Draft Background Review Document
Use of the Murine Local Lymph Node Assay (LLNA)
to Determine Skin Sensitization Potency Categories

January 2008

[This Page Intentionally Left Blank]

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ACD	Allergic contact dermatitis
AOO	Acetone: olive oil
BRD	Background Review Document
BT	Buehler Test
CASRN	Chemical Abstracts Service Registry Number
Conc.	Concentration tested
CPSC	U.S. Consumer Product Safety Commission
DMSO	Dimethyl sulfoxide
EC3	Estimated concentration needed to produce a stimulation index of three
ECVAM	European Centre for the Validation of Alternative Methods
EPA	U.S. Environmental Protection Agency
FDA	U.S. Food and Drug Administration
FHSA	Federal Hazardous Substances Act
FR	*Federal Register*
GHS	United Nations Globally Harmonized System for the Labelling and Classification of Chemicals
GLP	Good Laboratory Practice
GPMT	Guinea Pig Maximization Test
HCA	Hexyl cinnamic aldehyde
HMT	Human Maximization Test
HPTA	Human Patch Test Allergen
HRIPT	Human repeat-insult patch test
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IPCS	International Programme on Chemical Safety
IWG	Immunotoxicity Working Group

JaCVAM	Japanese Center for the Validation of Alternative Methods
K_{ow}	Octanol-water partition coefficient
LLNA	Local Lymph Node Assay
LOEL	Lowest observed effect level
LOEL/10	LOEL divided by a safety factor of 10
MeSH	Medical Subject Headings
NC	Not calculated
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NOEL	No observed effect level
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods
SI	Stimulation index
TG	Test Guideline
U.K.	United Kingdom
U.N.	United Nations
U.S.	United States
w/v	Weight to volume ratio
WHO	World Health Organization

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Designated Agency Representatives[1]

Agency for Toxic Substances and Disease Registry
• Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission
• Marilyn L. Wind, Ph.D. (Chair)
◊ Kristina Hatlelid, Ph.D.
* Joanna Matheson, Ph.D.

Department of Agriculture
• Jodie Kulpa-Eddy, D.V.M. (Vice-Chair)
◊ Elizabeth Goldentyer, D.V.M.

Department of Defense
• Robert E. Foster, Ph.D.
◊ Patty Decot
* Peter J. Schultheiss, D.V.M., D.A.C.L.A.M.
* Harry Salem, Ph.D.

Department of Energy
• Michael Kuperberg, Ph.D.
◊ Marvin Stodolsky, Ph.D.

Department of the Interior
• Barnett A. Rattner, Ph.D.
◊ Sarah Gerould, Ph.D.

Department of Transportation
• George Cushmac, Ph.D.
◊ Steve Hwang, Ph.D.

Environmental Protection Agency
Office of Science Coordination and Policy
• Karen Hamernik, Ph.D.
Office of Research and Development
◊ Julian Preston, Ph.D.
* Suzanne McMaster, Ph.D.
OECD Test Guidelines Program
* Jerry Smrchek, Ph.D.
Office of Pesticides Programs
* Amy Rispin, Ph.D.
* Deborah McCall

Food and Drug Administration
Office of Science
• Suzanne Fitzpatrick, Ph.D., D.A.B.T.
Center for Drug Evaluation and Research
◊ Abigail C. Jacobs, Ph.D.
Center for Devices and Radiological Health
* Melvin E. Stratmeyer, Ph.D.
Center for Biologics Evaluation and Research
* Richard McFarland, Ph.D., M.D.
* Ying Huang, Ph.D.
Center for Food Safety and Nutrition
* David G. Hattan, Ph.D.
* Robert L. Bronaugh, Ph.D.
Center for Veterinary Medicine
* Devaraya Jagannath, Ph.D.
* M. Cecilia Aguila, D.V.M.
National Center for Toxicological Research
* William T. Allaben, Ph.D.
* Paul Howard Ph.D.
Office of Regulatory Affairs
* Lawrence A. D'Hoostelaere, Ph.D.

National Cancer Institute
• Alan Poland, M.D.
◊ T. Kevin Howcroft, Ph.D.

National Institute of Environmental Health Sciences
• William S. Stokes, D.V.M., D.A.C.L.A.M.
◊ Raymond R. Tice, Ph.D.
* Rajendra S. Chhabra, Ph.D., D.A.B.T
* Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
• Paul Nicolaysen, V.M.D.
◊ K. Murali Rao, M.D., Ph.D.

National Institutes of Health
• Margaret D. Snyder, Ph.D.

National Library of Medicine
◊ Jeanne Goshorn, M.S.

Occupational Safety and Health Administration
• Surender Ahir, Ph.D.

• Principal Agency Representative
◊ Alternate Principal Agency Representative
* Other Designated Agency Representative

[1] Roster as of December 2007.

Acknowledgements

The following individuals are acknowledged for their contributions to the evaluation of the use of the LLNA to assess relative skin sensitization potency .

Interagency Coordinating Committee on the Validation on Alternative Methods (ICCVAM) Immunotoxicity Working Group (IWG)

U.S. Consumer Product Safety Commission
Joanna Matheson, Ph.D. (IWG Co-Chair)
Marilyn Wind, Ph.D.

U.S. Environmental Protection Agency
Karen Hamernik, Ph.D.
Masih Hashim, Ph.D.
Marianne Lewis
Deborah McCall
Timothy McMahon, Ph.D.
Amy Rispin, Ph.D
MaryJane Selgrade, Ph.D.
Marsha Ward, Ph.D.
Ronald Ward, Ph.D.

U.S. Food and Drug Administration
Ruth Barratt, Ph.D., D.V.M.
Paul Brown, Ph.D.
Abigail Jacobs, Ph.D. (Co-Chair)
Daniel Lyle, Ph.D.
Jiaqin Yao, Ph.D.

National Institute of Environmental Health Sciences
Dori Germolec, Ph.D.
William S. Stokes, D.V.M., D.A.C.L.A.M.
Raymond R. Tice, Ph.D.

National Institute for Occupational Safety and Health
Jean Meade, D.V.M., Ph.D.

ECVAM Liaison
Silvia Casati, Ph.D.

JaCVAM Liaison
Hajime Kojima, Ph.D.

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences

William Stokes, D.V.M., D.A.C.L.A.M.	Director; Project Officer
Raymond Tice, Ph.D.	Deputy Director
Deborah McCarley	Special Assistant; Asst. Project Officer

NICEATM Support Contract Staff (Integrated Laboratory Systems [ILS], Inc.)

David Allen, Ph.D.	Principal Investigator
Douglas Winters, M.S.	Project Manager
Neepa Choksi, Ph.D.	Senior Staff Toxicologist
Judy Strickland, Ph.D., D.A.B.T.	Senior Staff Toxicologist
Frank Deal, M.S.	Staff Toxicologist
Elizabeth Lipscomb, Ph.D.	Staff Toxicologist
Eleni Salicru, Ph.D.	Staff Toxicologist
Thomas Burns, M.S.	Senior Project Coordinator/Technical Writer
Michael Paris	Senior Project Coordinator/Technical Writer
Patricia Ceger, M.S.	Project Coordinator/Technical Writer
James Truax, M.A.	Project Coordinator/Technical Writer
Catherine Sprankle	Communications Specialist/Web Developer
Linda Litchfield	Meeting Planner and Coordinator

Other Acknowledgements

ICCVAM and NICEATM gratefully acknowledge the following individuals and institutions who submitted data to NICEATM used for the evaluation of the use of the LLNA to assess relative skin sensitization potency.

David Basketter, Ph.D.
Unilever Safety and Environmental Assurance Centre
Sharnbrook, UK

Phil Botham, Ph.D.
European Crop Protection Association
Brussels, Belgium

Eric Debruyne, Ph.D.
Bayer CropScience SA, Sophia Antipolis Cedex, France

George DeGeorge, Ph.D. and Melissa Kirk, Ph.D.
MB Research Labs
Spinnerstown, PA, USA

G. Frank Gerberick, Ph.D.
Procter and Gamble Company
Cincinnati, OH

Dori Germolec, Ph.D.
National Toxicology Program
Research Triangle Park, NC, USA

Ian Kimber, Ph.D.
Syngenta Central Toxicology Laboratory
Macclesfield, UK

Michael J. Olson, Ph.D.
GlaxoSmithKline
Research Triangle Park, NC, USA

Kirill Skirda, Ph.D.
TNO Quality of Life
Delft, Netherlands

Masahiro Takeyoshi, Ph.D.
Chemicals Evaluation and Research Institute
Oita, Japan

Peter Ungeheuer, Ph.D.
European Federation for Cosmetic Ingredients
Frankfurt, Germany

Hans Werner Vohr, Ph.D.
Bayer HealthCare
Wuppertal-Elberfeld, Germany

Preface

In 1998, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in conjunction with the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) evaluated the validation status of the murine local lymph node assay (LLNA) as an alternative to guinea pig test methods for assessing the allergic contact dermatitis (ACD) potential of substances. As described in the ICCVAM evaluation report[1], ICCVAM recommended that the LLNA could be used as a valid substitute for most testing situations.

Based on the ICCVAM recommendations, the ICCVAM member agencies that require the regulatory submission of ACD data accepted the LLNA, with the identified limitations, as an alternative to guinea pig tests for assessing ACD. In 2002, the LLNA was adopted as Test Guideline 429 by the 30-member countries of the Organisation for Economic Co-operation and Development (OECD). On January 10, 2007, the U.S. Consumer Product Safety Commission (CPSC) formally nominated several activities related to the LLNA for evaluation by ICCVAM and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). One of the nominated activities was an assessment of the validation status of the LLNA as a stand-alone assay for potency determinations for classification purposes.

After considering comments from the public and the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) on this nomination, ICCVAM assigned it a high priority, and directed NICEATM and the ICCVAM Immunotoxicity Working Group (IWG) to conduct a review of the current literature and an evaluation of the available data. The information described in this background review document (BRD) was compiled by ICCVAM in response to this nomination. ICCVAM and its IWG developed draft test method recommendations based on this evaluation. An independent peer review panel (Panel) will be convened to peer review the BRD, to assess the adequacy of the available data and information in the BRD, and to evaluate the extent to which this data and information support the draft ICCVAM recommendations regarding use of the LLNA for potency categories for

[1]ICCVAM (1999), available at http://iccvam.niehs.nih.gov/methods/immunotox/llna_PeerPanel98.htm

hazard classification purposes. ICCVAM will consider the conclusions and recommendations of the Panel, along with comments received from the public and SACATM, when developing a final BRD and final recommendations for each of the nominated activities.

We gratefully acknowledge the organizations and scientists who provided data and information for this document. We would also like to recognize the efforts of the individuals who contributed to the preparation of this BRD. These include David Allen, Ph.D., Thomas Burns, M.S., Neepa Choksi, Ph.D., Michael Paris, Eleni Salicru, Ph.D., Catherine Sprankle, Judy Strickland, Ph.D., and Doug Winters, M.S., of Integrated Laboratory Systems, Inc., the NICEATM Support Contractor, as well as the members of the ICCVAM IWG and the ICCVAM representatives who subsequently reviewed and provided comments throughout the process leading to this final draft version. We also want to thank Raymond Tice, Ph.D., Deputy Director of NICEATM, for his contributions to this project. Finally, we want to recognize the excellent leadership of the IWG Co-chairs, Abigail Jacobs, Ph.D. (FDA) and Joanna Matheson, Ph.D. (CPSC).

Marilyn Wind, Ph.D.

Deputy Associate Executive Director

Directorate for Health Sciences

U.S. Consumer Product Safety Commission

Chair, ICCVAM

William S. Stokes, D.V.M., D.A.C.L.A.M.

Rear Admiral, U.S. Public Heath Service

Director, NICEATM

Executive Director, ICCVAM

(insert date of final review here)

Executive Summary

In 1999, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) recommended that the murine local lymph node assay (LLNA) is a valid substitute for currently accepted guinea pig test methods to assess the allergic contact dermatitis (ACD) potential of many, but not all types of substances. The recommendation was based on a comprehensive evaluation that included an independent scientific peer review panel (Panel) assessment of the validation status of the LLNA (ICCVAM 1999[2]).

ICCVAM forwarded recommendations to U.S. Federal agencies that the LLNA should be considered for regulatory acceptance or other non-regulatory applications for assessing the ACD potential of substances, while recognizing that some testing situations would still require the use of traditional guinea pig test methods (ICCVAM 1999, Sailstad et al. 2001). The LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization (Organisation for Economic Co-operation and Development [OECD] Test Guideline 429 [OECD 2002]; International Standards Organization [ISO] 10993-10: Tests for Irritation and Sensitization [ISO 2002]; U.S. Environmental Protection Agency [EPA] Health Effect Testing Guidelines on Skin Sensitization [EPA 2003]).

On January 10, 2007, the U.S. Consumer Product Safety Commission (CPSC) formally nominated[3] several activities related to the LLNA for evaluation by ICCVAM and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). One of the nominated activities was an assessment of the usefulness and limitations of the LLNA as a stand-alone assay for potency determinations for classification purposes. The information described in this background review document (BRD) was compiled by ICCVAM and NICEATM in response to this nomination. The BRD provides a comprehensive review of available data and information regarding the use of the LLNA as a stand-alone assay for determining potency (including severity) for the purpose of hazard classification.

[2] available at http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf
[3] Available at http://iccvam.niehs.nih.gov/methods/immunotox/llnadocs/CPSC_LLNA_nom.pdf

For the purposes of this BRD, relative potency is defined as the concentration of a fixed volume of a substance that is required for either the induction or elicitation phases of a skin sensitization reaction. For induction, potency refers to the concentration of a substance needed to induce a sensitization response; the more potent the substance the smaller the quantity needed for induction. Likewise, for elicitation, potency refers to the concentration of a substance needed to elicit a response in a previously sensitized individual; the more potent a substance, the smaller the quantity required for elicitation.

The ICCVAM recommended LLNA protocol provides a detailed description of the conduct of the assay (ICCVAM 1999, Dean et al. 2001). A test-substance-induced increase in lymphocyte proliferation in the draining lymph nodes of the ear, the site of application, is used in the LLNA to identify chemical sensitizers. Mice are injected with radiolabeled thymidine (or an analogue of thymidine), which is incorporated into the DNA of proliferating cells. The Stimulation Index (SI), which is the ratio of incorporated radioactivity in the auricular lymph nodes of treated versus control mice, is used to assess the sensitizing potential of the test substance. An SI of three or greater is used to classify a test substance as a skin-sensitizing agent. The estimated concentration, in a volume of 25 μL/ear applied to both ears, needed to produce an SI =3 (i.e., the EC3) is used as the metric for predicting sensitization potency using the LLNA.

The information summarized in this BRD is based on a retrospective review of traditional LLNA data. Data were obtained from published reports and unpublished data submitted to NICEATM in response to a *FR* notice (Vol. 72, No. 95, pp. 27815-27817[4]). The information included in this BRD is based on a retrospective review of LLNA, guinea pig, and human data derived from a database of over 500 substances, 170 of which have comparative LLNA, guinea pig, and/or human data. Among these 170 substances, there are 112 substances with comparative human data (97 sensitizers, 15 non-sensitizers), 105 substances with comparative guinea pig data (52 sensitizers, 53 non-sensitizers), and 47 substances with comparative human and guinea pig data (34 sensitizers, 13 non-sensitizers).

[4] , available at http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf

The reference database for this evaluation consisted of human clinical studies (e.g., human repeat insult patch test and human maximization test) and the currently accepted guinea pig test methods for skin sensitization (i.e., the guinea pig maximization test [GPMT] and the Buehler test [BT]). For each substance with comparative LLNA and human data, potency was evaluated by comparing the LLNA EC3 concentration against the threshold concentration inducing the human response. For each substance with comparative LLNA and guinea pig data, potency was evaluated by comparing the LLNA EC3 concentration against the percentage of responding guinea pigs in the BT or GPMT and the associated induction concentration.

Ability of the LLNA to Predict Skin Sensitization Potency in Humans

In the current NICEATM LLNA database, there are 112 substances with both LLNA and human data, 81 of which are classified as sensitizers in both the LLNA and in the HMT and/or the HRIPT. Two approaches were used to evaluate the ability of the LLNA to predict sensitization potency in humans. In the first approach, for each substance classified as a sensitizer in both the LLNA and in humans, the LLNA EC3 concentration (expressed in $\mu g/cm^2$ and not as a percent) was correlated against the human threshold response (i.e., either the NOEL or LOEL/10, expressed in $\mu g/cm^2$). In the second approach, using the same set of 81 sensitizers used in the first approach, the human sensitizers were classified into strong or weak based on using either of two proposed decision criteria (strong sensitizers < 250 or < 500 $\mu g/cm^2$). Next, the optimal EC3 value that maximized obtaining the correct skin sensitization calls for strong and weak sensitizers (using one or the other proposed decision criterion) was pragmatically determined and the correct classification rate as well as the over- and under-classification rates calculated. In a variant of the second approach, substances that were classified in the LLNA as false positives (i.e., sensitizers in the LLNA but non-sensitizers in humans), false negatives (i.e., non-sensitizer in the LLNA but sensitizers in human tests), and non-sensitizers in both the LLNA and in human tests were included, the optimal EC3 values were re-calculated, and then the correct classification rate as well as the over- and under-classification rates re-calculated for each sensitization category (strong sensitizer, weak sensitizer, non-sensitizer).

A regression analysis of LLNA EC3 versus human threshold values for the 81 LLNA/human sensitizers, both scaled in μg/cm^2 and based on log transformed data, indicated a positive correlation with an R^2 value of 0.325 ($P<0.0001$) and 0.405 ($P<0.0001$) when either the most potent LLNA EC3 and human threshold values or the geometric mean for multiple test results was used, respectively. However, this correlation is not very strong, as evidenced by relatively low R^2 values. Based on an analysis of slope and intercept, the two regressions are not significantly different ($p = 0.125$ for slope and $p=0.620$ for intercept). However, based on the higher R^2 value (0.405) achieved when geometric means of multiply tested substances were calculated, this approach was carried forward through the remainder of the performance analyses.

These results were also compared to linear regression data for LLNA EC3 values versus various sets of human threshold data as published previously (Griem et al. 2003, Schneider and Akkam 2004, Basketter et al. **Appendix A**). There are differences in R^2 values among the various analyses, which presumably reflect differences in the number of substances with both LLNA EC3 and human sensitization threshold data, which human test is considered (i.e., HMT and/or HRIPT), how NOEL and/or LOEL values are used (e.g., LOEL or LOEL/10), and how data for substances tested multiply times are collapsed into a single value.

In the first approach to estimate the correct classification rate as well as the over- and under-classification rates for the LLNA based on the two proposed decision criteria for distinguishing between strong and weak sensitizers in humans, the analysis considered only those substances classified as sensitizers in both the LLNA and in humans based on the HMT and/or HRIPT. For these 81 LLNA/human sensitizers, the optimal EC3 values are 6.8% and 8.1% when 250 μg/cm^2 and 500 μg/cm^2, respectively, are used to distinguish between strong and weak skin sensitizers in humans. Using these two EC3 values, the correct classification rate was 74% (60/81) and 70% (57/81) for 250 and 500 μg/cm^2, respectively, while the over- and under-classification rates ranged from 28% (10/36) to 31% (9/29) and 24% (11/45) to 29% (15/52), respectively.

When the analysis took into consideration those substances classified in the LLNA as false positives and false negatives against human skin sensitization data, as well as those classified

as non-sensitizers in both the LLNA and in humans, the optimal EC3 value was 9.4% for either human threshold concentration. Using all 112 substances with both LLNA and human data, the correct classification rate was 62% (70/112) and 60% (67/112) for 250 and 500 $\mu g/cm^2$, respectively, while the over- and under-classification rates ranged from 26% (13/50) to 33% (5/15) and 21% (10/47) to 33% (14/43), respectively.

Ability of the LLNA to Predict Skin Sensitization Potency in Guinea Pigs

In the current NICEATM LLNA database, there are 105 substances with both LLNA and guinea pig test data, 52 of which are classified as sensitizers both in the LLNA and in the guinea pig. In this analysis, for multiply tested substances, the geometric mean LLNA EC3 value was used, while a weight-of-evidence evaluation was used to categorize the guinea pig test results. In this approach, test results from either GPMT or BT tests were considered together when assigning an overall classification category. Next, the correct classification rate as well as the over- and under-classification rates against the guinea pig results were calculated, using this optimal EC3 value. In a variant of this approach, substances that were classified in the LLNA as false positives (i.e., sensitizers in the LLNA but non-sensitizers in guinea pigs), false negatives (i.e., non-sensitizer in the LLNA but sensitizers in guinea pigs), and non-sensitizers in both the LLNA and in guinea pigs were included, the optimal EC3 value was re-calculated and then the correct classification rate as well as the over- and under-classification rates re-calculated for each sensitization category (strong sensitizer, weak sensitizer, non-sensitizer). In these various analyses, for substances that had more than one EC3 or guinea pig response, the geometric mean EC3 value and the weight of evidence GP classification category was used. Although the data generated by the GPMT and the BT is categorical, using the weight of evidence categorization provided some measure of a mean response across multiple studies.

In the first analysis, which focused only on substances classified as sensitizers in both the LLNA and in guinea pigs, overclassification means that weak sensitizers are missclassified as strong while underclassification means that strong sensitizers are missclassified as weak. Using the optimal EC3 value of 2.0%, the correct classification rate was 73% (38/52), while the over- and under-classification rates were 28% and 26%, respectively. The second analysis included substances classified as sensitizers in both the LLNA and in humans as well as

substances classified in the LLNA as false positives and false negatives compared to the human, and substances classified as non-sensitizers in both the LLNA and in humans. In this analysis, overclassification means that nonsensitizers are misclassified as weak or strong sensitizers and weak sensitizers are missclassified as strong while underclassification means that strong sensitizers are missclassified as weak or nonsensitizers and weak sensitizers are misclassified as nonsensitizers. Using the optimal EC3 value of 3.6%, the correct classification rate was 57% (60/105), while the over- and under-classification rates ranged from 25% (8/32) to 61% (30/49) and 9% (3/32) to 17% (4/24), respectively.

Ability of the LLNA versus the Guinea Pig to Predict Skin Sensitization Potency in Humans

In the current NICEATM LLNA database, there are 47 substances with human, LLNA, and guinea pig test results, 32 of which are classified as sensitizers in all three species. In this evaluation, the geometric mean EC3 and human threshold values and the weight of evidence sensitization classification in the guinea pig were used. In the first analysis, only those substances classified as sensitizers in all three species were evaluated. In the second analysis, all substances with data in all three species were evaluated. Based on the results of these analyses and depending on the decision criteria used in humans to distinguish between strong and weak sensitizers (i.e., 250 vs. 500 $\mu g/cm^2$) and whether only sensitizers in all three species or all data were used, the LLNA achieved the correct classification at a slightly higher rate, as evidenced by overall classification rates of 57% (27/47) to 72% (23/32) for the LLNA as compared to 47% (22/47) to 59% (19/32) for the guinea pig. The LLNA more accurately predicted strong human sensitizers than the guinea pig (maximum LLNA correct classification = 75% [18/24] using all data versus maximum guinea pig correct classification of 48% [11/23], both using 250 $\mu g/cm^2$ as the decision criteria in humans). In contrast, the guinea pig more accurately predicted the human weak sensitizers (maximum guinea pig correct classification = 89% [8/9] versus maximum LLNA correct classification 75% [6/8] both using sensitizer data only and 250 $\mu g/cm^2$ as the decision criteria in humans).

Test Method Reliability

An evaluation of the intralaboratory variability associated with 29 individual EC3 concentrations for isoeugenol (which ranged from 0.5% to 2.6%, when tested in a single laboratory) was conducted by Basketter and Cadby (2004). These data were used to support

435 the "often-mentioned perspective that the biological variation associated with the estimation
436 of EC3 values means that any particular EC3 can be halved or doubled" (Basketter et al.
437 2004). Additionally, Basketter et al. (2007) evaluated the interlaboratory reproducibility of
438 EC3 data for 17 sensitizers tested in at least two laboratories using the same vehicle. The
439 authors concluded that, although variability exists, it is less than an order of magnitude.

440 However, a number of analyses included in this BRD highlight the potential impact of the
441 vehicle used in the LLNA on EC3 values and potency classification. An evaluation of 31
442 substances in the NICEATM database for which data from tests in multiple vehicles were
443 available revealed that potency classifications would differ for 58% (18/31) of these
444 substances. For eight of these 31 substances (26%), a sensitizing or a nonsensitizing
445 classification could be assigned, depending on the vehicle used.

446 An evaluation of the variability for EC3 values calculated for the 31 substances for which
447 data from tests in multiple vehicles were available indicated that variability exceeded an
448 order of magnitude for 26% (8/31) of these substances.

449 In a separate analysis, a correlation was calculated for EC3 values from two vehicles (DMF
450 and acetone) when compared to the EC3 values for the same substance obtained with AOO
451 as the vehicle. These data indicate that EC3 values for substances tested in acetone and AOO
452 are similar while EC3 values for substances tested in DMF are consistently lower than those
453 obtained with AOO (i.e., the sensitizers are more potent in DMF than in AOO).

454 The proposal for using the LLNA for potency determinations does not impact its requirement
455 for using animals, or the number of animals that will be required. However, this application
456 could broaden the use of the LLNA protocol in place of the guinea pig tests and could
457 therefore further reduce the number of guinea pigs that are being used to assess skin
458 sensitization potency.

459 No changes to the LLNA protocol are being proposed and therefore, the transferability,
460 training requirements, and time and cost considerations for the LLNA remain unchanged to
461 the previous ICCVAM evaluation (ICCVAM 1999).

1.0 Introduction and Rationale for the Proposed Use of the Murine Local Lymph Node Assay (LLNA) for Potency Assessment

1.1 Introduction

1.1.1 Historical Background

In February 1998, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) received a submission from Drs. G. Frank Gerberick (Procter and Gamble, Cincinnati, United States [U.S.]), David Basketter (Unilever Safety and Environmental Assurance Centre, United Kingdom [U.K.]), and Ian Kimber (Syngenta Central Toxicology Laboratory, U.K.) to evaluate the murine local lymph node assay (LLNA) as an alternative to guinea pig tests (i.e., the Guinea Pig Maximization Test [GPMT], the Buehler Test [BT]) for assessing skin sensitization. The submission summarized the performance (i.e., relevance and reliability) of the LLNA as compared to the GPMT and the BT. An additional analysis was conducted by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to evaluate, where data existed, the comparative performance of the LLNA and the guinea pig tests against sensitization results obtained in humans. An independent expert peer review panel (Panel) was convened on September 17, 1998, to review the completeness of the submission, to determine whether the usefulness and limitations of the LLNA had been adequately described, and to decide whether its demonstrated performance supported recommending the LLNA as a stand-alone alternative to the GPMT and BT. The Panel also was asked to evaluate whether the LLNA offered advantages with regard to animal welfare considerations (i.e., refinement, reduction, or replacement[5]).

The Panel considered the performance of the LLNA to be similar to that of the GPMT and BT for identifying moderate to strong sensitizers. The Panel stated that the LLNA did not accurately predict all weak sensitizers, nor did it appear to adequately discriminate between strong skin irritants and skin sensitizers. The LLNA also produced false negative results with

[5] Refinement alternative is defined as a new or revised test method that refines procedures to lessen or eliminate pain or distress to animals, or enhances animal well-being; Reduction alternative is defined as a new or revised test method that reduces the number of animals required; Replacement alterative is defined as a new or revised test method that replaces animals with non-animal systems or one animal species with a phylogenetically lower one (e.g., a mammal is replaced with an invertebrate)(ICCVAM 1997).

some metals. It was recommended that future studies be conducted and/or workshops held to evaluate these issues. Furthermore, adequate data to support using the LLNA to test mixtures and aqueous solutions was not available and the number of pharmaceuticals tested was limited. Still, the Panel noted that when compared with the GPMT/BT methods, the LLNA appeared to provide equivalent prediction of sensitization potential, based on comparisons to available human data. In addition, the Panel concluded that the LLNA could be considered a refinement alternative to the GPMT and BT, because the pain and distress associated with the guinea pig methods could be virtually eliminated by using the LLNA. After consideration of the Panel report and public comments, ICCVAM recommended to U.S. Federal agencies that the LLNA could be used as an alternative to guinea pig tests for assessing sensitization potential. The Panel report and the ICCVAM recommendations (ICCVAM 1999) are available at the NICEATM/ICCVAM website[6]. The LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization (Organisation for Economic Co-operation and Development [OECD] Test Guideline [TG] 429 [OECD 2002]; International Standards Organization [ISO] 10993-10: Tests for Irritation and Sensitization [ISO 2002]; U.S. Environmental Protection Agency [EPA] Health Effect Testing Guidelines on Skin Sensitization [EPA 2003]).

1.1.2 Allergic Contact Dermatitis (ACD)

A skin sensitization reaction can result in allergic contact dermatitis (ACD). ACD is a frequent occupational health problem. In 2005, according to the U.S. Department of Labor Bureau of Labor Statistics, 980 cases of ACD involved time away from work. ACD develops in two phases, induction and elicitation. The induction phase occurs when a susceptible individual is exposed topically to a skin-sensitizing substance. Induction depends on the substance passing through the epidermis, where it forms a hapten complex with dermal proteins. The hapten complex is processed by the Langerhans cells, the resident antigen-presenting cells in the skin. The processed hapten complex then migrates to the draining lymph nodes. Antigen presentation to T-lymphocytes follows, which leads to the clonal expansion of these cells. At this point, the individual is sensitized to the substance (Basketter et al. 2003, Jowsey et al. 2006). Studies have shown that the magnitude of lymphocyte

[6] http://iccvam.niehs.nih.gov/methods/immunotox/llna_PeerPanel98.htm

proliferation correlates with the extent to which sensitization develops (Kimber and Dearman 1991, Kimber and Dearman 1996).

The elicitation phase occurs when the individual is again topically exposed to the same substance. As in the induction phase, the substance penetrates the epidermis, is processed by the Langerhans cells, and presented to circulating T-lymphocytes. The T-lymphocytes are then activated, which causes release of cytokines and other inflammatory mediators. This release produces a rapid dermal immune response that can result in ACD (ICCVAM 1999, Basketter et al. 2003, Jowsey et al. 2006).

1.1.3 Classification of Skin Sensitizers Based on Potency

Allergens are known to vary significantly in the potency with which they can induce skin sensitization. It has been suggested that skin sensitizing chemicals vary as much as 10,000-fold with respect to their relative sensitization potency (Kimber et al. 2003). For the purposes of this BRD, potency is defined as a function of the concentration of a substance that is required for either the induction or elicitation phases of a skin sensitization reaction. For induction, potency refers to the concentration of a substance needed to induce a sensitization response; the more potent the substance the smaller the quantity needed for induction. Likewise, for elicitation, potency refers to the concentration of a substance need to elicite a response in a previously sensitized individual; the more potent a substance, the smaller the quantity required for elicitation (ECETOC 2003).

The observed dose-response relationships that are associated with both induction and elicitation allow for thresholds of each phase to be determined (ECETOC 2003, Kimber et al. 2003). This includes thresholds for the level of exposure to a substance in a naïve individual below which sensitization will not likely be induced, or below which an elicitation reaction will not occur in a previously sensitized subject (Kimber at al. 1999). Although these thresholds are largely determined by the potency of a particular allergen, they vary due to vehicle effects and the extent of dermal exposure (Marzulli and Maibach 1976, Lea et al. 1999). Additionally, it has been suggested that:

- Induction thresholds for particular substances will be different from the elicitation threshold for the same substance (i.e., in general, higher levels are

546 needed for induction in a naïve individual than for elicitation in a previously
547 sensitized individual)

548 • Inter-individual variability in thresholds for elicitation exists and is largely
549 attributed to the extent to which an individual has been previously exposed
550 (Kimber et al. 1999, Basketter et al. 2003, ECETOC 2003).

551 Most authorities do not currently regulate products based on sensitization potency, instead
552 classifying them simply as "yes/no" designations. The U.S. Consumer Product Safety
553 Commission (CPSC), under the Federal Hazardous Substances Act (FHSA), currently
554 regulates products that are considered to be "strong" sensitizers following a weight of
555 evidence approach, taking into consideration frequency of occurrence and severity of
556 reaction. An OECD Task Force on Harmonization of Classification and Labeling, when
557 harmonizing existing hazard classification systems, originally suggested that differentiation
558 of skin sensitizers based on relative potency was not feasible because of the lack of
559 internationally accepted animal tests that could serve this purpose. Based on discussions
560 between CPSC and the international community during international workshops that have
561 recently been convened (see **Section 1.1.4**) a scheme has been proposed to subdivide
562 sensitization into two categories for purposes of hazard classification, based on a weight of
563 evidence evaluation combined with numerical guidance values for LLNA, guinea pig, and/or
564 human results (**Table 1-1**).

565 Kimber et al. (2003) have proposed a four-level classification scheme for potency based on a
566 log scale of EC3 values (**Table 1-2**). **Appendix A** contains a document provided by
567 Basketter et al. for consideration by ICCVAM and the European Centre for the Validation of
568 Alternative Methods (ECVAM) during their evaluations of the LLNA for potency
569 determinations. Similarly, a four-level classification scheme for assessing skin sensitization
570 potency has been proposed by the ECETOC Task Force on Contact Sensitization (see **Table**
571 **1-3**) (ECETOC 2003). However, in this evaluation, the ability of the LLNA to be used as a
572 stand-alone assay for determining sensitization potency is based on the proposed two-level
573 classification scheme (**Table 1-1**). Therefore, these other classification schemes are provided
574 for reference only.

575

575 **Table 1-1 Proposed Classification Categories for Sensitization**

Category	Criteria for Classification	LLNA EC3[2]	Human Threshold[1, 2]	GPMT Response[1]	BT Response
Category 1 (strong sensitizer)	EITHER: • 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	Option LLNA-A ≤ 1% Option LLNA-B ≤ 2%	Option Human-A: < 250 μg/cm^2 Option Human-B; < 500 μg/cm^2	EITHER: ≥ 60% responders at > 0.1% to ≤ 1.0% intradermal induction dose OR ≥ 30-60% responders at ≤ 0.1% intradermal induction dose	EITHER: ≥ 60% responders at > 0.2% to ≤ 20% topical induction dose OR ≥ 15% responders at ≤ 0.2% topical induction dose
Category 2 ([weak] sensitizer)[3]	EITHER: • 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	Option LLNA-A > 1% Option LLNA-B > 2%	Option Human-A: ≥ 250 μg/cm^2 Option Human-B; ≥ 500 μg/cm^2	EITHER: ≥ 30% responders at > 1.0% intradermal induction dose OR ≥ 30-60% responders at > 0.1% to ≤ 1.0% intradermal induction dose	EITHER: ≥ 15% to < 60% responders at > 0.2% to 20% topical induction dose OR ≥ 15% responders at > 20% topical induction dose

576 Abbreviations: BT = Buehler test; GPMT = Guinea pig maximization test; LLNA = Local Lymph Node Assay
577 [1]Proposed thresholds that are being considered; the expectation is that in the final version of this scheme, only
578 one LLNA EC3 and one human threshold value would be included.
579 [2]Human maximization test or human repeat-insult patch test induction threshold.
580 [3]For the purposes of this document, this category is being considered as "weak" sensitizers to clearly distinguish
581 it from "strong" sensitizers (i.e., Category 1).

582

582 **Table 1-2 Potency Categorization of Skin Sensitizers Based on LLNA EC3 Values[1]**

Potency Category	EC3 (%)[2]
Extreme	< 0.1
Strong	≥ 0.1 to < 1.0
Moderate	≥ 1.0 to < 10
Weak	≥ 10 to ≤ 100

583 [1]As proposed by Kimber et al. (2003).
584 [2]The LLNA EC3 value is the estimated concentration of a substance needed to
585 produce a stimulation index of three.

586 1.1.4 Use of the LLNA as a Stand-Alone Method for Potency Determinations

587 Traditional regulatory test methods for skin sensitization (i.e., GPMT, BT, LLNA) have
588 focused on "yes/no" determinations of sensitization hazard. In recent years, the LLNA has
589 been proposed as an effective method for determining sensitization potency because of the
590 dose response information that is generated. This concept was originally suggested by
591 Kimber and Basketter (1997) and was based on their characterization of the large difference
592 in LLNA threshold response between 2,4-dinitrochlorobenzene (DNCB) and hexyl cinnamic
593 aldehyde (HCA). A number of studies have been conducted in an attempt to support the use
594 of the LLNA for this purpose (see **Section 9.0** for the review articles on this topic).

595 **Table 1-3 Proposed Skin Sensitization Potency Categories Based on**
596 **Guinea Pig Data[1]**

Induction Concentration (%)	GPMT Incidence (%)		BT Incidence (%)	
	< 30 to < 60	≥ 60	< 15 to < 60	≥ 60
< 0.1	Strong	Extreme	Strong	Extreme
≥ 0.1 - < 1.0	Moderate	Strong	Moderate	Strong
≥ 1.0 - < 10.0	Weak	Moderate	Weak	Moderate
≥ 10 - ≤ 100	Weak	Weak	Weak	Weak

597 [1]Proposed by an ECETOC Task Force on Contact Sensitization (ECETOC 2003).

598 However, two international workshops on skin sensitization have stated that the LLNA has
599 yet to be adequately validated for classifying sensitizers according to potency. In July 2005,
600 the CPSC Sensitizer Scientific Panel recommended that a determination of risk be based on a

weight-of-evidence approach, utilizing all validated methods available, and that the LLNA alone was not adequate for this purpose. In October 2006, participants at the World Health Organization International Programme on Chemical Safety (IPCS) Workshop on Skin Sensitization in Chemical Risk Assessment concluded that the use of the LLNA for potency categorization of skin sensitization needs to be validated (WHO 2007).

1.1.5 U.S. Consumer Product Safety Commission (CPSC) Nomination

On January 10, 2007, CPSC formally nominated several activities related to the LLNA for evaluation by ICCVAM and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). One of the nominated activities was an assessment of the validation status of the LLNA as a stand-alone assay for determining the potency (including severity) of skin sensitizers for classification purposes. ICCVAM unanimously agreed that the nominated activity should have a high priority for evaluation. ICCVAM's advisory committee, the Scientific Advisory Committee on Alternative Toxicological Methods, also recommended that the nominated activity be undertaken, with a high priority. In response, ICCVAM directed its Immunotoxicity Working Group (IWG) to work with NICEATM in evaluating the validation status of the LLNA as a stand-alone assay for determining the potency of skin sensitizers for classification purposes. Both the European Centre for the Validation of Alternative Methods (ECVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM) have liaisons to the IWG to provide input during this evaluation.

1.1.6 Results of Peer Reviews on the Use of LLNA to Assess Sensitization Potency

A number of recent workshops have reviewed, among other tasks, the use of the LLNA to assess sensitization potency. These include the CPSC Sensitizer Scientific Panel, held July 2005 (CPSC 2005), the IPCS International Workshop on Skin Sensitization in Chemical Risk Assessment, held October 2006 (WHO 2007), and an OECD Expert Group on Sensitization that met February 2007. In each case, the participants concluded that the LLNA requires additional validation prior to being used as a stand-alone assay for skin sensitization potency determinations. To date, there have been no formal peer reviews of the use of the LLNA to assess the skin sensitization potency of test substances.

1.2 Validation of the LLNA for Skin Sensitization Potential

The ICCVAM Authorization Act (Sec. 4(c)) mandates that "[e]ach Federal Agency … shall ensure that any new or revised … test method … is determined to be valid for its proposed use prior to requiring, recommending, or encouraging [its use]." (ICCVAM 2000).

Validation is the process by which the reliability and relevance of an assay for a specific purpose are established, relevance is the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 1997). Reliability is defined as the reproducibility of a test method within and among laboratories. Reliability should be assessed by use of the test method with a diverse set of substances that are representative of both the types of chemical and product classes expected to be tested and the range of responses that needs to be identified. This validation process is intended to provide data and information to allow U.S. Federal agencies to develop guidance on the use of test methods in evaluating the skin sensitization potential of substances.

The first stage in this evaluation is the preparation of a Background Review Document (BRD) that provides a comprehensive review of a test method, including its mechanistic basis, proposed uses, data quality, and performance characteristics (i.e., relevance and reliability) (ICCVAM 1997). This BRD summarizes the available information on the use of the LLNA to assess the skin sensitization potential of a chemical. This BRD will also aid in day, yearidentifying any additional studies that should be considered during future development and validation activities.

1.3 Selection of Citations for the BRD

The test method data summarized in this BRD are based on information obtained from the peer-reviewed scientific literature identified through online searches via PubMed and SCOPUS, through citations in publications, and in response to a *Federal Register* (*FR*) notice requesting LLNA, guinea pig, and/or human skin sensitization data and experience (Vol. 72, No. 95, pp. 27815-27817[7]). The NICEATM database includes 345 published or unpublished references relevant to this evaluation. Key words used in the online searches for this evaluation were "LLNA" OR "Local Lymph Node" OR "Local lymph node" OR "local

[7] available at http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf

lymph node") AND (potency)"; the last comprehensive search was completed on January 15, 2008.

2.0 Test Method Protocol Components

The ICCVAM recommended LLNA protocol provides a detailed description of the conduct of the assay (ICCVAM 1999, Dean et. al. 2001). A test-substance-induced increase in lymphocyte proliferation in the draining lymph nodes of the ear, the site of test substance application, is used in the LLNA to identify chemical sensitizers. Mice are injected with radiolabeled compound (^{3}H-thymidine or ^{125}I-iodeoxyuridine), which is incorporated into the DNA of proliferating cells. The Stimulation Index (SI), which is the ratio of incorporated radioactivity in the auricular lymph nodes of treated versus control mice, is used to assess the sensitizing potential of the test substance. An SI of three or greater is used to classify a test substance as a skin-sensitizing agent. The estimated concentration needed to produce an SI of 3.0 (i.e., the EC3) is used as the metric for predicting sensitizaton potency using the LLNA.

The LLNA procedure described by ICCVAM (1999, Dean et al. 2001) and the EPA Health Effects Test Guidelines (EPA 2003) is identical but differ from the protocol described in OECD TG 429 (OECD 2002) in that they require the use of a concurrent positive control, the testing of five animals per dose group, and the collection and analysis of individual rather than pooled animal data. These differences are highlighted in **Appendix A** of the draft ICCVAM LLNA Performance Standards[8].

The method for determining the EC3 is a simple linear interpolation of the points in the dose response curve that lie immediately above and below an SI of 3, the classification threshold for sensitizers in the LLNA. This method was chosen from an evaluation of a variety of statistical approaches to derive EC3 values from LLNA dose-response data (Basketter et al. 1999c). When there are no data points that fall below an SI value of 3, a more complex log-linear extrapolation may be applied as described in Ryan et al. (2007) in which the two lowest test concentrations from the dose response curve are used.

[8] available at http://iccvam.niehs.nih.gov/methods/immunotox/llna_PeerPanel.htm

3.0 Substances Used for Validation of the LLNA for Potency Determinations

No additional LLNA studies were conducted for this evaluation; rather, data from available LLNA studies were evaluated retrospectively. Data were obtained from 58 different sources; these included published reports as well as unpublished data submitted to NICEATM in response to a *FR* notice (Vol. 72, No. 95, pp. 27815-27817[9]) requesting LLNA, guinea pig, and human sensitization study data. To be considered in this evaluation, LLNA data needed to be generated using the ICCVAM protocol (ICCVAM 1999, Dean et al. 2001, EPA 2003) or the protocol described in OECD TG 429.

The information included in this BRD is based on a retrospective review of LLNA, GP and human data derived from a database of over 500 substances, 170 of which have comparative LLNA, GP, and/or human data. Among these 170 substances, there are 112 substances with comparative human data (97 sensitizers, 15 non-sensitizers), 105 substances with comparative guinea pig data (52 sensitizers, 53 non-sensitizers), and 47 substances with comparative human and guinea pig data (34 sensitizers, 13 non-sensitizers) (**Appendix B**). **Appendix C** provides information on the physicochemical properties (e.g., octanol water partition coefficient), Chemical Abstracts Service Registry Number (CASRN), and chemical class for each substance tested. This information was obtained from the published reports, submitted data, or through online literature searches.

[9] available at http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf

4.0 Comparative *In Vivo* Reference Data

The reference data for this evaluation were human clinical studies (i.e., the human maximization test [HMT], the human repeat-insult patch test [HRIPT]) and the currently accepted guinea pig test methods for skin sensitization (i.e., the GPMT, the BT). National and international test guidelines and/or standardized protocols are available for these human (Stotts et al. 1980, Gerberick et al. 2000) and guinea pig (OECD 1992, EPA 2003) test methods.

Sensitization potency in humans was identified as the threshold concentration inducing a sensitizing response in either the HMT or HRIPT. For the purposes of this evaluation, the threshold for induction of skin sensitization in humans was considered to be the no observed effect level (NOEL, expressed as $\mu g/cm^2$) or, in the absence of a NOEL, the lowest observed effect level (LOEL, expressed as $\mu g/cm^2$) divided by a factor of 10 (CPSC 1992). Guinea pig potency, as determined from either the BT or the GPMT, was based on the percentage of responding guinea pigs and their associated induction concentration for each substance tested (see **Table 1-1**).

Ongoing efforts are being made by NICEATM to obtain the original records and/or reports for the human and guinea pig reference data used in this evaluation. These original records/reports have not yet been obtained. Ideally, all animal data supporting the validity of a test method should be obtained and reported from studies conducted in accordance with Good Laboratory Practice (GLP) guidelines, which are internationally recognized principles designed to produce high-quality laboratory records (OECD 1998; EPA 2006a, 2006b; FDA 2007a). The corresponding guidelines for human studies are Good Clinical Practices (GCP) (ICH 1996). Both the GLP and GCP guidelines provide an internationally standardized procedure for study conduct, reporting requirements, archival of study data and records, and information about the test protocol, in order to ensure the integrity, reliability, and accountability of a study.

The extent to which the human or guinea pig studies were compliant with GCP or GLP guidelines, respectively, is based on the information provided in published and submitted reports. Information on compliance with GLP guidelines was available for data obtained from GP studies submitted by E. Debruyne (Bayer CropScience SA) and P. Botham (ECPA).

733 None of the published references from which GP or human data were obtained have GCP or
734 GLP information specified.

5.0 LLNA Data and Results

5.1 Description of the LLNA Test Method Protocol Used to Generate Data

The studies included in this evaluation were reportedly conducted according to the recommended ICCVAM protocol (ICCVAM 1999, Dean et al. 2001) or following OECD TG 429 (OECD 2002). Where the OECD TG 429 was the reference protocol, specifics on the number of animals per dose group tested, whether or not lymph nodes were pooled within dose groups, and/or whether a concurrent positive control was used were generally not available. Where needed, this information has been requested, but all of these requests have not yet been answered.

5.2 Availability of Copies of Original LLNA Data Used to Evaluate Accuracy and Reliability

Copies of original data for the LLNA studies considered during the ICCVAM (1999) evaluation were made available previously to NICEATM. However, availability of the original data has not yet been determined for all of the other LLNA studies considered during this evaluation but requests to determine availability have been made.

5.3 Description of the Statistical Approach Used to Evaluate the Resulting Data

Section 2.0 describes the derivation of the SI and the estimated concentration needed to produce an SI =3 (i.e., the EC3). The EC3 (typically expressed as %) is the metric used to evaluate the ability of the LLNA to predict sensitization potency.

To evaluate the correlation between EC3 values and human threshold values (expressed in $\mu g/cm^2$), EC3 values (in percent) were converted to $\mu g/cm^2$ by multiplying by a factor of 250 (based on an exposed area of 1 cm^2 and a dosing volume of 25 μL in the LLNA) (Griem et al. 2003). For all other comparisons between LLNA and human or guinea pig test results, the EC3 was expressed in its traditional units (%).

5.4 Summary of Results

The data used for this evaluation were obtained from 59 sources (**Appendix C**). Where available, the information extracted for each substance includes its name, CASRN,

physicochemical properties (e.g., octanol water partition coefficient), and chemical class[10] (**Appendix B**). LLNA data for each substance (i.e., stimulation indices for each concentration tested and calculated EC3 values, where applicable), along with the NOEL or LOEL and the threshold dose for sensitizers in humans, where available, and the induction concentration, incidence of sensitized animals, and sensitization potency classification in guinea pigs, where available, and the corresponding data source are provided in **Appendix B**. Other than the information provided with the submitted data, no additional attempt was made to identify the source or purity of the test substance.

5.5 Use of Coded Chemicals

Coding of substances to avoid potential scoring bias did not occur for any of the substances tested in the LLNA and evaluated by ICCVAM in the original evaluation (ICCVAM 1999) or for any of the more recently obtained studies used in this evaluation.

5.6 Lot-to-Lot Consistency of Test Substances

Ideally, a single lot of each substance is used during the validation of a test method. In situations where multiple lots of a substance must be used, lot-to-lot consistency must be evaluated to ensure the consistency of the substance evaluated over the course of the study. There was no available information in any of the reports included in this evaluation with which to assess lot-to-lot consistency.

5.7 Availability of Data for External Audit

The data for the substances tested in the LLNA and previously evaluated by ICCVAM (1999) were audited during that evaluation. However, although requested, the availability of the original data for the other LLNA studies included in this evaluation has not yet been determined.

[10] Chemical classes were assigned by NICEATM based on the classification of the National Library of Medicine's Medical Subject Heading (available at http://www.nlm.nih.gov/mesh/meshhome.html).

6.0 Test Method Accuracy

In this section, the ability of the LLNA to accurately predict skin sensitization potency in humans, based on data generated by the HMT and HRIPT, is evaluated. Also, because sensitization potency in guinea pig tests are used as part of a weight-of-evidence approach for estimating sensitization potency in humans (see **Section 1.1.3**), the ability of the LLNA to accurately predict sensitization potency as determined in the GPMT and BT is evaluated. Finally, the comparative ability of the LLNA and guinea pig tests to predict skin sensitization potency in humans is examined for substances tested in all three species.

6.1 Ability of the LLNA to Predict Skin Sensitization Potency in Humans

Two approaches were used to evaluate the ability of the LLNA to predict sensitization potency in humans. In the first approach, for each substance classified as a sensitizer in both the LLNA and in humans, the LLNA EC3 concentration (expressed in $\mu g/cm^2$ and not as a percent) was correlated against the human threshold response (i.e., either the NOEL or LOEL/10, expressed in $\mu g/cm^2$). In the second approach, using the same LLNA/human sensitizers used in the first approach, the human sensitizers were classified into strong or weak based on using either of two proposed decision criteria (strong sensitizers < 250 or < 500 $\mu g/cm^2$, see **Table 1-1**). Next, the optimal EC3 value that maximized obtaining the correct skin sensitization calls for strong and weak sensitizers (using one or the other proposed decision criterion) was pragmatically determined and the correct classification rate as well as the over- and under-classification rates calculated. In a variant of the second approach, substances that were classified in the LLNA as false positives (i.e., sensitizers in the LLNA but non-sensitizers in humans), false negatives (i.e., non-sensitizer in the LLNA but sensitizers in human tests), and non-sensitizers in both the LLNA and in human tests were included, the optimal EC3 values were re-calculated, and then the correct classification rate as well as the over- and under-classification rates re-calculated for each sensitization category (strong sensitizer, weak sensitizer, non-sensitizer).

In these analyses, for substances that had more than one EC3 or human threshold value, two methods for arriving at a single EC3 or threshold value were used. First, the most potent (i.e., the lowest) LLNA EC3 or human threshold concentration was used. Second, the geometric mean of all LLNA EC3 or human threshold concentrations was used. In the latter case, the

HMT and the HRIPT were not classified as repeat tests for the same substance (i.e., geometric means were calculated only for repeat HMT or repeat HRIPT). The impact of variability in the EC3 is discussed in **Section 7.0**.

6.1.1 LLNA EC3 versus Human Threshold Concentration Regression analysis

In the current NICEATM LLNA database, there are 112 substances with both LLNA and human data, 81 of which are classified as sensitizers in both the LLNA and in the HMT and/or the HRIPT (**Appendix B**). The distribution of these sensitizers in the LLNA by the number of studies conducted and the solvent used is provided in **Table 6-1**.

Table 6-1 Distribution of LLNA/Human Sensitizers by the Number of LLNA Studies Conducted and the Solvent Used

Multiplicity of LLNA Studies for 81 Sensitizers					
1	**2**	**3**	**4**	**5**	**≥6**
13 (16%)	25 (31%)	12 (15%)	7 (9%)	2 (2%)	22 (27%)
Number of Sensitizers Tested in Each Solvent					
AOO	**ACE**	**DMF**	**DMSO**	**EtOH-DEP**	**Other[1]**
27 (33%)	1 (1%)	7 (9%)	4 (5%)	7 (9%)	35 (43%)

Abbreviations: AOO = acetone:olive oil; EtOH-DEP = 1:3 Ethanol:diethyl phthalate; LLNA = Local Lymph Node Assay; Data based on x substances classified as sensitizers in both the LLNA and in the HMT and/or HRIPT.

[1]Includes 1 substance tested in 80% EtOH and 35 substances for which vehicle information is not known.

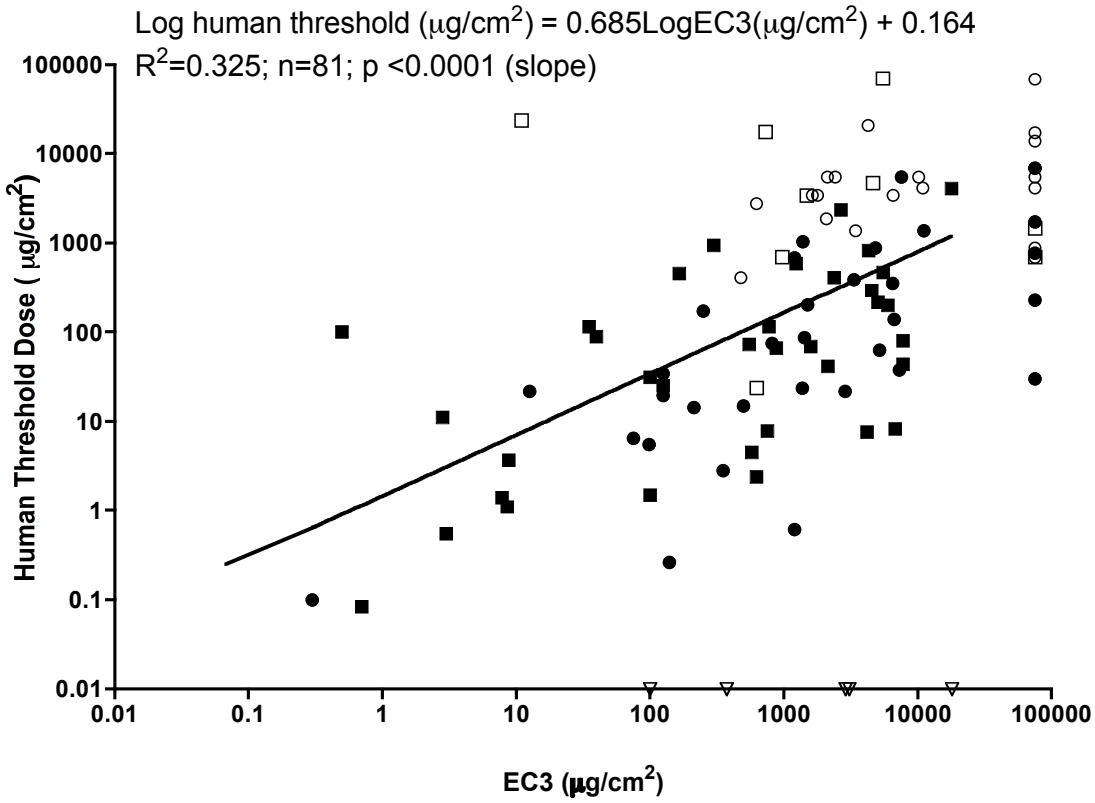
A regression analysis of LLNA EC3 versus human threshold values for these 81 LLNA/human sensitizers, both scaled in $\mu g/cm^2$ and based on log transformed data, indicated a positive correlation with an R^2 value of 0.325 ($P<0.0001$) (**Figure 6-1**) and 0.405 ($P<0.0001$) (**Figure 6-2**) when either the most potent LLNA EC3 and human threshold values or the geometric mean for multiple test results was used, respectively. The resulting regression equations are provided in **Table 6-2**. Based on an analysis of slope and intercept, the two regressions are not significantly different ($p = 0.125$ for slope and $p=0.620$ for intercept). However, based on the higher R^2 value (0.405) achieved when geometric means of multiply tested substances were calculated, this approach was carried forward through the remainder of the performance analyses.

Table 6-2 compares the correlation results obtained using the the NICEATM database of human skin sensitization data (**Appendix B**):

- when LLNA EC3 data were correlated against HMT threshold data, HMT NOEL data only, or HMT LOEL/10 data only
- when LLNA EC3 data were correlated against HRIPT threshold data, HRIPT NOEL data only, or HRIPT LOEL/10 data only
- for sensitizers tested in the LLNA using acetone:olive oil (AOO), the most common solvent used, when correlated against human threshold data

For comparative purposes, linear regression data for LLNA EC3 values versus various sets of human threshold data as published previously (Griem et al. 2003, Schneider and Akkam 2004, Basketter et al. **Appendix A**) are provided also in **Table 6-2**. All of the sensitizers in these data sets are included in the NICEATM database (**Appendix B**).

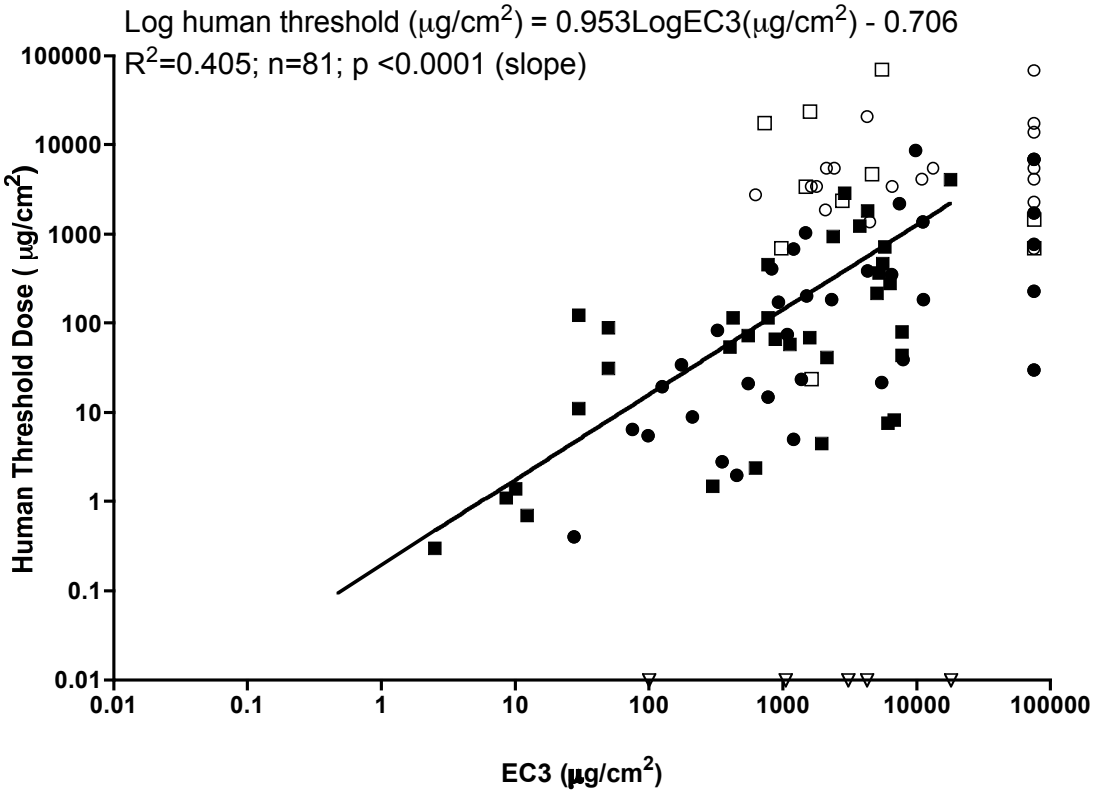
852 **Figure 6-1 LLNA EC3 versus Human Threshold Concentrations for LLNA/Human**
853 **Skin Sensitizers, Based on Using the Most Potent Concentration for a**
854 **LLNA EC3 or for a Human Threshold Response**



855

856 The solid line shows the regression line for the LLNA EC3 concentration versus the corresponding
857 human threshold concentration (both in μg/cm^2) for 81 sensitizers detected in the HMT (circles) and
858 the HRIPT (squares). For LLNA and human data, the lowest EC3 or threshold value, respectively,
859 was used if a substance had been tested more than once. Human NOEL data are indicated with open
860 circles/squares; human LOEL/10 data are indicated with closed circles/squares. Not included in the
861 regression analysis but provided to indicate the range of values obtained are an additional 16
862 substances (symbols aligned vertically along the 100000 EC3 value) that were negative in the LLNA
863 (SI < 3.0), but positive for sensitization in humans, and an additional 5 substances (inverted triangles)
864 that were sensitizers in the LLNA but not sensitizers in humans.

865 **Figure 6-2 LLNA EC3 versus Human Threshold Concentrations for LLNA/Human**
866 **Skin Sensitizers, Based on Using the Geometric Mean Concentration for**
867 **an LLNA EC3 or for a Human Threshold Response for Substances**
868 **Multiply Tested**



869

870 The solid line shows the regression line for the EC3 concentration versus the corresponding human
871 threshold concentration (both in μg/cm^2) for 81 sensitizers in the HMT (circles) and the HRIPT
872 (squares). For LLNA and human data, the geometric mean EC3 or threshold value, respectively, was
873 used if a substance had been tested more than once using the same test method. Human NOEL data
874 are indicated with open circles/squares; human LOEL/10 data are indicated with closed
875 circles/squares. Not included in the regression analysis but provided to indicate the range of values
876 obtained are an additional 16 substances (symbols aligned vertically along the 100000 EC3 value)
877 that were negative in the LLNA (SI < 3.0), but positive for sensitization in humans, and an additional
878 5 substances (inverted triangles) that were sensitizers in the LLNA but not sensitizers in humans.
879

880

880 **Table 6-2 Linear Regressions obtained for LLNA EC3 values versus Human**
881 **Threshold Values**

Comparison	N	Regression Coefficient (μg/cm^2)	Y-intercept	R^2	p-value
LLNA EC3 data vs human threshold data for sensitizers using most potent value	81	0.685	0.164	0.325	<0.0001
LLNA EC3 data vs human threshold data for sensitizers using geometric mean value for multiply tested substances	81	0.953	-0.706	0.405	<0.0001
Sensitizers tested in the LLNA using AOO vs geo mean human threshold data	34	1.010	-1.016	0.466	<0.0001
LLNA EC3 data vs HMT threshold data	39	1.261	-1.670	0.495	<0.0001
LLNA EC3 data vs HMT NOEL data only	11	0.174	2.994	0.050	0.510
LLNA EC3 data vs HMT LOEL/10 data only	28	1.045	-1.354	0.528	<0.0001
LLNA EC3 data vs HRIPT threshold data	42	0.809	-0.289	0.360	<0.0001
LLNA EC3 data vs HRIPT NOEL data only	8	0.917	0.526	0.069	0.529
LLNA EC3 data vs HRIPT LOEL/10 data only	34	0.693	-0.233	0.438	<0.0001
Griem et al. (2003) reported EC3 data vs. HMT/HRIPT LOEL data	23	0.783	0.682	0.657	<0.0001
Griem et al. (2003) reported EC3 data vs. HMT/HRIPT NOEL data	18	0.959	0.111	0.776	<0.0001
Griem et al. (2003) reported EC3 data vs. HMT/HRIPT LOEL and NOEL data	41	0.854	0.466	0.711	<0.0001
Schneider and Akkan (2004) reported EC3 data vs HRIPT thresholds	24	0.765	0.818	0.641	<0.0001
Schneider and Akkan (2004) reported EC3 data vs HMT thresholds	38	0.586	0.936	0.419	<0.0001
Basketter et al. (2005) reported EC3 data vs HRIPT threshold data	25	1.1	-0.53	0.72	<0.0001
Basketter et al. submission (**Appendix A**) reported EC3 data vs HRIPT/HMT threshold data	66	0.896	0.211	0.519	<0.0001

882 Abbreviations: AOO = Acetone: olive oil; EC3 = Estimated concentration needed to produce a stimulation
883 index of three; HMT = Human Maximization Test; HRIPT = Human Repeat-Insult Patch Test; LLNA = Local
884 Lymph Node Assay; LOEL = Lowest observed effect level; LOEL/10 = LOEL divided by a safety factor of 10;
885 NOEL = No observed effect level; R^2 = correlation coefficient.
886 *Basketter et al. (**Appendix A**) did not apply a safety factor of 10 to LOELs as was done by ICCVAM. This
887 additional analysis was conducted by ICCVAM as a point of reference..

888

As demonstrated in **Table 6-2**, there are differences in R^2 values (which is a measure of association between two sets of values) among the various analyses. These differences presumably reflect differences in the number of substances with both LLNA EC3 and human sensitization threshold data, which human test is considered (i.e., HMT and/or HRIPT), how NOEL and/or LOEL values are used (e.g., LOEL or LOEL/10), and how data for substances tested multiply times are collapsed into a single value. For example, the R^2 value generated with the overall NICEATM database (n=81) increased from 0.325 to 0.405 when geometric mean threshold values were used for multiply tested sensitizers instead of the most potent value. Additionally, when only the data for substances tested in the LLNA with the recommended vehicle (AOO) used, the R^2 value further increased to 0.466.

The R^2 values generated from data published in Griem et al. (2003), Schneider and Akkan (2004), and Basketter et al. (2005) are higher than the R^2 values from the NICEATM database. There may be several reasons for this apparent discordance including the fact that:

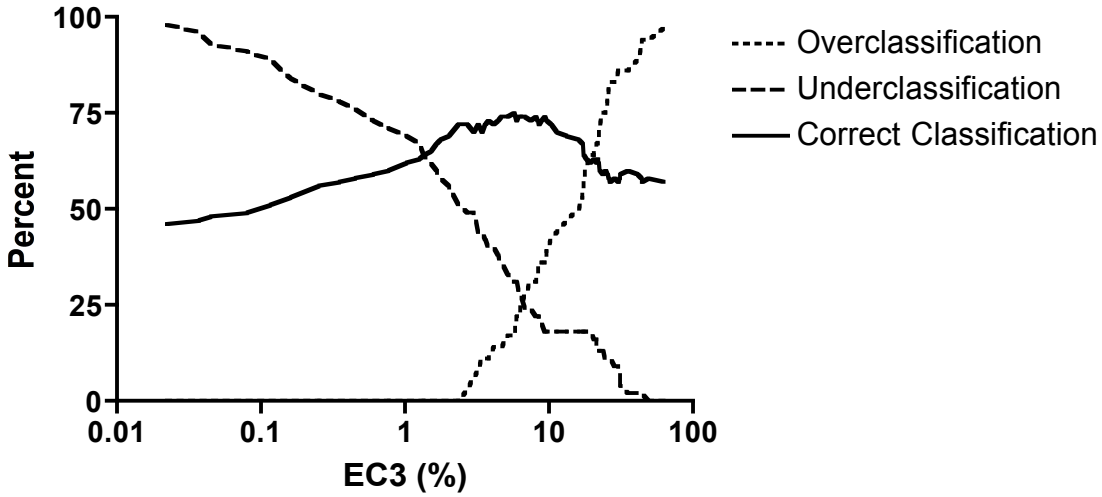
- The NICEATM database represents a larger set of substances (n=81) than the published datasets (n=24 to 41).
- Griem et al. (2003) and the ICCVAM analysis each included NOELs and LOELs from both HMT and HRIPT data. However, the ICCVAM used LOEL/10, while Griem et al. (2003) used LOEL.
- Schneider and Akkan (2004), Basketter et al. (2005), and Basketter et al. (**Appendix A**) used only HRIPT data, and they used LOEL instead of LOEL/10.
- The R^2 value generated from Basketter et al. (Appendix A), which used LOELs and NOELs from HRIPT studies only (n=66), was 0.519. However, when LOEL/10 and NOELs were used, the R^2 was reduced to 0.490.

These notable differences among the various databases should be taken into consideration when comparing the results in **Table 6-2**.

6.1.2 LLNA EC3 versus Human Threshold Concentration Analysis Based on Potency Classification

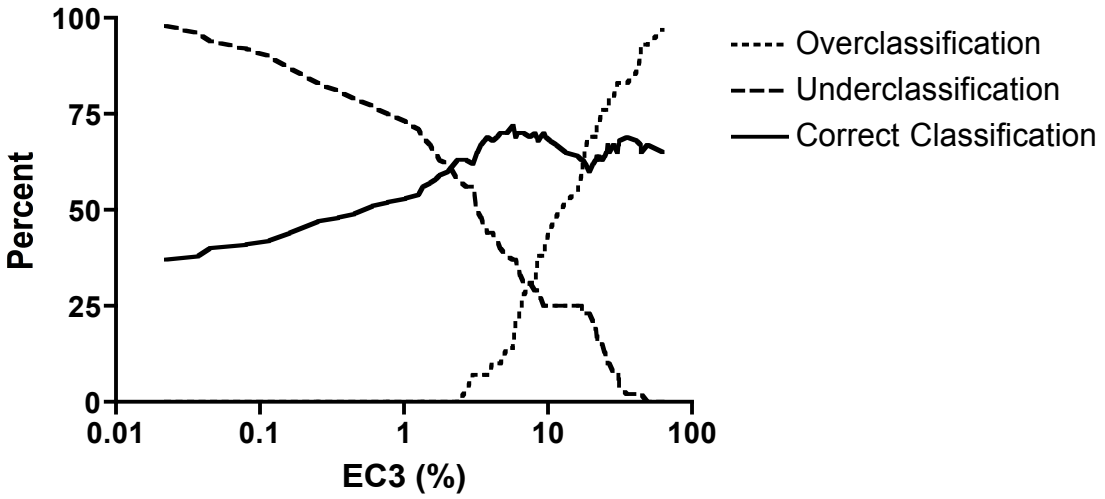
In this analysis, the ability of the LLNA EC3 value to correctly distinguish between weak and strong sensitizers in humans was evaluated using two different decision criteria for human threshold data (i.e., 250 and 500 μg/cm^2, see **Table 1-1**). In addition, two approaches were used to estimate the correct classification rate as well as the over- and under-classification rates based on the two proposed decision criteria for distinguishing between strong and weak sensitizers in humans. In the first approach, the classification analysis considered only those substances classified as sensitizers in both the LLNA and in humans based on the HMT and/or HRIPT. In the second approach, the analysis took into consideration those substances classified in the LLNA as false positives and false negatives against human skin sensitization data, as well as those classified as non-sensitizers in both the LLNA and in humans. Regardless of the approach, the first step was to determine the optimal LLNA EC3 value for distinguishing between strong and weak human sensitizers (i.e., the EC3 value that maximized the ability of the LLNA to correctly distinguish between sensitizers classified as strong or weak in humans. **Figures 6-3** and **6-4** show the relationship between the LLNA EC3 value and the overall correct human skin sensitization classification as well as the over- and under-classification rates, when the decision criteria for distinguishing between strong and weak sensitizers is 250 and 500 μg/cm^2, respectively. In this analysis, for multiply tested substances the geometric mean was calculated based on the improved (although not statistically significant) correlation obtained using the geometric mean values for LLNA and human results. For these data, the optimal EC3 values are 6.8% and 8.1% when 250 μg/cm^2 and 500 μg/cm^2, respectively, are used to distinguish between strong and weak skin sensitizers in humans. **Appendix D** contains the complete dataset for these analyses.

Figure 6-3 Correct Classification and Over- and Under-Classification Rates for the Ability of Different LLNA EC3 Values to Identify Strong and Weak Skin Sensitizers in Humans Based on a Human Threshold Concentration of 250 μg/cm^2, using Substances Classified in the LLNA and in Humans as Sensitizers



Analysis based on 81 substances identified as sensitizers in both the LLNA and in humans using the HMT and/or the HRIPT. For multiply tested substances, the most potent LLNA EC3 or human threshold value was used. In humans, sensitizers were classified as strong or weak if the threshold dose was < 250 μg/cm^2 or ≥250 μg/cm^2, respectively. EC3 = Estimated concentration needed to produce a stimulation index of three.

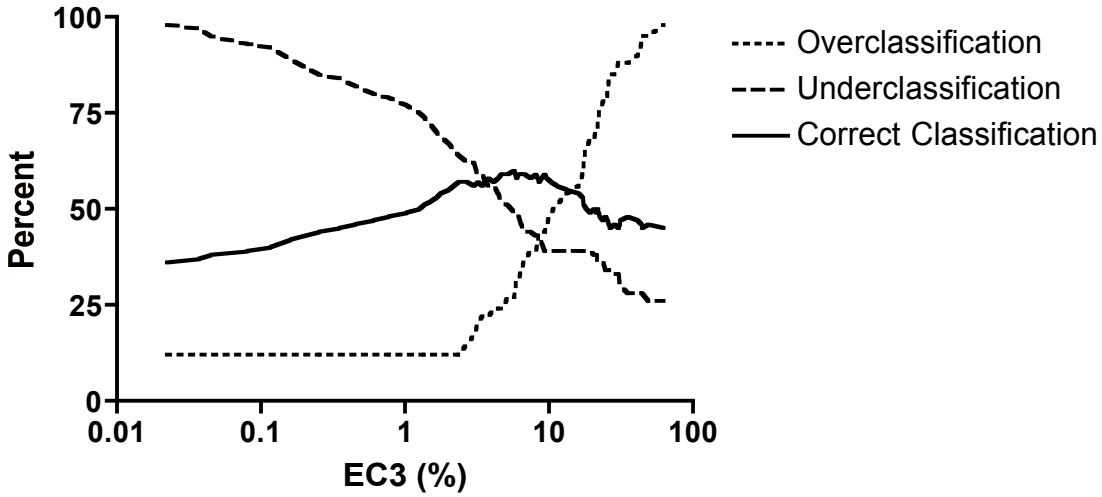
Figure 6-4 Correct Classification and Over- and Under-Classification Rates for the Ability of Different LLNA EC3 Values to Identify Strong and Weak Skin Sensitizers in Humans Based on a Human Threshold Concentration of 500 μg/cm^2, using Substances Classified in the LLNA and in Humans as Sensitizers



Analysis based on 81 substances identified as sensitizers in both the LLNA and in humans using the HMT and/or the HRIPT. For multiply tested substances, the most potent LLNA EC3 or human threshold value was used. In humans, sensitizers were classified as strong or weak if the threshold dose was < 500 μg/cm^2 or ≥ 500 μg/cm^2, respectively. EC3 = Estimated concentration needed to produce a stimulation index of three.

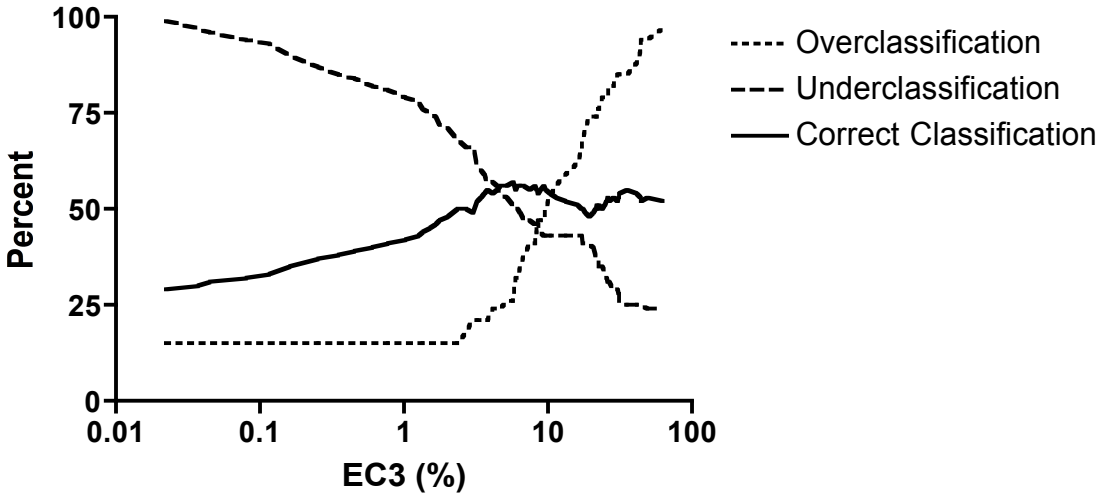
A second but similar analysis was conducted after including data for substances classified as false negatives and false positives in the LLNA compared to human results, as well as substances classified as non-sensitizers in both the LLNA and in humans. This increased the number of substances with comparative LLNA and human data from 81 to 102. The results of this analyses, using 250 and 500 μg/cm^2 as the decision criteria for distinguishing between strong and weak skin sensitizers in humans, are provided in **Figures 6-5** and **6-6**, respectively. For this dataset, the optimal LLNA EC3 was 9.35% when using either 250 or 500 μg/cm^2 as the decision criteria for distinguishing between strong and weak sensitizers in humans.

Figure 6-5 Correct Classification and Over- and Under-Classification Rates for the Ability of Different LLNA EC3 Values to Identify Strong and Weak Skin Sensitizers in Humans Based on a Human Threshold Concentration of 250 μg/cm^2, using Sensitizers, False Negatives, False Positives and Non-sensitizers



Analysis based on 102 substances identified as sensitizers in both the LLNA and in humans using the HMT and/or the HRIPT, as false positive or false negative in the LLNA compared to human results, and as non-sensitizers in both the LLNA and in humans. In humans, sensitizers were classified as strong or weak if the threshold dose was ≤250 μg/cm^2 or >250 μg/cm^2, respectively. EC3 = Estimated concentration needed to produce a stimulation index of three.

Figure 6-6 Correct Classification and Over- and Under-Classification Rates for the Ability of Different LLNA EC3 Values to Identify Strong and Weak Skin Sensitizers in Humans Based on a Human Threshold Concentration of 500 μg/cm^2, using Sensitizers, False Negatives, False Positives and Non-sensitizers



Analysis based on 112 substances identified as sensitizers in both the LLNA and in humans using the HMT and/or the HRIPT, as false positive or false negative in the LLNA compared to human results, and as non-sensitizers in both the LLNA and in humans. In humans, sensitizers were classified as strong or weak if the threshold dose was ≤500 μg/cm^2 or >500 μg/cm^2, respectively. EC3 = Estimated concentration needed to produce a stimulation index of three.

The correct classification and over- and under-classification rates obtained under different conditions using the optimal EC3 values associated with the two different decision criteria for distinguishing between strong and weak sensitizers in humans are provided in **Table 6-3**.

In the first analysis, which focused only on substances classified as sensitizers in both the LLNA and in humans, overclassification means that weak sensitizers are missclassified as strong while underclassification means that strong sensitizers are missclassified as weak. Using the optimal EC3 values identified when 250 or 500 μg/cm^2 was used as the decision criteria for distinguishing between sever and weak skin sensitizers in humans, the correct classification rate was 74% and 70% for 250 and 500 μg/cm^2, respectively, while the over- and under-classification rates ranged from 28% to 31% and 24% to 29%, respectively.

The second analysis included substances classified as sensitizers in both the LLNA and in humans as well as substances classified in the LLNA as false positives and false negatives compared to the human, and substances classified as non-sensitizers in both the LLNA and in

humans. In this analysis, overclassification means that nonsensitizers are misclassified as weak or strong sensitizers and weak sensitizers are misclassified as strong sensitizers. Likewise, underclassification means that strong sensitizers are misclassified as weak or nonsensitizers and weak sensitizers are misclassified as nonsensitizers. Using the optimal EC3 values identified when 250 and 500 $\mu g/cm^2$ was used as the decision criteria for distinguishing between sever and weak skin sensitizers in humans, the correct classification rate was 59% and 56% for 250 and 500 $\mu g/cm^2$, respectively, while the over- and under-classification rates ranged from 44% to 47% and 39% to 43%, respectively.

The minor differences in classification/misclassification rates between using 250 $\mu g/cm^2$ or 500 $\mu g/cm^2$ as the threshold value for distinguishing between strong and weak sensitizers in humans can be explained by the relative few substances (n = 7) that have threshold values that fall between these two criterion. Using a threshold of $\leq$500 $\mu g/cm^2$ to classify strong sensitizers would potentially classify more substances as strong sensitizers than the lower threshold of $\leq$250 $\mu g/cm^2$. A larger database is needed to determine which threshold value is optimum for classifying sensitizers as severe in humans.

Table 6-3 Correct Classification and Over- and Under-classification Rates when the Optimal LLNA EC3 Value is Used to Predict the Human Skin Sensitization Potency Classification[1]

Comparison	Overall Classification	Classification						
		Strong Sensitizer		Weak Sensitizer			Non-Sensitizer	
		Correct[2]	Under[2]	Over[2]	Correct[2]	Under[2]	Correct[2]	Over[2]
Sensitizers only: optimal LLNA EC3 (6.8%) vs. strong/weak sensitizers in humans using a decision criterion of 250 μg/cm^2	74% (60/81)	76% (34/45)	24% (11/45)	28% (10/36)	72% (26/36)	NA	NA	NA
Sensitizers only: optimal LLNA EC3 (8.1%) vs. strong/weak sensitizers in humans using a decision criterion of 500 μg/cm^2	70% (57/81)	71% (37/52)	29% (15/52)	31% (9/29)	69% (20/29)	NA	NA	NA
All data: optimal LLNA EC3 (9.4%) vs. strong/weak sensitizers and non-sensitizers in human using a decision criterion of 250 μg/cm^2 and taking into account LLNA false negative and false positive sensitizers compared to the human	62% (70/112)	79% (37/47)	21% (10/47)	26% (13/50)	46% (23/50)	28% (14/50)	67% (10/15)	33% (5/15)
All data: optimal LLNA EC3 (9.4%) vs. strong/weak sensitizers and non-sensitizers in human using a decision criterion of 500 μg/cm^2 and taking into account LLNA false negative and false positive sensitizers compared to the human	60% (67/112)	72% (39/54)	28% (15/54)	26% (11/43)	42% (18/43)	33% (14/43)	67% (10/15)	33% (5/15)

[1]Two human threshold values (250 and 500 μg/cm^2) are proposed for distinguishing between strong and weak sensitizers. The LLNA EC3 values used in this analysis are the ones that maximally reduce the over- and under-classification rates, based on the geometric mean LLNA EC3 or human threshold value for multiply tested substances, as determined in **Figures 6-1** through **6-4**.

[2]Overall Classification: The proportion of substances assigned to each hazard classification for skin sensitization potency (i.e., Strong, Weak, Non-Sensitizer). Correct**:** percentage of LLNA EC3 values that correctly predicted the human skin sensitization potency classification; Over**:** percentage of LLNA EC3 values that overclassified the human sensitization classification; Under**:** percentage of LLNA EC3 values that underclassified the human skin sensitization potency classification.

[3] The proportion on which the percentage calculation is based.

NA = Not applicable since only strong and weak sensitizers evaluated in the absence of non-sensitizers (i.e., Weak sensitizers can only be over-predicted and no non-sensitizers were evaluated)

6.2 Ability of the LLNA to Predict Skin Sensitization Potency in Guinea Pigs

As mentioned previously, sensitization potency in guinea pigs in the GPMT and the BT is used, as part of a weight-of-evidence approach, to classify substances as strong sensitizers in humans (see **Section 1.1.3**). Thus, it was deemed useful to evaluate the ability of the LLNA to agree with the sensitization potency classification assigned to a substance tested in the GPMT or the BT (see **Table 1-1**). Due to the categorical nature of the data collected in the guinea pig tests (i.e., incidence of sensitized animals at a particular test substance concentration), a regression analysis could not be conducted. However, substances detected as sensitizers in the GPMT or the BT could be assigned a severity classification (strong or weak) based on the decision criteria described in **Table 1-1**, and the ability of the LLNA EC3 (using an optimal value determined empirically) to correctly predict this classification evaluated.

In the current NICEATM LLNA database, there are 105 substances with both LLNA and guinea pig test data, 43 of which are classified as sensitizers both in the LLNA and in the guinea pig (**Appendix C**). The distribution of LLNA/guinea pig sensitizers by the number of LLNA studies conducted and the solvent use is provided in **Table 6-4**, while the distribution of corresponding sensitizers positive in the GPMT and/or the BT and the number of studies conducted are provided in **Table 6-5**.

Table 6-4 Distribution of LLNA/Guinea Pig Sensitizers by the Number of LLNA Studies Conducted and the Solvent Used

Multiplicity of LLNA Studies for 43 Sensitizers					
1	**2**	**3**	**4**	**5**	**≥6**
9 (17%)	14 (27%)	5 (10%)	1 (2%)	2 (4%)	21 (40%)
Number of Sensitizers Tested in Each Solvent					
AOO	**ACE**	**DMF**	**MEK**	**DMSO**	**Other[1]**
22 (42%)	1 (12%)	6 (12%)	1 (2%)	3 (6%)	19 (36%)

Abbreviations: AOO = acetone:olive oil; LLNA = Local Lymph Node Assay;
Data based on x substances classified as sensitizers in both the LLNA and in the GPMT and/or BT.
[1]Includes 3 substances tested in Pluronic L92, one each tested in 30% Ethanol and 1:3 Ethanol:diethyl phthalate, and 14 substances for which vehicle information is not known.

Table 6-5 Distribution of LLNA/Guinea Pig Sensitizers by Guinea Pig Test Method and the Number of Guinea Pig Studies Conducted

Test Method and Multiplicity of Studies								
GPMT			BT			GPMT + BT		
1	2	≥3	1	2	≥3	1	2	≥3
28	3	17	18	6	4	15	4	4

Abbreviations: GPMT = Guinea Pig Maximization Test; BT = Buehler Test.
Data based on 52. substances classified as sensitizers in both the LLNA and in the GPMT and/or BT.

Guinea pig sensitizers that were also sensitizers in the LLNA were classified into strong or weak based on using the proposed decision criteria in **Table 1-1**. Next, the optimal EC3 value that maximized the percentage of correct calls for strong and weak sensitizers was determined. **Figure 6-7** shows the relationship between the LLNA EC3 value and the correct guinea pig skin sensitization classification as well as the over- and under-classification rates, based on the GPMT and BT decision criteria for distinguishing between strong and weak sensitizers as described in **Table 1-1**. In this analysis, for multiply tested substances, the geometric mean LLNA EC3 value was used, while a weight-of-evidence evaluation was used to categorize the guinea pig test results. In this approach, test results from either GPMT or BT tests (i.e., as per the decision criteria in **Table 1-1**) were considered together when assigning an overall classification category according to Table 1-1.

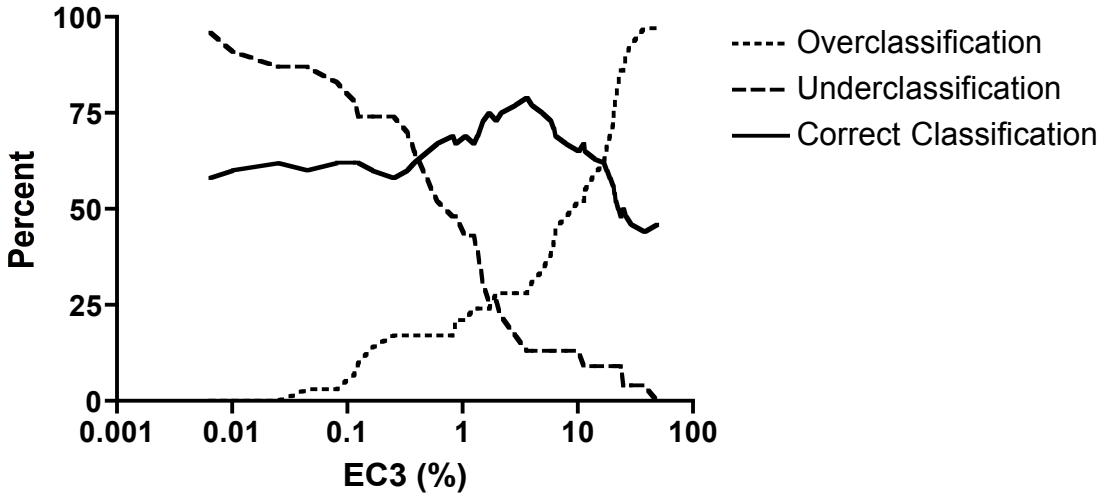
Consider the following example: Substance X has GPMT test results from two studies (one labeled as Category 1 based on GPMT results and one as Category 2) and BT results from one study (labeled as Category 1 based on BT results). Based on the two Category 1 studies vs. one Category 2 study, Chemical X would be labeled as Category 1.

Next, the correct classification rate as well as the over- and under-classification rates against the guinea pig results were calculated, using this optimal EC3 value. In a variant of this approach, substances that were classified in the LLNA as false positives (i.e., sensitizers in the LLNA but non-sensitizers in guinea pigs), false negatives (i.e., non-sensitizer in the LLNA but sensitizers in guinea pigs), and non-sensitizers in both the LLNA and in guinea pigs were included, the optimal EC3 value was re-calculated (**Figure 6-8**) and then the correct classification rate as well as the over- and under-classification rates re-calculated for each sensitization category (strong sensitizer, weak sensitizer, non-sensitizer). In these

various analyses, for substances that had more than one EC3 or guinea pig response, the geometric mean EC3 value and the weight of evidence GP classification category was used.. Although the data generated by the GPMT and the BT is categorical, using the weight of evidence categorization provided some measure of a mean response across multiple studies.

The results obtained from using the optimal EC3 values are provided in **Table 6-6**. In the first analysis, which focused only on substances classified as sensitizers in both the LLNA and in guinea pigs, overclassification means that weak sensitizers are missclassified as strong while underclassification means that strong sensitizers are missclassified as weak. Using the optimal EC3 value of 1.95 %, the correct classification rate was 73% (38/52), while the over- and under-classification rates were 28% (8/29) and 26% (6/23), respectively.

Figure 6-7 Correct Classification and Over- and Under-Classification Rates for the Ability of Different LLNA EC3 Values to Identify Strong and Weak Skin Sensitizers in Guinea Pigs Based on Criteria in Table 1-1, using only Substances Detected as Sensitizers in Both the LLNA and the GPMT/BT



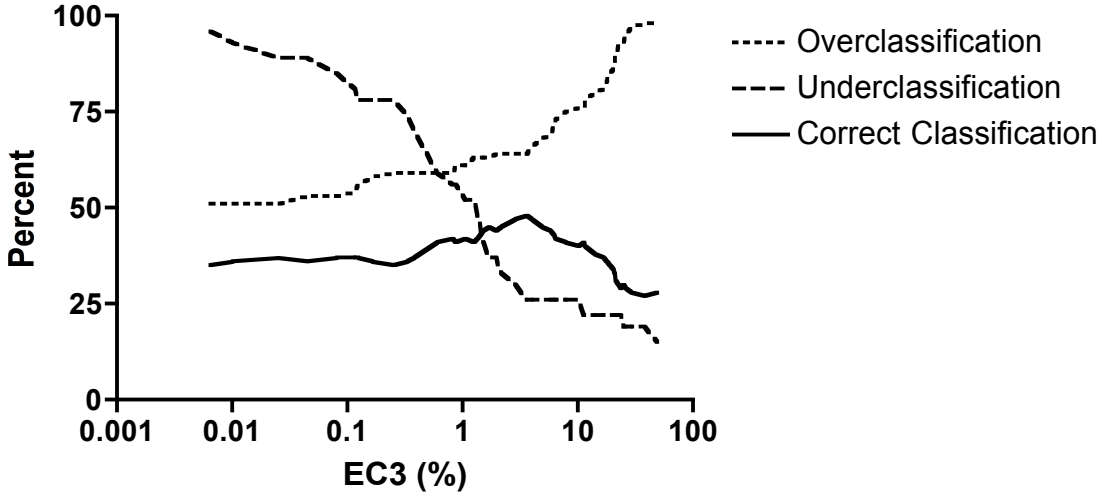
Analysis based on 52 substances identified as sensitizers in both the LLNA and in guinea pigs using the GPMT and/or the BT. In guinea pigs, sensitizers were classified as strong or weak based on the criteria in **Table 1-1**. EC3 = Estimated concentration needed to produce a stimulation index of three.

The second analysis included substances classified as sensitizers in both the LLNA and in humans as well as substances classified in the LLNA as false positives and false negatives compared to the human, and substances classified as non-sensitizers in both the LLNA and in humans. In this analysis, overclassification means that nonsensitizers are misclassified as

1105 weak or strong sensitizers and weak sensitizers are missclassified as strong while
1106 underclassification means that strong sensitizers are missclassified as weak or nonsensitizers
1107 and weak sensitizers are misclassified as nonsensitizers. Using the optimal EC3 value of
1108 3.6%, the correct classification rate was 57% (60/105), while the over- and under-
1109 classification rates ranged from 25% (8/32) to 61% (30/49) and 9% (3/32) to 17% (4/24),
1110 respectively.

1111

Figure 6-8 Correct Classification and Over- and Under-Classification Rates for the Ability of Different LLNA EC3 Values to Identify Strong and Weak Skin Sensitizers in Guinea Pigs Based on Criteria in Table 1-1, using Sensitizers, LLNA False Negatives, LLNA False Positives and Non-sensitizers against the Guinea Pig



Analysis based on 105 substances identified as sensitizers in both the LLNA and in guinea pigs using the GPMT and/or the BT, as false positive or false negative in the LLNA compared to guinea pig results, and as non-sensitizers in both the LLNA and in guinea pigs. In guinea pigs, sensitizers were classified as strong or weak based on the criteria in **Table 1-1**. EC3 = Estimated concentration needed to produce a stimulation index of three.

Table 6-6 Correct Classification and Misclassification Rates when the Optimal LLNA EC3 Value is Used to Predict the Guinea Pig Skin Sensitization Potency Classification[1]

Comparison	Overall Classification[2]	Classification						
		Strong Sensitizer		Weak Sensitizer			Non-Sensitizer	
		Correct[2]	Under[2]	Over[2]	Correct[2]	Under[2]	Correct[2]	Over[2]
Sensitizers only: optimal LLNA EC3 (2.0%) vs. strong/weak sensitizers in guinea pigs	73% (38/52)	74% (17/23)	26% (6/23)	28% (8/29)	72% (21/29)	NA	NA	NA
All data: optimal LLNA EC3 (3.6%) vs. strong/weak sensitizers and non-sensitizers in guinea pigs taking into account LLNA false negative and false positive sensitizers compared to the guinea pig	57% (60/105)	83% (20/24)	17% (4/24)	25% (8/32)	66% (21/32)	9% (3/32)	39% (19/49)	61% (30/49)

[1]The criteria for distinguishing between strong and weak sensitizers in the GPMT and the BT are provide in **Table 1-1**. The LLNA EC3 values used in this analysis are the ones that maximally reduce the over- and under-classification rates, based on the most potent LLNA EC3 or the most severe guinea pig classification for substances multiply tested, as determined in **Figures 6-7** and **6-8**.

[2]Overall Classification: The proportion of substances assigned to each hazard classification for skin sensitization potency (i.e., Strong, Weak, Non-Sensitizer). Correct**:** percentage of LLNA EC3 values that correctly predicted the guinea pig skin sensitization potency classification; Over**:** percentage of LLNA EC3 values that overclassified the guinea pig skin sensitization classification; Under**:** percentage of LLNA EC3 values that underclassified the guinea pig skin sensitization potency classification.

[3] The proportion on which the percentage calculation is based.

NA = Not applicable since only strong and weak sensitizers evaluated in the absence of non-sensitizers (i.e., Weak sensitizers can only be over-predicted and no non-sensitizers were evaluated)

6.3 Ability of the LLNA versus the Guinea Pig to Predict Skin Sensitization Potency in Humans

In this analysis, the comparative ability of the LLNA and the guinea pig to predict skin sensitization potency in humans was evaluated for substances with results in all three species. Due to the categorical nature of the data collected in the guinea pig tests (i.e., incidence of sensitized animals at a particular test substance concentration), the analysis was limited substances assigned a skin sensitization potency classification (strong versus weak) in humans.

In the current NICEATM LLNA database, there are 47 substances with human, LLNA, and guinea pig test results, 32 of which are classified as sensitizers in all three species (**Appendix B**). In this evaluation, the geometric mean EC3 and human threshold values and the weight of evidence sensitization classification in the guinea pig (see **Section 6.2**) were used. In the first analysis, only those substances classified as sensitizers in all three species were evaluated. In the second analysis, all substances with data in all three species were evaluated. The EC3 decision criteria was that determined against human data in **Section 6.1**. Based on the results of these analyses (**Table 6-7**), the LLNA achieved the correct potency classification at a slightly higher rate, as evidenced by overall classification rates of 57% to 72% for the LLNA as compared to 47% to 59% for the GP. While the LLNA more accurately predicted strong human sensitizers than the GP (71% to 75% correct versus 41% to 48% in the LLNA and GP, respectively), the GP more accurately predicted the human non-sensitizers than the LLNA (75% correct versus 50% correct in the GP and LLNA, respectively). Although the results were more variable for human weak sensitizers than the other two categories (i.e., strong and non-sensitizers), the GP was also better at predicting the human weak sensitizers (42% to 89% correct versus 42% to 75% correct in the GP and LLNA, respectively).

Table 6-7 Comparative Correct Classification and Over- and Under-classification Rates when the Optimal LLNA EC3 Value or the Guinea Pig Skin Sensitization Potency Classification is used to Predict the Human Skin Sensitization Classification[1]

Comparison	Overall Classification[2]	Classification						
		Strong Sensitizer		Weak Sensitizer			Non-Sensitizer	
		Correct[2]	Under[2]	Over[2]	Correct[2]	Under[2]	Correct[2]	Over[2]
Sensitizers only: optimal LLNA EC3 (6.8%) vs. strong/weak sensitizers in humans using a decision criterion of 250 μg/cm^2	72% (23/32)	71% (17/24)	29% (7/24)	25% (2/8)	75% (6/8)	NA	NA	NA
Sensitizers only: optimal LLNA EC3 (8.1%) vs. strong/weak sensitizers in humans using a decision criterion of 500 μg/cm^2	63% (20/32)	63% (17/27)	37% (10/27)	40% (2/5)	60% (3/5)	NA	NA	NA
Sensitizers only: guinea pig skin sensitization classification vs. strong/weak sensitizers humans using a decision criterion of 250 μg/cm^2	59% (19/32)	48% (11/23)	52% (12/23)	11% (1/9)	89% (8/9)	NA	NA	NA
Sensitizers only: guinea pig skin sensitization classification vs. strong/weak sensitizers humans using a decision criterion of 500 μg/cm^2	50% (16/32)	42% (11/26)	58% (15/26)	17% (1/6)	83% (5/6)	NA	NA	NA
All data: optimal LLNA EC3 (9.4%) vs. strong/weak sensitizers and non-sensitizers in humans using a decision criterion of 250 μg/cm^2 and taking into account LLNA false negative and false positive sensitizers compared to the human	64% (30/47)	75% (18/24)	25% (6/24)	13% (2/15)	53% (8/15)	33% (5/15)	50% (4/8)	50% (4/8)

Comparison	Overall Classification[2]	Classification						
		Strong Sensitizer		Weak Sensitizer			Non-Sensitizer	
		Correct[2]	Under[2]	Over[2]	Correct[2]	Under[2]	Correct[2]	Over[2]
All data: optimal LLNA EC3 (9.4%) vs. strong/weak sensitizers and non-sensitizers in humans using a decision criterion of 500 μg/cm^2 and taking into account LLNA false negative and false positive sensitizers compared to the human	57% (27/47)	67% (18/27)	33% (9/27)	17% (2/12)	42% (5/12)	42% (5/12)	50% (4/8)	50% (4/8)
All data: guinea pig sensitization classification vs. strong/weak sensitizers and non-sensitizers in humans using a decision criterion of 250 μg/cm^2 taking into account guinea pig false negative and false positive sensitizers compared to the human	53% (25/47)	46% (11/24)	54% (13/24)	7% (1/15)	53% (8/15)	40% (6/15)	75% (6/8)	25% (2/8)
All data: guinea pig sensitization classification vs. strong/weak sensitizers and non-sensitizers in humans using a decision criterion of 500 μg/cm^2 taking into account guinea pig false negative and false positive sensitizers compared to the human	47% (22/47)	41% (11/27)	59% (16/27)	8% (1/12)	42% (5/12)	50% (6/12)	75% (6/8)	25% (2/8)

[1]Two human threshold values (250 and 500 μg/cm^2) are proposed for distinguishing between strong and weak sensitizers. The LLNA EC3 values used in this analysis are the ones that maximally reduce the over- and under-classification rates, based on the most potent LLNA EC3 or human threshold value for multiply tested substances, as determined in **Figures 6-1** through **6-4**.

[2]Overall Classification: The proportion of substances assigned to each hazard classification for skin sensitization potency (i.e., Strong, Weak, Non-Sensitizer). Correct**:** percentage of LLNA EC3 values that correctly predicted the human skin sensitization potency classification; Over**:** percentage of LLNA EC3 values that overclassified the human sensitization classification; Under**:** percentage of LLNA EC3 values that underclassified the human skin sensitization potency classification.

[3] The proportion on which the percentage calculation is based.
[4]The criteria for distinguishing between strong and weak sensitizers in the GPMT and the BT are provided in **Table 1-1**. For substances multiply tested in the GPMT and/or BT, the majority classification category was used, or when an equal number of discordant classifications were recorded, the most severe classification category was used. The LLNA EC3 values used in this analysis are the ones that maximally reduce the over- and under-classification rates, as determined in **Figures 6-7** and **6-8**, and represent the geometric mean EC3 value if a substance had been tested more than once.
NA = Not applicable since only strong and weak sensitizers evaluated in the absence of non-sensitizers (i.e., Weak sensitizers can only be over-predicted and no non-sensitizers were evaluated)

7.0 Test Method Reliability

An assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM 2003). Repeatability refers to the closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM 1997, 2003). Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Interlaboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test substances, and indicates the extent to which a test method can be transferred successfully among laboratories.

As described in **Section 6.0**, the use of the LLNA for skin sensitization potency assessments depends on determining an accurate EC3 value for sensitizers. Thus, not only does the LLNA need to reproducibly achieve the correct call (i.e., sensitizer versus non-sensitizer), but it also needs to reproducibly assign the proper sensitization potency classification. An evaluation of the intralaboratory variability associated with 29 individual EC3 values for isoeugenol (which ranged from 0.5% to 2.6%, when tested in a single laboratory) was considered by Basketter and Cadby (2004) to support the "often-mentioned perspective that the biological variation associated with the estimation of EC3 values means that any particular EC3 can be halved or doubled" (Basketter and Cadby 2004).

Basketter et al. (2007) evaluated EC3 data for 17 sensitizers tested in at least two laboratories as a measure of interlaboratory reproducibility of the EC3 value (see **Appendix A-1**). The authors conclude that, although there is biological variation in the EC3 value among multiply tested substances using the same vehicle, this variation is less than an order of magnitude.

7.1 Vehicle Effects on LLNA Results

Although many factors impact skin sensitization, two important factors are the ability of the test substance to traverse the stratum corneum and reach the viable epidermis, as well as the efficiency of Langerhans cell migration from the skin. Both of these factors are susceptible to

vehicle effects, and therefore the vehicle chosen for testing in the LLNA can have an impact on results (Lea et al. 1999, Basketter et al. 2001, Wright et al. 2001, McGarry et al. 2007). Such effects need to be considered when evaluating the reproducibility of the LLNA in assigning the proper sensitization potency category. **Table 7-1** provides EC3 values for the 31 substances in the NICEATM database for which data from tests in multiple vehicles are available. These data indicate that the assigned sensitization potency classification would differ for 58% (18/31) of these substances when using the proposed two-level classification scheme (see **Table 1-1**). For eight of the 31 substances (26%), depending on the solvent used, the substances would have been classified as a sensitizer or a non-sensitizer in the LLNA.

McGarry et al. (2007) performed a similar analysis using the four-level system proposed by Kimber et al. (2003) (see **Table 1-2**) to demonstrate that the solvent used impacts on the EC3 value and the resulting sensitization classification of a substance. Among seven substances discussed for which data from tests in multiple solvents were available, six substances (86%) would have been assigned to different sensitization potency categories depending on the solvent used. When these data are applied to the proposed two-level classification scheme (**Table 1-1**), three substances (43%) would still have been assigned to different sensitization potency categories depending on the solvent used (**Table 7-2)**.

Wright et al. (2001) also investigated the influence of application vehicle on sensitizing potency, using the LLNA to examine the activity of four recognized human contact allergens: isoeugenol and cinnamic aldehyde and two fragrance chemicals; 3-dimethylaminopropylamine (a sensitizing impurity of cocamidopropyl betaine, a surfactant used in shower gel) and dibromodicyanobutane (the sensitizing component of Euxyl K 400, a preservative used in cosmetics). The four chemicals were applied in each of seven different vehicles (acetone: olive oil [4 : 1]; dimethyl sulfoxide: methyl ethyl ketone; dimethylformamide; propylene glycol; and both 50 : 50 and 90 : 10 mixtures of ethanol and water). It was found that the vehicle in which a chemical is presented to the epidermis can have a marked effect on sensitizing activity. EC3 values ranged from 0.9 to 4.9% for isoeugenol, from 0.5 to 1.7% for cinnamic aldehyde, from 1.7 to >10% for dimethylaminopropylamine and from 0.4 to 6.4% for dibromodicyanobutane. These authors confirm that the vehicle in which a chemical is encountered on the skin has an important

influence on the relative skin sensitizing potency of chemicals and may have a significant impact on the acquisition of ACD.

Figure 7-1 provides further indication that the vehicle used has pronounced effects on the predicted skin sensitization potency when based on LLNA EC3 values. Five representative substances were selected from those listed in **Table 7-1** based on available data from at least one LLNA test in multiple vehicles. These data demonstrate the potential impact of the vehicle on potency categorization when using the EC3 value, as greater than an order magnitude difference can be seen for all five substances. This is in contrast to the conclusions of Basketter et al. (2007) for multiple study data collected using the same solvent (i.e., that EC3 values are typically within an order of magnitude). Again, some substances were either negative or positive in LLNA, depending on the vehicle selected. One of these substances (1,4-dihydroquinone) is positive in the guinea pig (data are unavailable for a comparison to human). The other discordant substance (methyl salicylate) is positive in humans, but negative in the guinea pig.

1253 **Table 7-1 LLNA EC3 Values for Skin Sensitizers Tested in Different Vehicles (from the NICEATM Database)**

Substance	LLNA Vehicle and Associated EC3 Value (%)											GP Result	Human Result
	AOO[1]	DMF[2]	MEK[3]	PG[4]	DMSO[5]	A-AOO	ACE	DMF/H_2O	EtOH/DEP	L92	PE		
2-Amino-6-chloro-4-nitrophenol					**6.85**	**0.68**						NA	+
3-Aminophenol	**3.2**	**0.24**										+	NA
Benzocaine	**22**	**18**					**NC**					+	+
Butyl acrylate	11						24.4					+	NA
(Chloro)methylisothiazolinone	0.01*	0.009*		0.055*			0.005					+	+
Cinnamal	**3**	**0.05**										+	+
Cinnamic aldehyde	**2.2**	**0.19***										+	+
Citral	10.9*								6.3			+	+
Coumarin	**NC**	**29.6***										NA	+
Dihydrocoumarin	5.6	3.3										+	+
1,4-Dihydroquinone	**0.14***	**0.21***	**0.09***	**NC**								+	NA
2,4-DNCB	0.05*				0.015		0.012					+	+
Ethyl Acrylate	**32.5***						**NC**					-	+
Ethylenediamine	**2.2**						**NC**					+	+
Eugenol	10.4*						18.2		5.4			+	+
Formaldehyde	**0.4***	**0.21**		**2.8**			**0.44**			**7.03**		+	+
Geraniol	34.1*								17.2*			+	+
Glutaraldehyde	**0.12***	**0.02**		**1.5**			**0.06**	**2.1**				+	+
Glyoxal	1.4	0.6										NA	+
HCA	**5.6***						**1.2**			**10.14**		+	+
Hydroxycitronellal	23.25*	18.8										+	+
2-Mercaptobenzothiazole	**9.79***	**1.78***										+	+
Methylhydrocinnamal	17.36*	23.1										NA	+
Methylisothiazolinone	**0.87***			**2.2**								NA	+
Methyldibromoglutaronitrile	**2.2**	**2**									**1.9**	NA	NA
Methylmethacrylate	90						60					+	NA
Methyl salicylate	**NC**	**25**	**11.5**				**NC**					-	+
Nickel sulfate					**4.8, NC**					**NC**		+	+
Oxazolone	0.002*						0.002*					+	NA
Salicylic acid	**NC**						**12.2**					-	-
SLS		4.7*			2.5							-	-

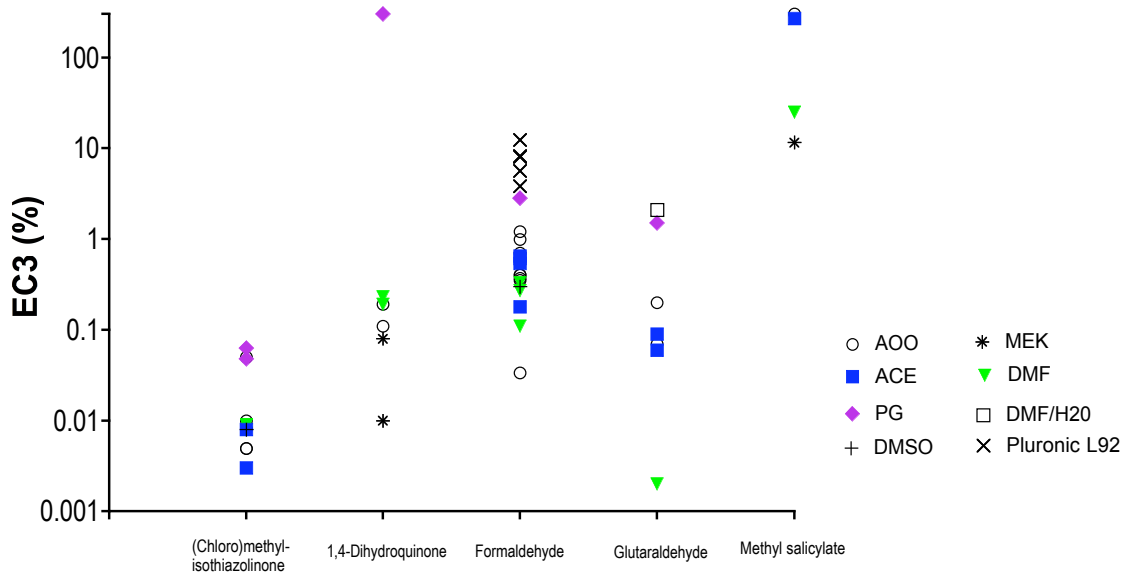
1254 Abbreviations: A-AOO = Aqua Acetone Olive:Oil; AOO = Acetone:Olive Oil; DMF = Dimethylformamide; DMF/H_2O = DMF/Water; DMSO = Dimethylsulfoxide; EtOH/DEP = Ethanol/ ; L92 = 1%
1255 Pluronic L92; MEK = Methyl ethyl ketone; NA = Not available; NC = Not calculated since SI < 3.0; PG = Propylene glycol * = Value represents a geometric mean of n ≥ 2 EC3 values
1256 Vehicles recommended by OECD TG 429 are indicated by a superscript and are listed in order of preference (OECD 2002); TG 429 also indicates that other vehicles may be used with sufficient
1257 scientific rationale.
1258 Bolded text highlights substances for which discordant classifications (using the proposed EC3 cutoffs of 1% or 2%, see **Table 1-2**) would be assigned depending on the vehicle used in the LLNA.

1259 **Table 7-2 LLNA EC3 Values for Skin Sensitizers Tested in Different Vehicles (from McGarry 2007)**

Substance	LLNA Vehicle and Associated EC_3 value (%)									GP Result	Human Result
	AOO[1]	DMF[2]	MEK[3]	PG[4]	DMSO[5]	ACE	L92	EtOH/H_2O (90:10)	EtOH/H_2O (50:50)		
Cinnamic aldehyde	**1.7**	**0.5**	**1.1**	**1.4**	**0.9**			**1.6**	**1.2**	+	+
1,4-Dihydroquinone	0.15	0.21	0.09		0.35	0.08				+	**NA**
3-Dimethylpropylamine	**2.2**	**1.7**	**1.8**	**> 10**	**3.2**			**4.1**	**7.1**	**NA**	**NA**
Isoeugenol	**1.0**	**1.4**	**1.0**	**2.5**	**0.9**			**1.8**	**4.9**	+	+
(Chloro)methylisothiazolinone/ Methylisothiazolinone	0.0049	0.0075	0.0068	0.048	0.0075	0.0076				+	+
Nickel sulfate		> 5.0			4.8		2.5			+	+
Potassium dichromate		0.0327			0.05		0.17			+	+

1260 L92; MEK = Methyl ethyl ketone; NA = Not available; PG = Propylene glycol
1261 Vehicles recommended by OECD TG 429 are indicated by a superscript and are listed in order of preference (OECD 2002); TG 429 also indicates that other vehicles may be used
1262 with sufficient scientific rationale.
1263 Bolded text highlights substances for which discordant classifications (using the proposed EC3 cutoffs of 1% or 2%, see **Table 1-2**) would be assigned depending on the vehicle
1264 used in the LLNA.

Figure 7-1 Representative Substances and Their Respective EC3 Values When Tested in Different Vehicles

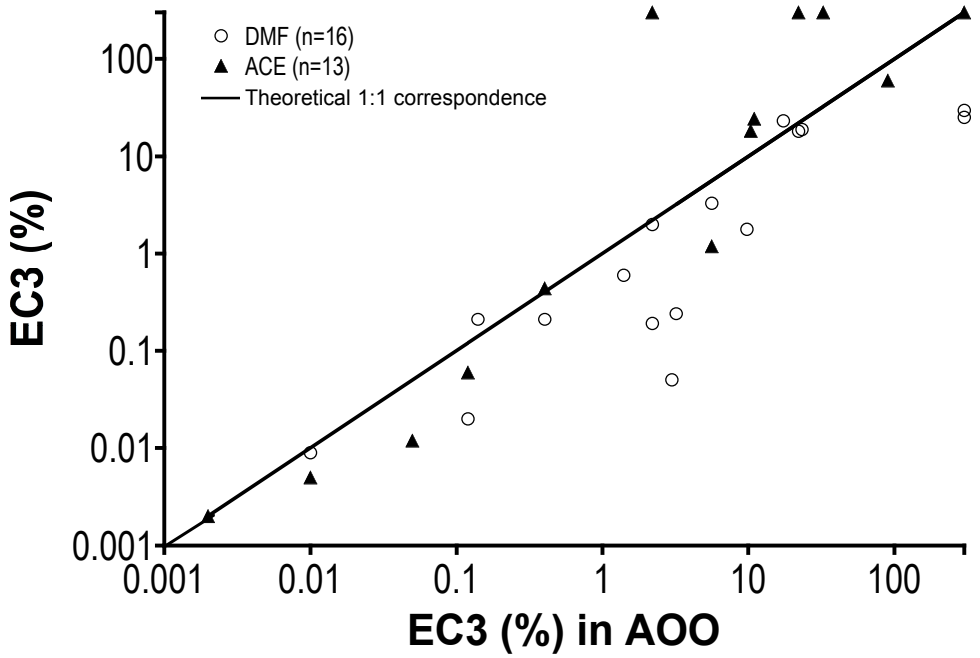


Note: Values above 100 indicate studies where the substance was classified as a nonsensitizer.

As another indicator of variability in EC3 values depending on the vehicle used, a correlation was calculated for EC3 values from two vehicles (DMF and acetone) when compared to the EC3 values for the same substance obtained with AOO (where a 1:1 correspondence would indicate that identical EC3 values had been obtained with the different solvents) (**Figure 7-2**). These data from a limited number of substances suggest that substances tested in acetone more closely correlate with those tested in AOO than do those tested in DMF. These data also suggest that EC3 values obtained with DMF are consistently lower than those obtained with acetone or AOO (i.e., the sensitizer is more potent when tested using this vehicle).

Four substances were negative when tested in acetone, one of which was also negative in AOO when tested at an even higher concentration (only two were also tested in DMF; both were positive). These are indicated on **Figure 7-2** as points that extend beyond the y-axis. Two of these four substances (benzocaine and ethylenediamine) are also positive in the guinea pig and human, while the remaining two substances (ethyl acrylate and methyl salicylate) are positive in humans, but negative in the guinea pig.

Figure 7-2 Correlation of EC3 Values from LLNA Tests with DMF or Acetone and Acetone:Olive Oil



Abbreviations: ACE = Acetone; AOO = Acetone: Olive Oil; DMF = Dimethylformamide; EC3 = the calculated concentration of a substance that induces a three-fold increase in the stimulation index.
NOTE: Symbols extending beyond the y-axis represent negative LLNA results (i.e., no EC3 calculated because stimulation index < 3)

8.0 LLNA Data Quality

8.1 Adherence to National and International GLP Guidelines

Based on the available information, the published papers, and data submissions, information on compliance with GLP guidelines was available for data obtained from Gerberick et al. (2005), E. Debruyne (Bayer CropScience SA), and P. Botham (ECPA).

A formal assessment of the quality of the remainder of the LLNA data considered here was not feasible. The published data on the LLNA were limited to tested concentrations and calculated SI and EC3 values. Auditing the reported values would require obtaining the original individual animal data for each LLNA experiment, which have been requested, but not yet obtained. However, many of the studies were conducted according to GLP guidelines, which implies that an independent quality assurance audit was conducted. The impact of any deviations from GLP guidelines cannot be evaluated for the data reviewed here, since no data quality audits was obtained.

As noted in **Section 5.0,** the original records were not obtained for the studies included in this evaluation. Data were available for several of the substances included in the ICCVAM (1999) evaluation and thus some of the raw data for these substances were available for review.

8.2 Data Quality Audits

Formal assessments of data quality, such as a quality assurance audit, generally involve a systematic and critical comparison of the data provided in a study report to the laboratory records generated for a study.

Much of the data published by Gerberick et al. (2005) was conducted following GLP guidelines or were conducted in GLP-compliant facilities. Therefore, it was previously inferred that data audits were conducted on the data (ICCVAM 1999).

A formal assessment of the quality of the remainder of the LLNA data included in this BRD was not feasible. The published data on the LLNA were limited to tested concentrations and calculated SI and EC3 values. Auditing the reported values would require obtaining the original individual animal data for each LLNA experiment. Such data were not obtained.

However, as stated in **Section 8.1**, many of the studies were conducted according to GLP guidelines, which implies that an independent quality assurance audit was conducted.

8.3 Impact of Deviations from GLP Guidelines

The impact of any deviations from GLP guidelines cannot be evaluated for the data reviewed in this BRD, since no information on data quality audits was obtained.

8.4 Availability of Laboratory Notebooks or Other Records

As noted in **Section 5.2,** the original records were not obtained for the studies included in this evaluation. Data were available for the substances included in the ICCVAM (1999) evaluation and thus some of the raw data for these substances were available for review.

9.0 Other Scientific Reports and Reviews

Several published studies have discussed the potential for using the LLNA to assess the relative skin sensitization potential of chemicals. The following section summarizes reviews that have been published on the use of LLNA for skin sensitization potency classifications. Since many of these reviews originate with collaborating scientists, the reviews are grouped together by these authors and arranged by date.

9.1 Basketter, Gerberick, Kimber, and Colleagues

9.1.1 Basketter et al. (2003)

The review discusses the usefulness of the LLNA for hazard identification and its current regulatory status. The review then also discusses the potential usefulness of the method to assess relative potency of chemicals and incorporation of the data into risk assessments.

The authors indicate that the use of LLNA to assess potency has been extensively evaluated in recent years. It is noted that factors to consider in the use of LLNA data for potency assessments include (1) how the potency is estimated from the LLNA, (2) the robustness of the estimation, (3) the relevance of the estimation, and (4) how the potency estimation is applied for risk assessment purposes. The authors note that several studies have shown that the calculated EC3 values, as discussed in Basketter et al. (1999b), correlate well with human potency classifications (Basketter et al. 2000, Ryan et al. 2000, Gerberick et al. 2001).

The authors note that for the LLNA potency information to be useful, it should be capable of being incorporated into risk assessments. Various proposals have been published which discuss incorporation of EC3 values into risk assessments (Robinson et al. 2000; Gerberick et al. 2001; Basketter et al. 2001b). It is proposed that combining various potential exposure conditions with calculated EC3 values would provide a way to incorporate the information into risk assessments (Basketter et al. 2002; Felter et al. 2002, 2003).

9.1.2 Kimber et al. (2003)

This review summarizes the efforts of a European Centre for Ecotoxicology and Toxicology (ECETOC) Task Force (see also ECETOC [2003]) that was charged with recommending approaches for the measurement of potency and defining thresholds for skin sensitization. The task force focused primarily on categorization of sensitizers and the identification of

thresholds with respect to the induction phase of skin sensitization. Based on their deliberations, the task force concluded that the LLNA is the method of choice for prospective skin sensitization potency assessments. The task force proposed a sensitization potency classification based on the EC3 values as follows:

- Extreme: EC3 ≤ 0.1%
- Strong: 0.1% ≤ EC3 < 1%
- Moderate: 1% ≤ EC3 < 10%
- Weak: 10% ≤ EC3 ≤ 100%

The authors recognized that available data from guinea pig tests provide valuable information for such assessments.

9.1.3 Jowsey et al. (2006)

This article discusses strategies for assessing skin sensitization without the use of animals. However, included in this discussion is a summary of the use of the LLNA for assessing relative skin sensitization potential of chemicals. The authors note that LLNA is useful for hazard characterization since it models all the events that occur during the process of skin sensitization and the extent to which sensitization will develop (i.e., magnitude of lymphocyte proliferation is an indicator of the extent of skin sensitization) (Kimber and Dearman 1991). Using this observation, it was proposed that using EC3 values derived from LLNA studies could be useful in assessing skin sensitization potency (Kimber and Basketter 1997, Basketter et al. 2001b). They also cite studies that demonstrate the accuracy and reliability of the EC3 value, and state that it consistently correlates with clinical estimates of human skin sensitization potential (Dearman et al. 1998, Warbrick et al. 1999, Basketter et al. 2000, Gerberick et al. 2001).

9.1.4 Basketter et al. (2007)

This review provides an overview of the available data that the authors consider to be supportive of the validity of the LLNA for assessments of sensitization potency. In the article, the authors discuss the relevance of the LLNA EC3 value to evaluating human skin sensitization potency, the reliability of the EC3 value, and the interlaboratory transferability of the method based on EC3 values.

Most studies attempt to assign chemicals to various categories (e.g., non-sensitizers, weak sensitizers, strong sensitizers) based on predefined cut-off EC3 values. While these studies tend to show good correlation between LLNA outcomes and human skin sensitization potential, more recent studies have attempted to correlate experimental thresholds in humans (e.g., no effect levels in human repeated insult patch tests) with the LLNA EC3 value. Although the outcomes are dependent on exposure conditions used in the patch tests, the authors conclude that the studies showed that there was a good relationship between EC3 values and the evaluated threshold levels (Schneider and Akkan 2004, Griem et al. 2003, Basketter et al. 2005).

The authors conclude that the EC3 value is a useful metric with which to predict the skin sensitization potential of chemicals in humans, and that intra- and inter-laboratory studies have shown that the EC3 value is reproducible within and among laboratories. The authors therefore propose that integration of the LLNA for potency identification in risk assessments would assist in developing more accurate hazard identification and risk management strategies.

9.1.5 Gerberick et al. (2007)

In this review, the authors discuss the concept of using the LLNA to assess skin sensitization potential of chemicals in humans. They cite several advantages of the LLNA (e.g., provides dose response data, allows for quantification of threshold values) that make it amenable to potency determinations. They also cite several studies that have evaluated the accuracy and reliability of the EC3 value for assessing potency (Warbrick et al. 1999, Dearman et al. 2001, Basketter and Cadby 2004). These and other studies have reportedly demonstrated good correlation between LLNA potency estimates and human potency, as assessed by clinical studies and experience (Basketter et al. 2000; Gerberick et al. 2001).

Based on these findings, the authors conclude that the LLNA should be considered the preferred method for skin sensitization hazard identification and that it can provide important additional information regarding sensitization potency that facilitates scientifically sound risk assessments.

9.2 McGarry (2007)

This review provides an overview of concerns that have been raised regarding the use of the LLNA upon implementation of the European chemicals legislation on the registration, evaluation, authorization, and restriction of chemicals (REACH). These concerns include that the LLNA is susceptible to vehicle effects (refer also to **Section 7.0**), it has not been validated for testing mixtures, and may result in a number of false positive response when tested with skin irritants. The author states that these concerns have become heightened given the current requirements in the REACH legislation for skin sensitization testing, which specifies that the LLNA must be used for new *in vivo* testing of skin sensitization hazards, and only under "exceptional circumstances" can another method used.

This intent of this review is to address these concerns from a European regulatory perspective, and to discuss the potential utility of the LLNA to provide information on skin sensitization potency of substances. Evidence of vehicle effects, both on overall LLNA results (i.e., "yes/no" decisions) and on potency estimations (i.e., EC3 values), is described for several commonly used vehicles. Problems associated with testing mixtures and formulations (e.g., compatibility with traditional LLNA vehicles, alteration of the active substance's bioavailability by excipients) are also described. The author concludes with a discussion of the potential utility of the LLNA for estimating sensitization potency, while cautioning that the EC3 should not be considered a measure of absolute potency.

9.3 Schlede et al. (2003)

This article is the culmination of a 16-year collaboration among dermatologists, industry representatives, and regulators to assign potency rankings to chemicals with skin sensitizing properties. Clinical and experimental data on humans and results of animal tests from the scientific literature were collected on 244 substances (i.e., technically produced chemicals as well as chemically defined single ingredients of natural products). Based primarily on “expert judgment” and in combination with reviews of the published literature, each substance was allocated into one of three defined categories (i.e., significant contact allergen [Category A], solid-based indication for contact allergenic effects [Category B], and Insignificant contact allergen or questionable contact allergenic effect [Category C]).

Published human data were obtained with the HMT or HRIPT, while the animal data were obtained with the GPMT, BT and/or the LLNA. Most of the human experimental data correlate with positive and positive/negative animal data. However, the authors state that published data on experimental human testing are limited in most cases to older studies with insufficient experimental design and/or limited documentation.

The authors conclude that results obtained with animal data are reliable and sensitive indicators for the prediction of skin sensitization potential in humans.

10.0 Animal Welfare Considerations

The proposal for using the LLNA for potency determinations does not impact its requirement for using animals, or the number of animals that will be required; these are defined in the recommended LLNA protocol (ICCVAM 1999, Dean et al. 2001). However, this application could broaden the use of the LLNA protocol in place of the guinea pig tests and could therefore further reduce the number of guinea pigs that are being used to assess skin sensitization potency.

11.0 Practical Considerations

Several issues are taken into account when assessing the practicality of using an alternative to an existing test method. In addition to performance evaluations, assessments of the laboratory equipment and supplies needed to conduct the alternative test method, level of personnel training, labor costs, and the time required to complete the test method relative to the existing test method are necessary. The time, personnel cost, and effort required to conduct the proposed test method(s) must be considered to be reasonable when compared to the existing test method it is intended to replace. No changes to the LLNA protocol are being proposed and therefore, the transferability, training requirements, and time and cost considerations for the LLNA remain unchanged (see ICCVAM 1999).

12.0 References

Basketter DA, Gerberick F, Kimber I. 2007. The local lymph node assay and the assessment of relative potency: status of validation. Contact Dermatitis 57:70-75.

Basketter D A, Clapp C, Jefferies D, Safford R J, Ryan C A, Gerberick G F, Dearman R J, Kimber I. 2005. Predictive identification of human skin sensitisation thresholds. Contact Dermatitis 53: 260–267.

Basketter D A, Cadby P. 2004. Reproducible prediction of contact allergenic potency using the local lymph node assay. Contact Dermatitis 50: 15–17.

Basketter DA, Evans P, Gerberick GF, Kimber IAN. 2002. Factors affecting thresholds in allergic contact dermatitis: Safety and regulatory considerations. Contact Dermatitis 47(1):1-6.

Basketter D A, Wright Z M, Warbrick E V, Dearman R J, Kimber I, Ryan C A, Gerberick G F, White I R. 2001a. Human potency predictions for aldehydes using the local lymph node assay. Contact Dermatitis 45: 89–94.

Basketter DA, Gerberick GF, Kimber I. 2001b. Measurement of allergenic potency using the local lymph node assay. Trends Pharmacol. Sci. 22: 264–265.

Basketter DA, Balikie L, Dearman RJ, Kimber I, Ryan CA, Gerberick GF, Harvey P, Evans P, White IR, Rycroft RJ. 2000. Use of the local lymph node assay for the estimation of relative contact allergenic potency. Contact Dermatitis 42: 344–348.

Basketter DA, Rodford R, Kimber I, Smith I, Wahlberg JE. 1999a. Skin sensitization risk assessment: a comparative evaluation of 3 isothiazolinone biocides. Contact Dermatitis 40: 150–154.

Basketter DA, Lea LJ, Cooper K, Stocks J, Dickens A, Pate I, Dearman RJ, Kimber I. 1999b. Thresholds for classification as a skin sensitiser in the local lymph node assay: a statistical evaluation. Food Chem Toxicol 37:1167–74.

Basketter DA, Lea LJ, Dickens A, Briggs D, Pate I, Dearman RJ, Kimber I. 1999c. A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. J Appl Toxicol. 19:261-266.

CPSC. 1992. Guidelines for Determining Chronic Toxicity of Products Subject to the FHSA. *Federal Register* Vol. 57, No. 197, Friday, October 9. 1992.

CPSC. 2006. CPSC Staff Report on the Draft Proposed Revision of the FHSA "Strong Sensitizer" Supplemental Definition. Available http://www.cpsc.gov/library/foia/foia07/os/StrongSensitizer.pdf

De Jong W H, van Och F M, Den Hartopg Jager C F, Spiekstra S W, Slob W, Vandebriel R J, van Loveren H J. 2002. Ranking of allergenic potency of rubber chemicals in a modified local lymph node assay. Toxicol Sci 66: 226–232.

Dean J, Twerdok L, Tice R, Sailstad D, Hattan D, and Stokes WS. 2001. ICCVAM Evaluation of the Murine Local Lymph Node Assay (LLNA) II: Conclusions and Recommendations of an Independent Scientific Peer Review Panel. Reg Toxicol Pharmacol 34:258-273.

Dearman R J, Wright Z M, Basketter D A, Ryan C A, Gerberick G F, Kimber I. 2001. The suitability of hexyl cinnamic aldehyde as a calibrant for the using local lymph node assay. Contact Dermatitis 44: 357–361.

Dearman RJ, Hilton J, Evans P, Harvey P, Basketter DA, Kimber I. 1998. Temporal stability of local lymph node assay responses to hexyl cinnamic aldehyde. J. Appl. Toxicol. 18: 281–284.

ECETOC. 2003. Contact Sensitization: Classification According to Potency. Technical Report No. 87. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

EPA. 2003. Health Effects Test Guideline, OPPTS 870.2600. Skin Sensitization EPA 712–C–03–197. Washington, DC: U.S. Environmental Protection Agency. Available: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Revised/870r-2600.pdf [accessed 20 September 2007].

EPA. 2006a. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR 792. Available: http://www.access.gpo.gov/nara/cfr/waisidx_06/40cfr792_06.html [accessed 20 September 2007].

EPA. 2006b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and Rodenticide Act. 40 CFR 160.

http://www.access.gpo.gov/nara/cfr/waisidx_06/40cfr160_06.html [accessed 20 September 2007].

FDA. 2007a. Good laboratory practice for nonclinical laboratory studies. 21 CFR 58.

Felter SP, Ryan CA, Basketter DA, Gerberick GF. 2003. Application of the risk assessment paradigm to the induction of allergic contact dermatitis. Regulatory Toxicol Pharmacol. 37:1-10.

Felter SP, Robinson MK, Basketter DA, Gerberick GF. 2002. A review of the scientific basis for default uncertainty factors for use in quantitative risk assessment of the induction of allergic contact dermatitis. Contact Dermatitis 47:257-266.

Gerberick GF, Ryan CA, Dearman RJ, Kimber I. 2007. Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. Methods 41(1):54-60.

Gerberick GF, Robinson MK, Ryan CA, Dearman RJ, Kimber I, Basketter DA, Wright Z, Marks J. 2001. Contact allergenic potency: correlation of human and local lymph node assay data. Am J Contact Dermat 12: 156–161.

Gerberick GF, Robinson MK. 2000. A skin sensitization risk assessment approach for the evaluation of new ingredients and products. Am J Contact Dermat. 11:65-73.

Griem P, Goebel C, Scheffler H. 2003. Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data. Regul Toxicol Pharmacol 38: 269–290.

Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz GY, Basketter DA. 2005. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. Dermatitis 16(4):157-202.

Hilton J, Dearman R J, Harvey P, Evans P, Basketter D, Kimber I. 1998. Estimation of the relative skin sensitizing potency using the local lymph node assay: a comparison of formaldehyde with glutaraldehyde. Am J Contact Dermat 9: 29–33.

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No.: 97-3981. Research Triangle Park:National Toxicology Program.

ICCVAM 1999. The murine local lymph node assay: A test method for assessing the allergic contact dermatitis potential of chemical/compounds. NIH Publication No. 99-4494. Research Triangle Park, NC: National Toxicology Program. Available http://iccvam.niehs.nih.gov/methods/immunotox/llna_PeerPanel98.htm (accessed January, 2008)

ICCVAM Authorization Act of 2000. 2000. Public Law 106-545. [114 Stat. 2721]. Available: http://iccvam.niehs.nih.gov/docs/about_docs/PL106545.pdf (accessed January 2008).

ICCVAM 2001. ICCVAM Immunotoxicity Working Group Recommended Protocol for the Murine Local Lymph Node Assay (LLNA): Testing of Chemicals for Contact Sensitizing (Allergic Contact Dermatitis [ACD]) potential. NIH Publication. January 2008. Research Triangle Park, NC: National Toxicology Program.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park:National Toxicology Program.

ICH. 1996. Guidance for Industry - E6 Good Clinical Practice: Consolidated Guidance. Available http://www.fda.gov/cder/guidance/959fnl.pdf

Jowsey IR, Basketter DA, Westmoreland C, Kimber I. 2006. A future approach to measuring relative skin sensitising potency: A proposal. Journal of Applied Toxicology 26(4):341-350.

Kimber I, Dearman RJ. 1991. Investigation of lymph node cell proliferation as a possible immunological correlate of contact sensitizing potential. Food and Chemical Toxicology 29(2):125-129.

Kimber I, Hilton J, Dearman R J et al. 1995. An international evaluation of the murine local lymph node assay and comparison of modified procedures. Toxicology 103:63–73.

Kimber I, Dearman RJ. 1996. Contact hypersensitivity: immunological mechanisms. In: Toxicology of Contact Hypersensitivity (I K, Maurer T, eds). London:Taylor and Francis, 4-25.

Kimber I, Basketter DA. 1997. Contact sensitization: a new approach to risk assessment. Human Ecological Risk Assessment. 3:385-395.

1580 Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Newsome C, Steiling W, Vohr
1581 HW. 2003. Classification of allergens according to potency: proposals. Food Chem Toxicol.
1582 41:1799-1809.

1583 Lalko J, Api AM. 2006. Investigation of the dermal sensitization potential of various
1584 essential oils in the local lymph node assay. Food and Chemical Toxicology 44(5):739-746.

1585 Lea JL, Warbrick EV, Dearman RJ, Kimber I, Basketter DA. 1999. The impact of vehicle on
1586 assessment of relative skin sensitization potency of 1,4-dihydroquinone in the local lymph
1587 node assay. Am J Contact Dermatitis. 10:213-218.

1588 Loveless SE, Ladics GS, Gerberick GF, Ryan CA, Basketter DA, Scholes EW, House RV,
1589 Hilton J, Dearman RJ, Kimber I. 1996. Further evaluation of the local lymph node assay in
1590 the final phase of an international collaborative trial. Toxicology 108:141–152.

1591 Marzulli FN, Maibach HI. 1976. Effects of vehicles on elicitation concentration in contact
1592 dermatitis testing. Contact Dermatitis. 2:325-329.

1593 McGarry. 2007. The murine local lymph node assay: Regulatory and potency considerations
1594 under REACH. Toxicology. 238:71-89.

1595 OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance
1596 Monitoring Number 1: OECD principles on Good Laboratory Practice (as revised in 1997).
1597 ENV/MC/CHEM(98)17. Paris:OECD.

1598 OECD. 2002. Test guideline 429. Skin Sensitisation: Local Lymph Node Assay, adopted
1599 April 24, 2002. In: OECD Guidelines for Testing of Chemicals. Paris:OECD.

1600 Robinson MK, Gerberick GF, Ryan CA, McNamee P, White IR, Basketter DA. 2000. The
1601 importance of exposure estimation in the assessment of skin sensitization risk. Contact
1602 Dermatitis 2000; 42: 251–9

1603 Ryan CA, Gerberick GF, Cruse LW, Basketter DA, Lea L, Blaikie L, Dearman RJ, Warbrick
1604 EV, Kimber I. 2000. Activity of human contact allergens in the murine local lymph node
1605 assay. Contact Dermatitis 43: 95–102.

1606 Schneider K, Akkan Z. 2004. Quantitative relationship between the local lymph node assay
1607 and human skin sensitization assays. Regul Toxicol Pharmacol 39: 245–255.

Schlede E, Aberer W, Fuchs T, Gerner I, Lessmann H, Maurer T, et al. 2003. Chemical substances and contact allergy - 244 Substances ranked according to allergenic potency. Toxicology 193(3): 219-259UN. 2005. Globally Harmonised System of Classification and Labelling of Chemicals (GHS). New York & Geneva, United Nations Publications.

Stotts J. Planning, conduct and interpretation of human predictive sensitization patch tests. In: Current Concepts in Cutaneous Toxicity, Drill VA, Lazar P (eds): Orlando, Academic Press.

Van Och FMM, Slob W, De Jong WH, Vandebriel RJ, Van Loveren H. 2000. A quantitative method for assessing the sensitizing potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of the uncertainty margins. Toxicology 146:49-59.

Warbrick E V, Dearman R J, Lea L J, Basketter D A, Kimber I. 1999. Local lymph node assay responses to paraphenylenediamine: intra- and inter-laboratory evaluations. J Appl Toxicol 19: 255–260.

WHO. 2007. IPCS Project on the Harmonization of Approaches to the Assessment of Risk from Exposure to Chemicals. General Conclusions and Recommendations of an IPCS International Workshop on Skin Sensitization in Chemical Risk Assessment. Available http://www.who.int/ipcs/methods/harmonization/areas/sensitization_summary.pdf.

Wright ZM, Basketter DA, Blaikie L, Cooper KJ, Warbrick EV, Dearman RJ, Kimber I. 2001. Vehicle effects on skin sensitizing potency of four chemicals: Assessment using the local lymph node assay. International Journal of Cosmetic Science. 23(2): 75-83.

[This Page Intentionally Left Blank]

13.0 GLOSSARY

Accuracy[11]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of *relevance*. The term is often used interchangeably with "concordance" (see also *two-by-two table*). Accuracy is highly dependent on the prevalence of positives in the population being examined.

Allergic Contact Dermatitis (ACD): A Type IV allergic reaction of the skin that results from skin contact with an allergen. Symptoms of ACD include, but are not limited to, development of erythema (redness) and edema (swelling).

Assay[14]: The experimental system used. Often used interchangeably with *test* and *test method*.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

Concordance[14]: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of *relevance*. The term is often used interchangeably with *accuracy* (see also *two-by-two table*). Concordance is highly dependent on the prevalence of positives in the population being examined.

EC3: the concentration of a substance estimated from the dose response curve to produce a three-fold increase in stimulation index, as compared to the concurrent vehicle control.

Essential test method component[14]: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components is necessary when the acceptability of a proposed test method is being evaluated based on performance standards derived from

mechanistically and functionally similar validated test method. [Note: Previously referred to as *minimum procedural standards*]

False negative[14]: A substance incorrectly identified as negative by a test method.

False negative rate[14]: The proportion of all positive substances falsely identified by a test method as negative (see *two-by-two table*). It is one indicator of test method accuracy.

False positive[14]: A substance incorrectly identified as positive by a test method.

False positive rate[14]: The proportion of all negative substances that are falsely identified by a test method as positive (see *two-by-two table*). It is one indicator of test method accuracy.

Good Laboratory Practices (GLP)[14]: Regulations promulgated by the U.S. Food and Drug Administration and the U.S. Environmental Protection Agency, and principles and procedures adopted by the Organization for Economic Cooperation and Development and Japanese authorities that describe record keeping and quality assurance procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Hazard[14]: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Human threshold response: In the evaluation included in this BRD, the threshold for induction of skin sensitization was considered to be the no observed effect level (NOEL, expressed as $\mu g/cm^2$) or, in the absence of negative data, the lowest observed effect level (LOEL, expressed as $\mu g/cm^2$), as described by Basketter et al. (2005).

Interlaboratory reproducibility[14]: A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test method can be transferred successfully among laboratories.

---

[11] Definition used by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM 2003).

Intralaboratory repeatability[14]: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Intralaboratory reproducibility[14]: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

Immunological: Relating to the immune system and immune responses.

In vivo: In the living organism. Refers to assays performed in multicellular organisms.

Local Lymph Node Assay (LLNA): An *in vivo* test method used to assess the skin sensitization potential of a substance by measuring the proliferation of lymphocytes in the lymph nodes draining the ears (i.e., auricular lymph nodes) of mice, subsequent to topical exposure on the ear to the substance. The LLNA relates lymphocyte proliferation to the incorporation of tritiated thymidine (^{3}H) into the cells of the draining lymph nodes.

Lymphocyte: A white blood cell found in the blood, lymph, and lymphoid tissues, which regulates and plays a role in acquired immunity.

Negative predictivity[14]: The proportion of correct negative responses among substances testing negative by a test method (see *two-by-two table*). It is one indicator of test method accuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

Non-sensitizer: A substance that does not cause skin sensitization following skin contact.

Performance[14]: The accuracy and reliability characteristics of a test method (see *accuracy*, *reliability*).

Positive control: A substance known to induce a positive response, which is used to demonstrate the sensitivity of the test method and to allow for an assessment of variability in the conduct of the assay over time. For most test methods, the positive control substance is tested concurrently with the test substance and the vehicle/solvent control. However, for

some *in vivo* test methods, periodic studies using a positive control substance is considered adequate by the Organisation of Economic Co-operation and Development (OECD).

Positive predictivity[14]: The proportion of correct positive responses among substances testing positive by a test method (see *two-by-two table*). It is one indicator of test method accuracy. Positive predictivity is a function of the sensitivity of the test method and the prevalence of positives among the substances tested.

Potency: For the purposes of this BRD, potency is defined as a function of the concentration of a substance that is required for either the induction or elicitation phases of a skin sensitization reaction. For induction, potency refers to the concentration of a substance needed to induce a sensitization response; the more potent the substance the smaller the quantity needed for induction. Likewise, for elicitation, potency refers to the concentration of a substance need to elicite a response in a previously sensitized individual; the more potent a substance, the smaller the quantity required for elicitation.

Prevalence[14]: The proportion of positives in the population of substances tested (see *two-by-two table*).

Protocol[14]: The precise, step-by-step description of a test, including the listing of all necessary reagents, criteria and procedures for the evaluation of the test data.

Quality assurance[14]: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures is assessed independently by individuals other than those performing the testing.

Reduction alternative[14]: A new or modified test method that reduces the number of animals required.

Reference test method[14]: The accepted *in vivo* test method used for regulatory purposes to evaluate the potential of a test substance to be hazardous to the species of interest.

Refinement alternative[14]: A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

Relevance[14]: The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the *accuracy* or *concordance* of a test method.

Reliability[14]: A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

Replacement alternative[14]: A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility[14]: The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

Sensitivity[14]: The proportion of all positive substances that are classified correctly as positive in a test method. It is a measure of test method accuracy (see *two-by-two table*).

Skin sensitizer: A substance that induces an allergic response following skin contact. (U.N. 2005)

Specificity[14]: The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy (see *two-by-two table*).

Stimulation Index (SI): A value calculated for the Local Lymph Node Assay, to assess the skin sensitization potential of a test substance. The value is calculated as the ratio of radioactivity incorporated into the auricular lymph nodes of a group of treated mice to the radioactivity incorporated into the corresponding lymph nodes of a group of vehicle control mice. For the LLNA, an SI equal to or greater than 3 classifies a substance as a skin sensitizer.

Test[14]: The experimental system used; used interchangeably with *test method* and *assay*.

Test method[14]: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a

substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with *test* and *assay*. See also *validated test method* and *reference test*.

Transferability[14]: The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two table[14]: The two-by-two table can be used for calculating accuracy (concordance) ([a+d]/[a+b+c+d]), negative predictivity (d/[c+d]), positive predictivity (a/[a+b]), prevalence ([a+c]/[a+b+c+d]), sensitivity (a/[a+c]), specificity (d/[b+d]), false positive rate (b/[b+d]), and false negative rate (c/[a+c]).

		New Test Outcome		
		Positive	**Negative**	**Total**
Reference Test Outcome	Positive	a	c	a + c
	Negative	b	d	b + d
	Total	a + b	a + d	a + b + c + d

Validated test method[14]: An accepted test method for which validation studies have been completed to determine the relevance and reliability of this method for a specific proposed use.

Validation[14]: The process by which the reliability and relevance of a procedure are established for a specific purpose.

Vehicle control: An untreated sample containing all components of a test system, including the vehicle that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the same vehicle.

Weight-of-evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.