Belgium;

⁴NICEATM/DNTP/NIEHS/NIH/HHS, RTP, NC, USA

Abstract

Regulatory authorities require testing to identify substances with the potential to cause allergic contact dermatitis so that appropriate labeling alerts users to the hazard and precautions necessary to minimize exposure. To reduce or eliminate animal use in testing, integrated test strategies (ITS) that combine *in silico* and *in vitro* test methods have been proposed. Scientists at the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and Procter and Gamble (P&G) are developing an open-source version of a previously published ITS for skin sensitization. The original ITS is based on a Bayesian network (BN ITS-2) using *in silico* and *in vitro* models that map to the OECD Adverse Outcome Pathway for skin sensitization. BN ITS-2 was developed using a commercial software package. To increase accessibility and algorithmic transparency, NICEATM and P&G developed open-source ITS-2 (OS ITS-2) with tools in R software for building and performing exact inference using a Bayesian network. R versions of widely used algorithms for supervised discretization and latent class learning were substituted for proprietary algorithms. The overall classification accuracies for the OS ITS-2 and the BN ITS-2 were the same, with three compounds misclassified by both methods. Two case studies of representative substances, chlorobenzene and 2-mercaptobenzothiazole, were evaluated using NICEATM's skin sensitization database, and value of information was assessed for the *in vitro* assays and *in silico* inputs. The OS ITS-2 provides availability and transparency, and represents a major step in allowing the ITS to be reproduced and tested, which is essential for use in a regulatory framework. The model is available on the NTP website (<http://ntp.niehs.nih.gov/go/its>).

Introduction

- The evaluation and promotion of alternative test methods for regulatory use in assessing skin sensitization hazards are a priority of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM).
 - The murine local lymph node assay (LLNA), the first alternative test method evaluated by ICCVAM, has been accepted internationally since 2002 for assessing skin sensitization hazard (OECD 2010).
 - Compared with guinea pig methods, the LLNA reduces the use of animals and eliminates the potential pain and distress associated with a positive response.
- To further reduce and potentially eliminate animal use for skin sensitization testing, potency results from the LLNA were used as the target endpoint to develop an integrated testing strategy (ITS) using a Bayesian network (BN) (Jaworska et al. 2011, 2013).
- The BN ITS:
 - Combines relevant *in silico* and *in vitro* data to make probabilistic predictions of skin sensitization potency category
 - Is aligned with the adverse outcome pathway (AOP) for substances that initiate the skin sensitization process by crossing the skin barrier and covalently binding to skin proteins (OECD 2012)

Methods

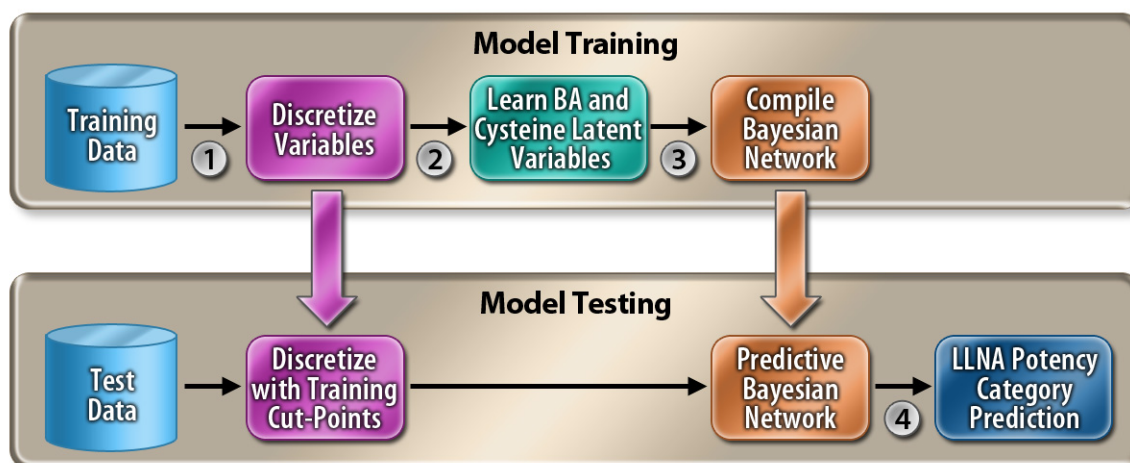
- The original and more recent versions of the BN ITS (Jaworska et al. 2011, 2013) used commercial software.
- We have developed an open-source (OS) version of the more recent BN ITS, ITS-2, using the free and open-source statistical programming language R (R v3.0.1, GNU Public License v3) (R Development Core Team 2008).
- Refinements to the published BN ITS-2 for skin sensitization (Jaworska et al. 2013) made in the OS ITS-2 include:
 - Correction of two errors in the experimental data
 - A change in the method for calculating the bioavailability parameters to improve transparency (to assure public access to all of the calculations) and consistency of predictions
 - The skin diffusion pathway for polar substances was eliminated from the calculation as it remains under development and is not yet publicly available. The bioavailability for the lipid diffusion pathway was calculated using a tool available on the National

Institute for Occupational Safety and Health website

(<http://www.cdc.gov/niosh/topics/skin/finiteSkinPermCalc.html>).

- The prediction strategy for physicochemical properties was revised to consider the following parameters:
 - LogP (i.e., calculated via EpiSuite or ACD/Labs v 12.0 predicted value)
 - Water solubility (S_w)
 - Vapor pressure (P_{vp})
 - Density, pKa value(s), Log D, MW (i.e., from ACD/Labs v 12.0)
 - EpiSuite calculated melting point
- The refined version of OS ITS-2 is referred to as OS ITS-2 lipid and is posted on the NTP website at <http://ntp.niehs.nih.gov/go/its>.
- The process for building and testing OS ITS-2 lipid is shown in **Figure 1**.

Figure 1. Primary Computational Steps for Building OS ITS-2 Lipid



Abbreviations: BA = bioavailability. Variables are defined in **Table 1** and the structure of the network is shown in **Figure 2**.

- The OS ITS-2 lipid model was trained to the target variable, LLNA potency category, with 124 substances: 36 nonsensitizers, 28 weak sensitizers, 35 moderate sensitizers, and 25 strong or extreme sensitizers.
 - LLNA potency categories are defined by EC3, or the effective concentration that produces a stimulation index of 3.
 - The potency categories are “strong or extreme” ($EC3 < 1\%$), “moderate” ($1\% \leq EC3 < 10\%$), “weak” ($EC3 \geq 10\%$), and “nonsensitizer” (no EC3).
- The *in vitro* and *in silico* data variables relevant to skin sensitization used to train the model are shown in **Table 1**. The structure of the OS ITS-2 lipid model is shown in **Figure 2**.

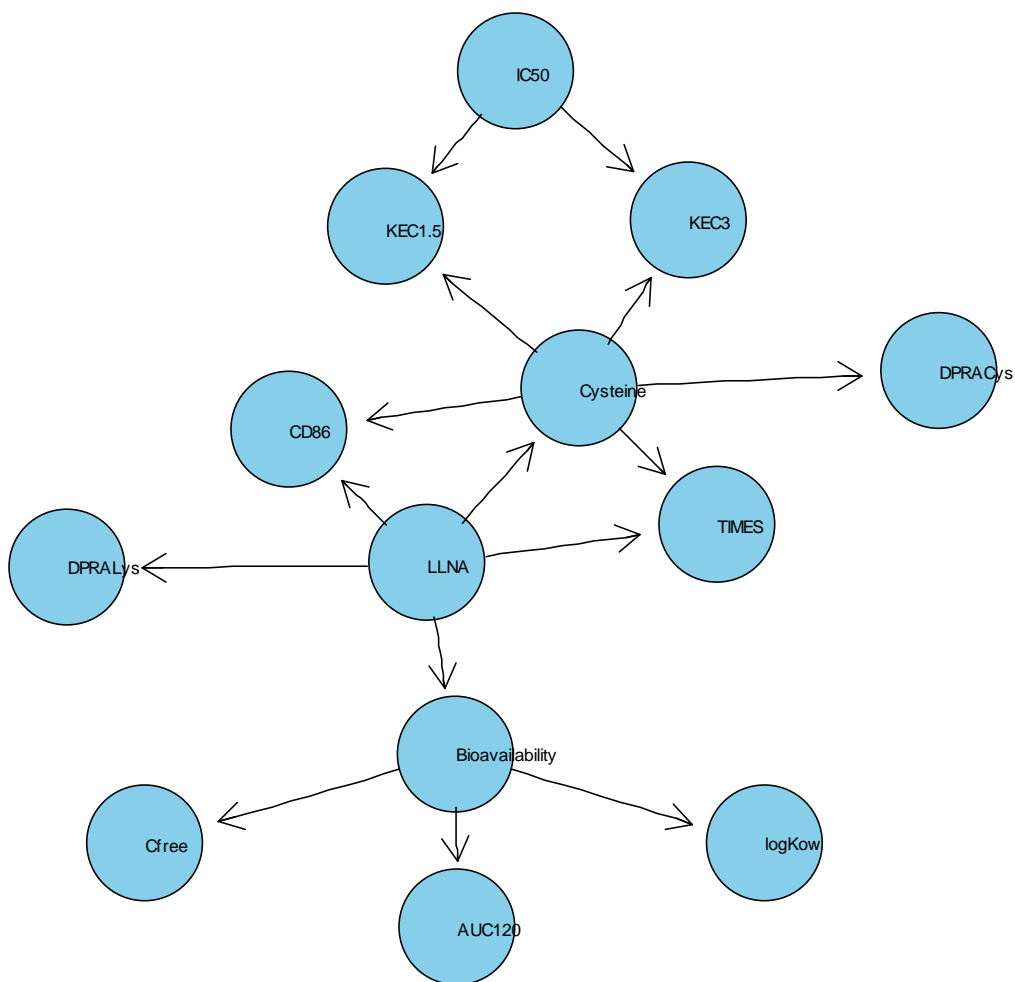
- The categorical LLNA potency predictions of the model were tested using 21 substances in an external test set: 6 nonsensitizers, 5 weak sensitizers, 5 moderate sensitizers, and 5 strong or extreme sensitizers.

Table 1. Variables for the Open-Source ITS-2 Lipid Model

Variable	Description	Measurement	Abbreviation in Figure 2
LLNA	Potency classification in four categories, based on the EC3 ranges described above	1 = nonsensitizer 2 = weak sensitizer 3 = moderate sensitizer 4 = strong or extreme sensitizer	LLNA
U937 Activation Test	<i>In vitro</i> test that uses the human myeloid cell line U937	EC150 (μM) for CD86 cell surface marker expression	CD86
Direct Peptide Reactivity Assay	<i>In chemico</i> method that measures peptide remaining after the test substance binds to two model heptapeptides	1) Percent cysteine peptide remaining 2) Percent lysine peptide remaining	1) DPRACys 2) DPRALys
KeratinoSens Assay	<i>In vitro</i> test that detects electrophiles using the Nrf2 electrophile-sensing pathway in the HaCaT (immortalized keratinocyte) cell line	1) Average concentration that produces 1.5-fold enhanced activity (μM) 2) Average concentration yielding 3-fold enhanced activity (μM) 3) Concentration producing 50% cytotoxicity (μM)	1) KEC1.5 2) KEC3 3) IC50
Physicochemical Property	Octanol–water partition coefficient	Log K_{ow}	log K_{ow}
Bioavailability	Concentration of chemical reaching the mid-epidermal layer of skin calculated using a transdermal transport model (Kasting et al. 2008).	1) Free test substance concentration in mid-epidermis multiplied by thickness of viable epidermis (0.01 cm) expressed as percent of applied dose 2) Area under the flux curve at 120 h (percent of applied dose)	1) Cfree 2) AUC120
TIMES-M	<i>In silico</i> categorical prediction of skin sensitization potency using TIMES (Tissue Metabolism Simulator) software (V.2.25.7), an expert system that makes predictions based on knowledge about the parent compound and potential skin metabolites (Dimitrov et al. 2005).	Three categories: nonsensitizer, weak sensitizer, and moderate/strong/extreme sensitizer	TIMES

Abbreviations: EC150 = effective concentration that produces a 1.5-fold increase in the CD86 cell surface marker expression, the threshold for a positive response; EC3 = effective concentration that produces a stimulation index of 3, the threshold for a positive response in the LLNA; LLNA = murine local lymph node assay.

Figure 2. Structure of the OS ITS-2 Lipid



The arrows show the conditional dependencies of the variables that impact murine local lymph node assay (LLNA) potency. LLNA potency category is the target variable. The remaining variables are manifest variables. Bioavailability and Cysteine are latent variables for bioavailability and cysteine binding, respectively. The abbreviations for all variables are listed in **Table 1**.

Results

- The LLNA potency category predictions of the OS ITS-2 lipid model using R for discretization with the Class-attribute Interdependence Maximization (CAIM) algorithm and latent class learning using the poLCA package are shown in **Tables 2** and **3** for the training sets and test sets, respectively. The bold red numbers in the tables show the results of the commercial software in cases where there is a difference between the OS ITS-2 lipid model and the commercial BN ITS-2 lipid model.
 - For the training set, the accuracy of LLNA potency category predictions was greater for the OS ITS-2 lipid model: 78% (97/124) vs. 76% (94/124) for the commercial BN ITS-2 lipid model.
 - Using the OS ITS-2 lipid model, 15 substances (12%) were overclassified (predicted category was more severe than observed in the LLNA) and 12 substances (10%) were underclassified (predicted category was less severe than observed in the LLNA).
 - Using the commercial BN ITS-2 lipid model, 21 substances (17%) were overclassified and 9 substances (7%) were underclassified.
 - For the test set, the accuracy of potency category predictions was identical for the OS ITS-2 lipid model: 86% (18/21) vs. 86% (18/21) for the commercial BN ITS-2 lipid model.
 - Using the OS ITS-2 lipid model, no substances were overclassified and 3 substances (14%) were underclassified.
 - Using the commercial BN ITS-2 lipid model, 1 substance (18%) was overclassified and 2 substances (10%) were underclassified.

Table 2. Confusion Matrix for LLNA Potency Category Predictions on the Training Set of 124 Substances

Predicted Potency Category ¹	Observed Potency Category ¹			
	Nonsensitizer (36)	Weak Sensitizer (28)	Moderate Sensitizer (35)	Strong/Extreme Sensitizer (25)
Nonsensitizer (36) (32)	31 29	2 1	1	2 1
Weak Sensitizer (27) (26)	3	22 21	2	0
Moderate Sensitizer (35)	1 3	3 4	26 24	5 4
Strong/Extreme Sensitizer (26) (31)	1	1 2	6 8	18 20

Abbreviations: LLNA = murine local lymph node assay.

¹ The numbers in parentheses show the total number of chemicals predicted or observed in each category. Numbers in **bold red** show the different values yielded by the BN ITS-2 lipid model developed using commercial software (Jaworska et al. 2013).

Table 3. Confusion Matrix for LLNA Potency Category Predictions on the Test Set of 21 Substances

Predicted Potency Category ¹	Observed Potency Category ¹			
	Nonsensitizer (6)	Weak Sensitizer (5)	Moderate Sensitizer (5)	Strong/Extreme Sensitizer (5)
Nonsensitizer (7)	6	1	0	0
Weak Sensitizer (5) (4)	0	4	1 0	0
Moderate Sensitizer (5)	0	0	4	1
Strong/Extreme Sensitizer (4) (5)	0	0	0 1	4

Abbreviations: LLNA = murine local lymph node assay.

¹ The numbers in parentheses show the total number of chemicals predicted or observed in each category. Numbers in **bold red** show the different values yielded by the BN ITS-2 lipid model developed using commercial software (Jaworska et al. 2013).

Case Studies

- Chlorobenzene and 2-mercaptobenzothiazole are two case studies that illustrate how the OS ITS-2 lipid model can use existing information to determine the *in vitro* or *in silico* tests that would be most effective for determining the potency classification.
- Value of information (VoI) from all possible sources determines which variable provides the most information about the target. VoI was assessed by calculating the mutual information between variables, which determines the uncertainty in one variable that is reduced by knowing the results from another variable.

1. Chlorobenzene

- Chlorobenzene is a solvent and chemical intermediate. It is typically a nonsensitizer in the LLNA and in guinea pig skin sensitization tests (ICCVAM 2009). It is assumed to be a nonsensitizer in humans due to a lack of evidence for skin sensitization (ICCVAM 2009).
- Testing Strategy
 - When the OS ITS-2 lipid model is trained to the training set of 124 substances, the variable with the highest mutual information, 0.74, is TIMES (**Figure 3a**).

- Because physicochemical properties may be obtained without wet laboratory work, assume that $\log K_{ow}$, and other physicochemical properties for calculating the bioavailability of chlorobenzene in skin are known and applied to the model. C_{free} and AUC120, measures of the bioavailability of chlorobenzene in the skin, are included in the model. Assume that the TIMES result, an *in silico* prediction, is applied to the model.

Potency Category Probabilities			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.82	0.084	0.072	0.028

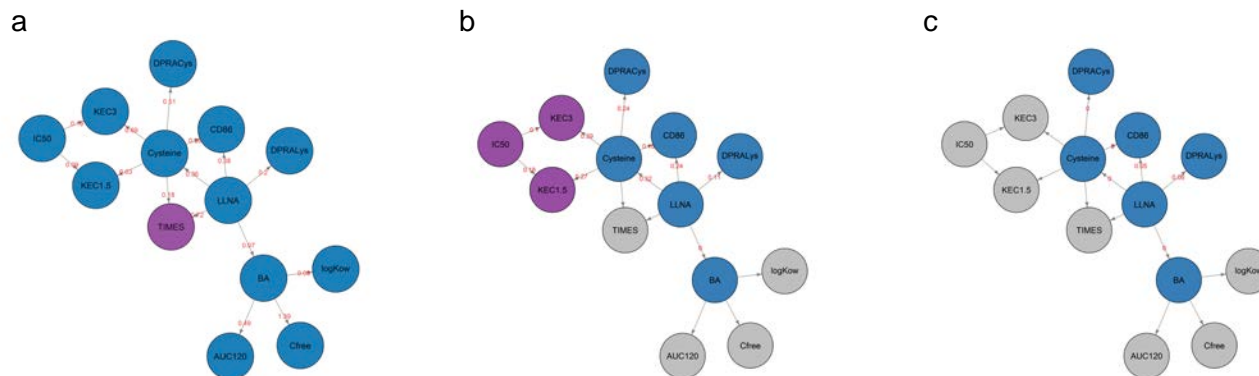
- The latent variable Cysteine has the highest mutual information for the LLNA, 0.32 (**Figure 3b**). The KeratinoSens variables, KEC1.5 and KEC3, have the highest mutual information for Cysteine (0.27 and 0.39, respectively).
- After obtaining the KeratinoSens data, including the IC50, the remaining variables have small mutual information values. Thus, no further testing is needed (**Figure 3c**).

Potency Category Probabilities (KeratinoSens Data)			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.92	0.049	0.00097	0.031

- When information on all the variables is applied, the probability for the nonsensitizer category increases by a small amount.

Potency Category Probabilities (All Variables)			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.97	0.018	0.00020	0.0072

Figure 3. Testing Strategy for Chlorobenzene



The abbreviations for the variables are listed in **Table 1**, except for BA = bioavailability. Blue indicates undefined variables, purple indicates the variables with the highest mutual information, and gray indicates variables with known values. **(A)** With no information on chlorobenzene, the variable with the highest mutual information is TIMES. **(B)** When the TIMES, $\log K_{ow}$, and bioavailability (Cfree and AUC120) are known, the KeratinoSens data have the highest mutual information for the latent variable Cysteine. **(C)** After KeratinoSens data are applied, the mutual information for the remaining variables is small.

2. 2-Mercaptobenzothiazole

- 2-Mercaptobenzothiazole is used in manufacturing to accelerate the vulcanization of rubber products. It is classified as a moderate sensitizer according to LLNA results. 2-Mercaptobenzothiazole is also a sensitizer in human and guinea pig tests (ICCVAM 2011).
- Testing Strategy
 - Assume that the *in silico* information is available: log K_{ow} , C_{free} , AUC120, and TIMES (**Figure 4a**)

Potency Category Probabilities			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.07	0.13	0.43	0.37

- The variable CD86 has the highest mutual information for the LLNA, 0.28 (**Figure 4a**).
- When probabilities are recalculated after obtaining the U937 activation test data:

Potency Category Probabilities (U937 Activation Test Data)			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.011	0.069	0.61	0.31

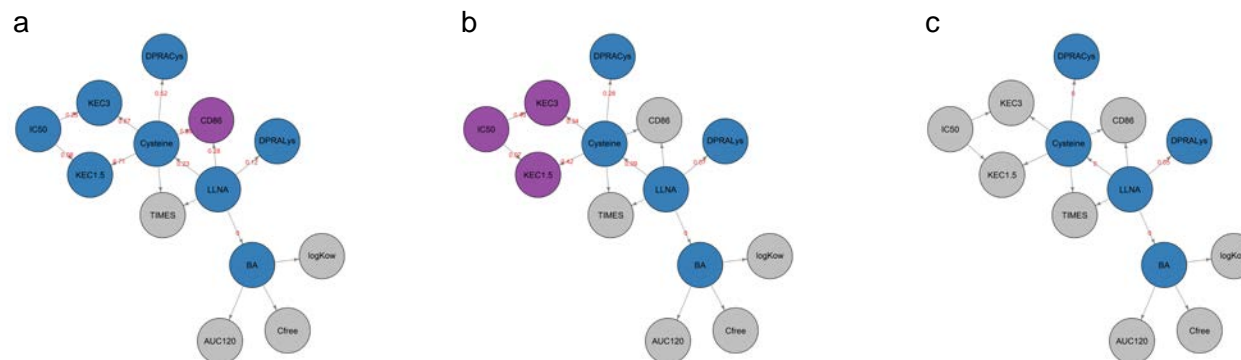
- The Cysteine latent variable has the highest mutual information for the LLNA, 0.09, and the KeratinoSens variables, KEC1.5 and KEC3, have the highest mutual information for Cysteine (0.42 and 0.34, respectively) (**Figure 4b**).
- After obtaining the KeratinoSens data, the probability for the moderate category increases. The remaining variables have small mutual information values (**Figure 4c**). Thus, no further testing is needed.

Potency Category Probabilities (KeratinoSens Data)			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.000045	0.036	0.67	0.29

- When all information is included, the probability for the moderate category increases again slightly.

Potency Category Probabilities (All variables)			
Nonsensitizer	Weak	Moderate	Strong/Extreme
0.000096	0.053	0.71	0.24

Figure 4. Testing Strategy for 2-Mercaptobenzothiazole



The abbreviations for the variables are listed in **Table 1**, except for BA = bioavailability. Blue indicates undefined variables, purple indicates the variables with the highest mutual information, and gray indicates variables with known values. **(A)** When the TIMES, logK_{ow}, and bioavailability (C_{free} and AUC120) are known, the CD86 data have the highest mutual information for the LLNA. After the CD86 data are applied, the highest mutual information for the LLNA is yielded by the latent variable Cysteine. **(B)** KeratinoSens data have the highest mutual information for Cysteine. **(C)** After KeratinoSens data are added, the mutual information for the remaining variable with value for the LLNA, DPRALys, is small.

Conclusions

- The OS ITS-2 lipid model for skin sensitization potency adequately reproduces the BN ITS-2 lipid model developed using commercial software.
- The open-source model
 - Increases the availability and transparency of the ITS
 - Represents a major step in allowing the ITS to be reproduced and tested, properties that are essential for implementation in a regulatory framework
- OS ITS-2 lipid is available to the public for testing at <http://ntp.niehs.nih.gov/go/its>.
- Future work will
 - Substitute the human cell line activation test for the U937 assay
 - Evaluate open source replacements for the TIMES-M *in silico* predictions and open sources for physicochemical properties needed for the bioavailability calculations
 - Add additional substances to the trained model as data are collected

References

- Dimitrov SD, Low LK, Patlewicz GY, et al. 2005. Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *Int J Toxicol* 24: 189–204.
- ICCVAM. 2009. Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication No. 09-7357. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available at <http://ntp.niehs.nih.gov/go/40465>
- ICCVAM. 2011. ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans. NIH Publication No. 11-7709. Research Triangle Park, NC:National Institute of Environmental Health Sciences. Available at <http://ntp.niehs.nih.gov/go/40458>
- Jaworska J, Harol A, Kern PS, Gerberick GF. 2011. Integrating non-animal test information into an adaptive testing strategy—skin sensitization proof of concept case. *ALTEX* 28: 211–225.
- Jaworska J, Dancik Y, Kern P, Gerberick GF, Natsch A. 2013. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. *J Appl Toxicol* 33: 1353–1364.

Kasting GB, Miller MA, Nitsche JM. 2008. Absorption and evaporation of volatile compounds applied to skin. In: *Dermatologic, Cosmeceutic and Cosmetic Development* (Walters KA and Roberts MS, eds). New York:Informa Healthcare USA, 385–400.

OECD. 2010. Test No. 429. Skin Sensitisation: Local Lymph Node Assay [adopted 22 July 2010]. In: *OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects*. Paris:OECD Publishing. Available: <http://dx.doi.org/10.1787/9789264071100-en>

OECD. 2012. OECD Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, Part 1: Scientific Assessment. Paris:OECD Publishing. Available: <http://www.oecd.org/env/ehs/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm> [accessed 2 Dec 2013]

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing (ISBN 3-900051-07-0). Vienna, Austria:R Foundation for Statistical Computing. Available: www.R-project.org

Acknowledgements

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by ILS under NIEHS contracts N01-ES 35504 and HHSN27320140003C and SSS, Inc., under NIEHS contract GS-23F-9806H.

The views expressed above do not necessarily represent the official positions of any Federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.



A summary of NICEATM and ICCVAM activities at SOT 2014 is available on the National Toxicology Program website at <http://ntp.niehs.nih.gov/go/41297>