# Open Source Software Implementation of an Integrated Testing Strategy for Skin Sensitization Potency Based on a Bayesian Network

*Jason R. Pirone[1], Marjolein Smith[1], Nicole C. Kleinstreuer[2], Thomas A. Burns[2], Judy Strickland[2], Yuri Dancik[3], Richard Morris[1], Lori A. Rinckel[2], Warren Casey[4] and Joanna S. Jaworska[3]*

[1]Social and Scientific Systems, Inc., Durham, NC, USA; [2]Integrated Laboratory Systems Inc., Morrisville, NC, USA; [3]Procter & Gamble NV, Strombeek-Bever, Belgium; [4]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

## Summary

*An open source implementation of a previously published integrated testing strategy (ITS) for skin sensitization using a Bayesian network has been developed using R, a free and open source statistical computing language. The ITS model provides probabilistic predictions of skin sensitization potency based on in silico and in vitro information as well as skin penetration characteristics from a published bioavailability model (Kasting et al., 2008). The structure of the Bayesian network was designed to be consistent with the adverse outcome pathway published by the OECD (Jaworska et al., 2011, 2013). In this paper, the previously published data set (Jaworska et al., 2013) is improved by two data corrections and a modified application of the Kasting model. The new data set implemented in the original commercial software package and the new R version produced consistent results. The data and a fully documented version of the code are publicly available (http://ntp.niehs.nih.gov/go/its).*

*Keywords: skin sensitization, integrated testing strategy, allergic contact dermatitis, Bayesian network, R*

Toxicity testing in the 21st century is purposefully transitioning from traditional disease-related observations in animal models towards the use of mechanism-based outcomes from cell-based assays and *in silico* models. However, it is unlikely that a single assay or *in silico* model will provide sufficient information on the risk or hazard posed by a chemical. Therefore, data from multiple inputs need to be integrated in a way that maximizes the utility of the available information. A Bayesian network is a graphical model that enables integration of data from multiple sources in a transparent and intuitive way. In situations where available data is incomplete or uncertain, Bayesian networks provide a coherent probabilistic framework for reasoning and guiding decisions on the classification of a substance or the need for additional testing.

The integrated testing strategy (ITS) using a Bayesian network for skin sensitization was previously developed using commercial software (Jaworska et al., 2011, 2013). The use of commercial software is convenient in some settings, but can limit the utility and awareness of an approach by obscuring the details of the analysis. Without full access to the code and data used to generate the model, it is difficult for others to test, verify and build on the model. Transparency was identified as one of the most important conceptual requirements of a successful ITS (Jaworska and Hoffmann, 2010). Accordingly, we developed an implementation of the Bayesian network ITS for skin sensitization using the free and open source statistical programming language R (R v3.0.1, GNU Public License v3).

A categorical representation of a compound's potency in the murine local lymph node assay (LLNA) is used as the target endpoint (Tab. 1) in the original Bayesian network ITS models (Jaworska et al., 2011, 2013). Relative to other *in vivo* sensitization assays, the LLNA achieves a reduction in number of animals used, less discomfort associated with a positive response and time required for completion, as well as providing a quantitative measure of skin sensitization potency. The LLNA is an internationally accepted method for assessing skin sensitization hazard (OECD, 2010).

The structure of the Bayesian network was designed to be consistent with the adverse outcome pathway (AOP) for substances that initiate the skin sensitization process by covalently binding to skin proteins (Jaworska et al., 2011, 2013). There are four key events in the AOP. In order of occurrence they are: 1)

**Tab. 1: LLNA EC3 correspondence to skin sensitization potency categories**

| Category Number | Category Description | EC3 Range |
|---|---|---|
| 1 | Nonsensitizer | No EC3 |
| 2 | Weak | EC3 ≥ 10% |
| 3 | Moderate | 1% ≤ EC3 < 10% |
| 4 | Strong and extreme | EC3 < 1% |

Abbreviations: 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

covalent binding to skin proteins, 2) inflammatory responses in the keratinocyte, 3) activation of dendritic cells and 4) T-cell proliferation (OECD, 2012). Table 2 links these events to the nodes (variables) found in the ITS structure (Jaworska et al., 2013) shown in Figure 1.

In a previous paper that described the ITS-2 model (Jaworska et al., 2013), both lipid and polar skin diffusion pathways were used for the bioavailability calculations and incorporated in an MS Excel version of the Kasting skin penetration model (Dancik et al., 2013). The bioavailability calculations for the lipid diffusion pathway are publicly accessible on the National Institute for Occupational Safety and Health website[1], but the polar skin diffusion pathway module is under development and not yet publicly available. Upon re-evaluation of the model, the contributed value of the polar skin diffusion pathway was not clear. Therefore, that pathway was dropped and the (minor) changes to the bioavailability nodes are contained in the current data set. Additionally, two errors in the direct peptide reactivity assay (DPRA) data for benzoic acid (training set) and imida-
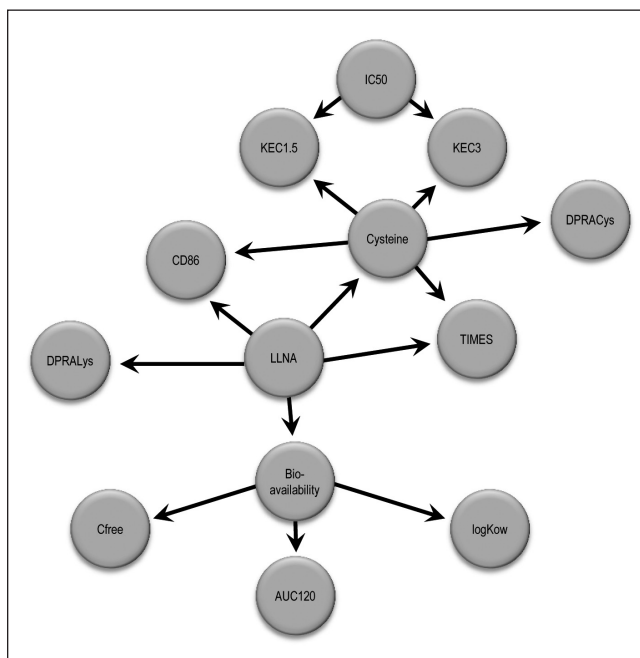


**Fig. 1: Directed acyclic graph representing the relationships among the variables in the ITS-2 lipid models**
Bioavailability and Cysteine represent latent (unobserved) variables (see text). All other nodes correspond to manifest (observed) variables and are described in Table 2.

zolidinyl urea (test set) were corrected. This revised model is referred to as ITS-2 Lipid.

Application of the Bayesian network requires three distinct computational steps as outlined in Figure 2. First, a supervised discretization algorithm is used to find cut-points that bin the
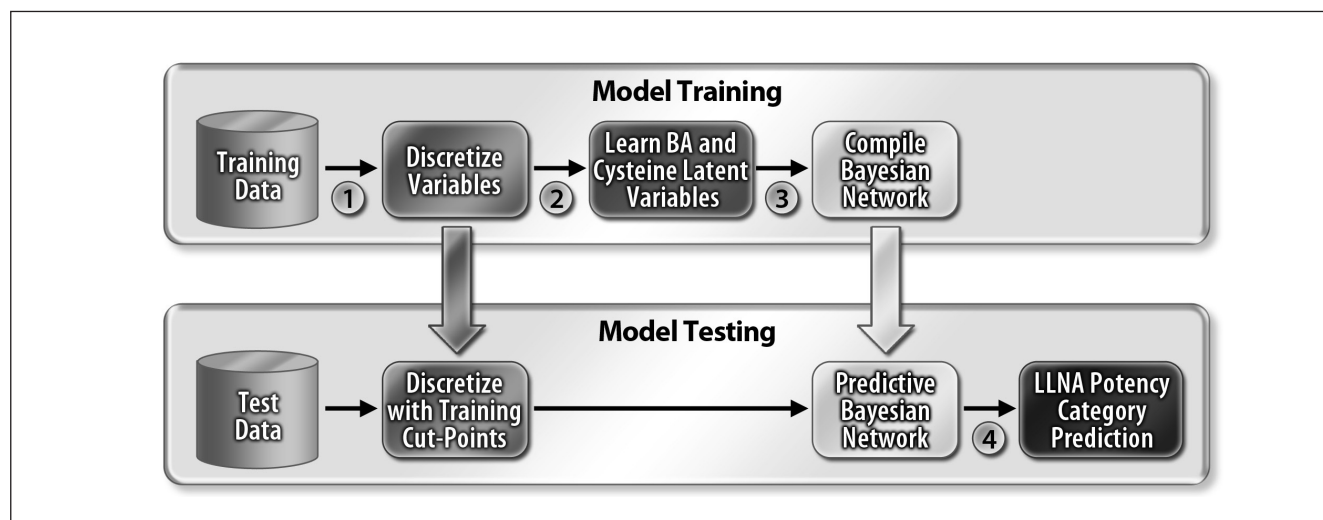


**Fig. 2: Diagram showing the primary computational steps of the ITS-2 lipid modeling process**
Abbreviations: BA, bioavailability; LLNA, murine local lymph node assay

---

[1] http://www.cdc.gov/niosh/topics/skin/finiteSkinPermCalc.html

**Tab. 2: Variables for the open source ITS-2 lipid model**

| Measurement | Description | Model Variable | AOP Key Event |
|---|---|---|---|
| Physicochemical property | Octanol-water partition coefficient | **logKow:** Log $K_{ow}$ | Substance must penetrate the stratum corneum (step 1 of the AOP; not a key event) |
| Epidermal bioavailability | Concentration of chemical reaching the mid-epidermal layer of skin calculated using a transdermal transport model (Kasting et al., 2008) | 1) **Cfree:** free test substance concentration in midepidermis multiplied by thickness of viable epidermis (0.01 cm) expressed as percent of applied dose<br>2) **AUC120:** area under the flux curve at 120 h as percent of applied dose | Substance must penetrate the stratum corneum (step 1 of the AOP; not a key event) |
| Direct Peptide Reactivity Assay (DPRA) | *In chemico* method that measures peptide remaining after the test substance binds to two model heptapeptides | 1) **DPRACys:** percent cysteine peptide remaining<br>2) **DPRALys:** percent lysine peptide remaining | 1) Binding to skin proteins |
| KeratinoSens Assay | *In vitro* test that detects electrophiles using the Nrf2 electrophile-sensing pathway in KeratinoSens cells, a reporter cell line derived from immortalized HaCaT keratinocytes | 1) **KEC1.5:** average concentration that produces 1.5-fold enhanced activity ($\mu$M)<br>2) **KEC3:** average concentration yielding 3-fold enhanced activity ($\mu$M)<br>3) **IC$_{50}$:** concentration producing 50% cytotoxicity ($\mu$M) | 2) Keratinocyte inflammatory responses |
| U937 Activation Test | *In vitro* test that uses the human myeloid cell line U937 | **CD86:** EC$_{150}$ ($\mu$M) for CD86 cell surface marker expression | 3) Dendritic cell activation |
| LLNA | *In vivo* test for skin sensitization; . EC3 is used to categorize potency as noted in Table 1 | **LLNA:** categorical representation of LLNA potency<br>1 = nonsensitizer<br>2 = weak sensitizer<br>3 = moderate sensitizer<br>4 = strong and extreme sensitizers | 4) T-cell proliferation |
| TIMES-M | *In silico* 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). | **TIMES:** three categories: nonsensitizer, weak sensitizer, and moderate/strong/extreme sensitizer | Not a key event of the AOP. Model links parent and metabolite structures to skin sensitization outcomes in animals and humans |

Abbreviations: AOP, adverse outcome pathway (OECD, 2012); EC$_{150}$, 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.

continuous assay data in the training data into intervals. The test data set cannot be used to find the discretization cut points, since doing so would result in biased and overly optimistic prediction results. The cut-points found for the training data are used to discretize the test data. Second, mechanistically related assays are clustered to form latent (unobserved) variables. The discretized *logKow*, *AUC120* and *Cfree* variables are clustered to form the *Bioavailability* latent variables. Similarly, the discretized results from the U937 assay, *CD86*, and the KeratinoSens assay, *KEC3* and *KEC1.5*, are clustered to form the *Cysteine* latent variable. Forming latent variables increases the interpretability of the network, while at the same time reducing its computational complexity. Finally, the relationships among variables in the discretized training data (including the latent variables) are described and quantitated using a Bayesian network. The Bayesian network has a qualitative and quantitative component. The qualitative part consists of a directed acyclic graph where each node represents an assay and each edge (arrow) indicates that there is a relationship between the variables it connects (Fig. 1). The strength of each

interaction is given by a set of conditional probability tables, one for each node, which makes up the quantitative portion of the network. The resulting model can be used to make LLNA potency category predictions for new chemicals.

As statistics and machine learning are the primary application domains of R, we were able to find high-quality R packages implementing each of the steps shown in Figure 2. The discretization package (Kim, 2012) contains implementations of several algorithms for supervised discretization (step 1). The *Bioavailability* and *Cysteine* latent variables were learned (step 2) using tools from the poLCA package (Linzer and Lewis, 2011). Finally, gRbase (Dethlefsen and Højsgaard, 2005) and gRain (Højsgaard, 2012) supply the functions for constructing, parameterizing and performing inference on Bayesian networks (step 3). These packages and their use are discussed in detail in the code documentation found on the National Toxicology Program (NTP) website[2].

Construction and validation of the open source ITS network occurred in three stages. First, we demonstrated that the tools available in R for building the probability tables and performing inference on Bayesian networks could give identical results to those obtained using the commercial software. This was done by feeding the results of steps 1 and 2 obtained from the commercial software into step 3 of the R model. With the same discretization cut-points and latent node values, the LLNA category predictions for the training and test datasets were identical for the R implementation and the commercial software. In the second stage, we compare the latent variable learning algorithm used in the R model with that used by the commercial software package. The exact algorithm used by the commercial software is not known, but we compared the methods by using the dataset discretized by the commercial software package as the input to the latent variable learning algorithm implemented in R (step 2). The R and commercial algorithms grouped the training set chemicals in the same way to form the latent variables. Finally, we used R versions of widely used algorithms for supervised discretization and latent class learning (steps 1 and 2), and then built the network in R as well (step 3). The overall classification accuracies between the R-based method and the commercial software package were the same, with three compounds misclassified by both methods. The predictions were not identical, however, as two compounds were classified differently by the two methods. Dihydroeugenol (2-methoxy-4-propyl-phenol) (CASRN 2785-87-7) was correctly classified as a moderate sensitizer by the R-based method and incorrectly classified as a strong sensitizer by the commercial software. Citral (CASRN 5392-40-5) was incorrectly classified as a weak sensitizer by the R-based method and correctly classified as a moderate sensitizer by the commercial software package.

The open source R-based model with computational details is available on the NTP website[2], where interested users with intermediate R programming skills can access: 1) the R code and detailed documentation, 2) the current data set (with corrections) used to train and test the model and 3) output data to

verify their computations. The model is documented using the Sweave application (Leisch, 2002), which produces a "dynamic document" that integrates R code and expository text, thereby making each analysis step explicit. The Sweave document coupled with input and output datasets makes the Bayesian network ITS model independently reproducible (Fomel and Claerbout, 2009; Koenker and Zeileis, 2009; Walters, 2013; Peng, 2009).

The modular nature of the network programmed in R allows for alternative *in silico* and *in chemico* modules to serve as inputs to the Bayesian network ITS, which would then need to be retrained. Substituting *in vitro* modules could be more challenging, as that could potentially affect the structure of the network, but represents another opportunity provided by this transparent and flexible approach. A forum to exchange information and updates to the R code for the user community interested in Bayesian approaches to skin sensitization modeling is also available at the NTP website. This open source version provides the transparency necessary to make the model reproducible and accessible to others within the scientific and regulatory communities.

The present work reflects an ongoing collaboration between (NICEATM) and scientists from Procter & Gamble and Social and Scientific Systems, Inc, to develop and promote Bayesian networks as an operational framework for an ITS. NICEATM primarily provides technical and scientific support for the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), however it also conducts various collaborative projects that are independent of this committee, an example of which is presented here. Finally, ICCVAM is also committed to the evaluation of skin sensitization test methods and testing strategies that do not require the use of animals, however its plans to address such methods have not yet been finalized (NIEHS, 2013).

## References

Dancik, Y., Miller, M. A., Jaworska, J. et al. (2013). Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. *Adv Drug Deliv Rev 65*, 221-236. http://dx.doi.org/10.1016/j.addr.2012.01.006

Dethlefsen, C. and Højsgaard, S. (2005). A common platform for graphical models in R: The gRbase package. *J Stat Softw 14*, 1-12.

Dimitrov, S. D., Low, L. K., Patlewicz, G. Y. et al. (2005). Skin sensitization: Modeling based on skin metabolism simulation and formation of protein conjugates. *Int J Toxicol 24*, 189-204. http://dx.doi.org/10.1080/10915810591000631

Fomel, S. and Claerbout, J. F. (2009). Guest editors' introduction: Reproducible research. *Comput Sci Eng 11*, 5-7. http://dx.doi.org/10.1109/MCSE.2009.14

Højsgaard, S. (2012). Graphical independence networks with the gRain package for R. *J Stat Softw 46*, 1-26.

Jaworska, J. and Hoffmann, S. (2010). Integrated Testing Strategy (ITS) – Opportunities to better use existing data and guide

---

[2] http://ntp.niehs.nih.gov/go/its

future testing in toxicology. *ALTEX 27*, 231-242. http://www.altex.ch/Current-issue.50.html?iid=121&aid=1 (accessed 27.03.2014)

Jaworska, J., Harol, A., Kern, P. S. et al. (2011). Integrating non-animal test information into an adaptive testing strategy – Skin sensitization proof of concept case. *ALTEX 28*, 211-225. http://dx.doi.org/10.14573/altex.2011.3.211

Jaworska, J., Dancik, Y., Kern, P. et al. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. *J Appl Toxicol 33*, 1353-1364.

Kasting, G. B., Miller, M. A. and Nitsche, J. M. (2008). Absorption and evaporation of volatile compounds applied to skin. In K. A. Walters and M. S. Roberts (eds.), *Dermatologic, Cosmeceutic and Cosmetic Development*. New York, USA: Informa Healthcare.

Kim, H. J. (2012). Discretization: Data reprocessing, discretization for classification. R package version 1.0-1. http://cran.r-project.org/web/packages/discretization/discretization.pdf

Koenker, R. and Zeileis, A. (2009). On reproducible econometric research. *J Appl Econom 24*, 833-847. http://dx.doi.org/10.1002/jae.1083

Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle and B. Rönz (eds.), *Compstat 2002 – Proceedings in Computational Statistics*. Heidelberg, Germany: Physica Verlag.

Linzer, D. A. and Lewis, J. (2011). "poLCA: an R Package for Polytomous Variable Latent Class Analysis." *J Stat Softw 42*, 1-29. http://www.jstatsoft.org/v42/i10

NIEHS (2013). Request for information on alternative skin sensitization test methods and testing strategies and for comment on ICCVAM's proposed activities. *Fed Regist 78*, 68076-68077.

OECD (2010). Test No. 429. Skin Sensitisation: Local Lymph Node Assay. In (eds.), *OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects*. Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264071100-en

OECD Publishing (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.

Peng, R. (2009). Reproducible research and Biostatistics. *Biostatistics 10*, 405-408. http://dx.doi.org/10.1093/biostatistics/kxp014

Walters, W. (2013). Modeling, informatics, and the quest for reproducibility. *J Chem Inf Model 53*, 1529-1530. http://dx.doi.org/10.1021/ci400197w

## Correspondence to

Warren Casey
NICEATM, National Institute of Environmental
Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park
NC 27709, USA
Phone: +1 919 316 4729
Fax: +1 919 541 s0947
e-mail: warren.casey@nih.gov