Background Review Document of Existing Methods for Eye Irritation Testing:

Silicon Microphysiometer and Cytosensor Microphysiometer

Contract No. CCR.IHCP.C431305.X0

Prepared for:

European Commission
Directorate-General JRC
Joint Research Centre
Institute for Health and Consumer Protection

Prepared by:

Institute for In Vitro Sciences, Inc.
30 W. Watkins Mill Road, Suite 100
Gaithersburg, MD 20878
USA

Authors:

Rodger D. Curren, Ph.D.
Angela Sizemore, B.S.
Jennifer Nash, M.S.
Greg Mun, B.S
Amanda Ulrey, RQAP-GLP
John W. Harbell, Ph.D.

Final Revision

20 August 2008

TABLE OF CONTENTS

List of Tables

List of Figures

List of Annexes

List of Abbreviations

ATP	Adenosine triphosphate
BRD	Background review document
CAM	Chroioallantoic Membrane Vascular Assay
CM	Cytosensor microphysiometer
CMC	Critical Micelle Concentration
COLIPA	European Cosmetic, Toiletry, and Perfumery Association
CRO	Contract Research Organization
CTFA	U.S. Cosmetics, Toiletries, and Fragrance Association
CV	Coefficient of variance
DNA	Deoxyribonucleic acid
EC/HO	European Commission/British Home Office
ECVAM	European Centre for the Validation of Alternative Methods
EU	European Union
GHS	Globally Harmonized System
GLP	Good laboratory practices
HCE	Human Cornea Epithelium
ICCVAM	Interagency Co-ordinating Committee on the Validation of Alternative Methods
INVITTOX	Compendium of *In vitro* Toxicology protocols maintained by ECVAM
IRAG	U.S. Interagency Regulatory Alternatives Group
LVET	Low volume eye test
MAS	Maximum Average Score
MMAS	Modified Maximum Average Score
MRD_{50}	Metabolic rate decrement of 50%
MT	Management Team
NPV	Negative predictive value
pH	An acidity/alkalinity index; the reciprocal of the hydrogen ion concentration
PPV	Positive predictive value
QC	Quality Control
SD	Standard Deviation
SEM	Standard Error of the Mean
SLS	Sodium lauryl sulfate
SM	Silicon microphysiometer
SOP	Standard Operating Procedures
USEPA	United States Environmental Protection Agency

Acknowledgements

The generous contributions from the following companies and individuals who provided *in vitro* and/or *in vivo* data considered in this Background Review Document are gratefully acknowledged:

The Procter and Gamble Company, Cincinnati, OH USA
Pauline McNamee and Rosemarie Osborne

L'Oreal Corporation, Paris, France
Odile deSilva

Institute for In Vitro Sciences, Inc., Gaithersburg, MD USA
Greg Mun and Hans Raabe

1. Data collection

1.1 Description of the methods used to collect data, including literature search or other sources, and number of studies collected

Retrospective analysis of a method requires a complete search for available data. The method under review here utilizes an instrument which performs real time measurements of the metabolic rate of a cell population. This instrument has existed in two forms, and thus a search for data must include reference to the names of both instruments. The use of the silicon microphysiometer (SM) (the prototype instrument) began in approximately 1989 and the Cytosensor microphysiometer (the commercial instrument) (hereafter referred to as Cytosensor or CM) became available in 1995. In 1993, the sensor chambers for the silicon microphysiometer were changed from the coverslip configuration to the transwell configuration in preparation for the introduction of the commercial instrument. At that point, the standard protocol was developed for the transwell exposures and that protocol has been used for both the transwell-equipped silicon microphysiometer and the commercial Cytosensor. Most of the data provided in this BRD come from the transwell configuration of the silicon microphysiometer and Cytosensor instruments and its assay protocol. The Cytosensor is the instrument in current use.

Several of the authors of this Background Review Document (BRD) have been involved with the SM and subsequently the CM since 1990. Thus, they are aware of many of the previous studies on the evaluation of eye irritation potential. However, a full literature search (conducted through NERAC, Inc., Tolland, CT) was undertaken using “microphysiometer” and “Cytosensor” search terms. The databases searched by NERAC included, but were not limited to, Biobusiness, Biological Abstracts, Medline, Embase, and Life Sciences Collection. As the Cytosensor was primarily intended as a drug development tool, most of the references provided by the search dealt with specific receptor binding assays and were not applicable to this BRD.

In addition, both Toxnet and Pubmed were searched directly, and the following unique references were found. There were no references found in the NERAC search that were different than those found in Toxnet and Pubmed listed below. References that were similar between search terms are listed only once below. Table 1.1 describes the number of unique returns with each keyword for each database. The numbers in bold indicate articles returned with reasonable relevancy. Relevant articles are those defined as either having information on the toxicity of chemicals or formulations, or those giving important background information on the functioning of the SM (or CM) or how cellular changes in metabolic rate could be interpreted.

Table 1.1 Search items returned in Toxnet and Pubmed

Search Term	Database	
	TOXNET	PUBMED
Microphysiometer	**17** (55)	**20** (138)
Cytosensor	**5** (35)	**15** (97)
Silicon Microphysiometer	**8** (22)	**5** (24)

Toxnet:

Parce, J.W., et al., *The Microphysiometer and Its Application in Irritancy Testing. In vitro* Cell Dev Biol, 1990. **26**(3 Part 2): p. 35A.

Bruner, L.H., et al., *Evaluation of Seven In vitro Alternatives for Ocular Safety Testing.* Fundam Appl Toxicol, 1991. **17**(1): p. 136-149.

Bagley, D.M., et al., *An Evaluation of Five Potential Alternatives In vitro to the Rabbit Eye Irritation Test In vivo.* Toxicology *In vitro*, 1992. **6**(4): p. 275-284.

Calvin, G., *New Approaches to the Assessment of Eye and Skin Irritation.* Toxicology Letters, 1992. **64/65**: p. 157-164.

Parce, J.W., *Cells on Silicon Bioassays with a Microphysiometer.* FED AM SOC EXP BIOL, 1992. **6**(1): p. A5.

Catroux, P., et al., *The Silicon Microphysiometer for Testing Ocular Toxicity In vitro.* Toxicol *In vitro*, 1993. **7**(4): p. 465-469.

Harvell, J.D., et al., *An In vivo Correlation with Three In vitro Assays to Assess Skin Irritation Potential.* Journal of Toxicology - Cutaneous and Ocular Toxicology, 1994. **13**(2): p. 171-183.

Hirst, M.A., C.E. Green, and C. Tyson, .A., *Initial Studies of the Effects of Toxic Agents on Hepatocytes Using the Cytosensor Microphysiometer System. In vitro* Toxicology, 1994. **7**(2): p. 136.

Ajilore, O.A. and R.M. Sapolsky, *Application of Silicon Microphysiometery to Tissue Slices: Detection of Metabolic Correlates of Selective Vulnerability.* Brain Research, 1997. **752**(7-2): p. 99-106.

Botham, P., et al., *Cell Function-Based Assays.* Food and Chemical Toxicology, 1997. **35**(1): p. 67-77.

Botham, P., et al., *IRAG Working Group 3. Cell Function-based Assays. Interagency Regulatory Alternatives Group.* Food and Chemical Toxicology, 1997. **35**(1): p. 67-77.

Cao, C.J., et al., *Toxicity of Sea Nettle Toxin to Human Hepatocytes and the Protective Effects of Phosphorylating and Alkylating Agents.* Toxicon, 1998. **36**(2): p. 269-281.

Gronert, K., S.P. Colgan, and C.N. Serhan, *Characterization of Human Neutrophil and Endothelial Cell Ligand-Operated Extracellular Acidification Rate by Microphysiometry: Impact of Reoxygenation.* J Pharmacol Exp Ther, 1998. **285**(1): p. 252-261.

Jordan, R.E., et al., *Activation of the Cloned Human NK3 Receptor in Chinese Hamster Ovary Cells Characterized by the Cellular Acidification Response Using the Cytosensor Microphysiometer.* British Journal of Pharmacology, 1998. **125**(4): p. 761-766.

Cao, C.J., et al., *Cytotoxicity of Organophosphate Anticholinesterases. In vitro* Cell Dev Biol Anim, 1999. **35**(9): p. 493-500.

Cooke, D. and R. O'Kennedy, *Comparison of the Tetrazolium Salt Assay for Succinate Dehydrogenase with the Cytosensor Microphysiometer in the Assessment of Compound Toxicities.* Analytical Biochemistry, 1999. **274**(2): p. 188-194.

Harbell, J.W., et al., *Assessment of the Cytosensor™ Microphysiometer Assay in the COLIPA In vitro Eye Irritation Validation Study.* Toxicology *In vitro*, 1999. **13**: p. 313-323.

Koebe, H.G., et al., *In vitro Toxicology in Hepatocyte Bioreactors-Extracellular Acidification Rate (EAR) in a Target Cell Line Indicates Hepato-activated Transformation of Substrates.* Toxicology, 2000. **154**(1-3): p. 31-44.

Burvall, K., L. Palmberg, and K. Larsson, *Metabolic Activation of A549 Human airway Epithelial Cells by Organic Dust: A Study Based on Microphysiometery.* Life Sciences, 2002. **71**(3): p. 299-309.

Deglmann, C.J., et al., *A New Bioassay Including a Small Scale Hepatocyte Bioreactor for Hepato-mediated Toxicity Testing in a Target Cell Line.* Int J Artif Organs, 2002. **25**(10): p. 975-984.

Silbergeld, E.K., *Neurotoxicology Studies.* CRISP.

Pubmed:

Owicki, J.C. and J.W. Parce, *Bioassays with a Microphysiometer.* Nature, 1990. **344**(6263): p. 271.

Bruner, L.H., et al., *Evaluation of Seven In vitro Alternatives for Ocular Safety Testing.* Fundam Appl Toxicol, 1991. **17**(1): p. 136-149.

McConnell, H.M., et al., *The Cytosensor Microphysiometer: Biological Applications of Silicon Technology.* Science, 1992. **257**(5078): p. 1906-12.

Botham, P., et al., *IRAG Working Group 3. Cell Function-based Assays. Interagency Regulatory Alternatives Group.* Food and Chemical Toxicology, 1997. **35**(1): p. 67-77.

Cao, C.J., et al., *Toxicity of Sea Nettle Toxin to Human Hepatocytes and the Protective Effects of Phosphorylating and Alkylating Agents.* Toxicon, 1998. **36**(2): p. 269-281.

Cao, C.J., et al., *Cytotoxicity of Organophosphate Anticholinesterases. In vitro* Cell Dev Biol Anim, 1999. **35**(9): p. 493-500.

Cooke, D. and R. O'Kennedy, *Comparison of the Tetrazolium Salt Assay for Succinate Dehydrogenase with the Cytosensor Microphysiometer in the Assessment of Compound Toxicities.* Analytical Biochemistry, 1999. **274**(2): p. 188-194.

Ren, X., D. Wang, and H. Li, *Study on Electrochemical Behavior of HL-60 Cells During the Etioposide-inducing Apoptosis.* Zhonghua Yi Xue Za Zhi, 1999. **20**(2): p. 82-84.

Hafner, F., *Cytosensor Microphysiometer: Technology and Recent Applications.* Biosens Bioelectron, 2000. **15**(3-4): p. 149-158.

Koebe, H.G., et al., *In vitro Toxicology in Hepatocyte Bioreactors-Extracellular Acidification Rate (EAR) in a Target Cell Line Indicates Hepato-activated Transformation of Substrates.* Toxicology, 2000. **154**(1-3): p. 31-44.

Chen, Z.W., K. Yang, and Y. Wang, *Microphysiometer-a real-time, Sensitive Method for Evaluation of the Functional Activity of Cells.* Sheng Li Ke Xue Jin Zhan, 2001. **32**(3): p. 243-245.

Luckie, D.B., et al., *CFTR Activation Raises Extracellular pH of NIH/3T3 Mouse Fibroblasts and C127 Epithelial Cells.* J Membr Biol, 2001. **179**(3): p. 275-284.

Burvall, K., L. Palmberg, and K. Larsson, *Metabolic Activation of A549 Human airway Epithelial Cells by Organic Dust: A Study Based on Microphysiometery.* Life Sciences, 2002. **71**(3): p. 299-309.

Cai, B., et al., *Apoptosis-inducing Activity of Extract from Chinese Herb, Albizzia Lucidior I. Nielsen.* Ai Zheng, 2002. **21**(4): p. 373-378.

Deglmann, C.J., et al., *A New Bioassay Including a Small Scale Hepatocyte Bioreactor for Hepato-mediated Toxicity Testing in a Target Cell Line.* Int J Artif Organs, 2002. **25**(10): p. 975-984.

Landwojtowicz, E., P. Nervi, and A. Seelig, *Real-time Monitoring of P-glycoprotein Activation in Living Cells.* Biochemistry, 2002. **41**(25): p. 8050-8057.

Park, T.H. and M.L. Shuler, *Integration of Cell Culture and Microfabrication Technology.* Biotechnol Prog, 2003. **19**(2): p. 243-253.

Wille, K., L.A. Paige, and A.J. Higgins, *Application of the Cytosensor Microphysiometer to Drug Discovery.* Receptors Channels, 2003. **9**(2): p. 125-131.

Eklund, S.E., et al., *A Microphysiometer for Simultaneous Measurement of Changes in Extracellular Glucose, Lactate, Oxygen, and Acidification Rate.* Anal Chem, 2004. **76**(3): p. 519-527.

Erxleben, H.A., et al., *A Novel Approach for Monitoring Extracellular Acidification Rates: Based on Bead Injection Spectophotometry and the Lab-on-valve System.* Analyst, 2004. **129**(3): p. 205-212.

Gatlik-Landwojtowicz, E., P. Aanismaa, and A. Seelig, *The Rate of P-glycoprotein Activation Depends on the Metabolic State of the Cell.* Biochemistry, 2004. **43**(46): p. 14840-14851.

Gatlik-Landwojtowicz, E., P. Aanismaa, and A. Seelig, *Quantification and Characterization of P-glycoprotein-substrate Interactions.* Biochemistry, 2006. **45**(9): p. 3020-3032.

1.1.1 Studies identified outside of database searches

In addition to the references identified by the standard database searches, the authors of this BRD were aware of other articles dealing with the performance of the SM or CM in ocular irritation studies. Those articles are listed below.

Balls, M., P. A. Botham, et al. *The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test*. Toxicology *In vitro*, 1995. **9**(6): p. 871-929.

Brantom, P. G., L. H. Bruner, et al. *A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test*. Toxicology *In vitro*, 1997. **11**: p. 141-179.

Bruner, L. H., K. M. Miller, et al. *Testing Ocular Irritancy In vitro with the Silicon Microphysiometer*. Toxicology *In vitro*, 1991. **5**: p. 277-284.

Gettings, S. D., R. A. Lordo, et al. *The CTFA Evaluation of Alternatives Program: An Evaluation of In vitro Alternatives to the Draize Primary Eye Irritation Test. (Phase III) Surfactant-based Formulations.* Food Chem Toxicol, 1996. **34**(1): p. 79-117.

Parce, J. W., J. C. Owicki, et al. *Detection of Cell-affecting Agents with a Silicon Biosensor.* Science, 1989. **246**: 243-247.

1.2 Brief description of data collected on overall study management

The major focus of the data collection was on data that provided parallel animal and *in vitro* data. Two major categories are data generated in 3rd party evaluation/validation studies and in-house data from individual companies (e.g., product safety data). To this end, the Cosmetics, Toiletries, and Fragrance Association (CTFA) Phase III study (Gettings, Lordo et al. 1996), European Commission/British Home Office (EC/HO) (Balls, Botham et al. 1995), and COLIPA Eye Irritation Validation study (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999) data have been compiled as examples of the first category. The EC/HO and COLIPA studies provide data for the evaluation of between-laboratory consistency as well. For the second category, the largest data set (~80 paired data sets) comes from the in-house data of Company # 1. This corporate data set from Company # 1 uses the LVET as the reference data. The EC/HO, COLIPA, and CTFA use the Draize test as the *in vivo* reference assay. For comparison purposes, the CTFA study tested the materials in their study with both the Low Volume Eye Test (LVET) (Griffith, Nixon et al. 1980; Freeberg, Nixon et al. 1986) and the Draize test (Freeberg, Nixon et al. 1986).

The positive control for both the SM and CM has generally been sodium lauryl sulfate (SLS). Tracking of these results over time gives valuable information concerning the reproducibility of the assay within a laboratory. These data were generated both at Company # 4 (1994-1997) and at Company # 3 (1997-2006). Therefore, a historical control database for each assay has been provided using the in-house information from the archives of Company # 3.

Several studies provided only summary *in vivo* data (i.e., MMAS) and so the analysis of Globally Harmonized System (GHS), European Union (EU) and US Environmental Protection Agency (USEPA) Categories is not possible. The *in vivo* and *in vitro* data from the two Bruner, et al. studies (Bruner, Kain et al. 1991); (Bruner, Miller et al. 1991) are the same and use the standard silicon microphysiometer protocol. Company # 2 has provided summary *in vivo* data on the ingredients tested (Catroux, Rougier et al. 1993). Three data sets were summarized in the Interagency Regulatory Alternatives Group (IRAG) (Brantom, Bruner et al. 1997) report. These include the CTFA Phase III study (Company # 4) (Gettings,

Lordo et al. 1996), the data from Company # 2 (Catroux, Rougier et al. 1993), and early data from Company # 1.

Two additional studies are included to compare the silicon microphysiometer and the Cytosensor instrumentation. The report of Bagley, Bruner et al. 1992 contains summary MAS values only but does compare the MRD_{50} values for the silicon microphysiometer with coverslip (standard protocol with a 500-second exposure at each concentration) and the silicon microphysiometer with transwell (500 second exposure). An unpublished study provided by Company # 1 has been included to compare the MRD_{50} values obtained with the silicon microphysiometer with the glass coverslip (standard protocol) and with the Cytosensor with the transwell (with its standard protocol). This study was performed to develop a translation factor between the SM and CM MRD_{50} values.

Table 1.2 summarizes the studies included in this Background Review Document. Approximately 200 paired data sets with complete *in vivo* and at least summary *in vitro* data are available for review. Additional data sets include summary *in vivo* and *in vitro* data.

Table 1.2 Brief description of data collected with some examples

Sponsor	Management	Chemical / formulation Selection / Coding	Process of data collection and statistical analysis	Standardized Prediction model	Quality of data (*in vivo* / *in vitro*) (GLP/non-GLP)	Standardized protocol available	Format of data available (*in vivo* / *in vitro*)
Company # 1[1] (Bruner, Miller et al. 1991)	Corporate participants	Selected from the corporate database of personal care, household cleaning products, and ingredients; coding unknown	Internal	No	Unknown	Not available for this BRD	Summary (MAS LVET) and mean MRD_{50} values. GLP status of either data set is unknown
Company # 1[1] (Bruner, Kain et al. 1991)	Corporate participants	These chemicals and data are the same as those in the Bruner, et al. 1991 publication	Internal	No	Unknown	Not available for this BRD	Summary (MAS LVET) and mean MRD_{50} values. GLP status of either data set is unknown
Several companies[2] (Bagley, Bruner et al. 1992)	Corporate participants	Selected from household cleaning products, personal care, and cosmetics categories plus ingredients for these categories	Unknown	No	Unknown	Not available for this BRD	Summary MAS (Draize and LVET) and MRD_{50} values (silicon microphysiometer and Cytosensor, both 500 sec exposure).
Company # 2[3] (Catroux, Rougier et al. 1993)	Corporate participants	Selected from the corporate database (ingredients and formulations) coding unknown	Internal	Not prestudy, but one proposed posthoc	Unknown	Not available for this BRD	Summary MAS and MRD_{50} values in publication. More complete data for 20 of the 53 materials obtained for this BRD.
EC/HO[4] (Balls, Botham et al. 1995)	Independent MT; laboratories selected based on experience	Selected by the management team based on the availability of good animal data (ECETOC) and a range of chemical classes/coded	Independent contractor	None	*In vivo* – Unknown *In vitro* – non GLP	Yes, but level of adherence to protocol not documented;	*In vivo* – all *In vitro* – summaries (means) for 3 labs, raw data available from 1 lab

CTFA[5] (Gettings, Lordo et al. 1996)	Independent	Independently selected (based on a range of irritancy) from the categories surfactants and surfactant-containing products, supplied coded	Independent	None	*In vivo* - GLP *In vitro* – non GLP; concurrent testing *in vivo* / *in vitro*	Yes (only one lab running the assay); not available for this BRD	*In vivo* – all *In vitro* – each trial
COLIPA[6] (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999)	Independent MT; laboratories selected based on experience	Independent chemical / formulation selection/ coded	Independent	Yes, based on MMAS	*In vivo* – GLP (some concurrent / some historical) *In vitro* – non GLP	Yes	*In vivo* – all *In vitro* – each trial
Company # 3[6 & 7] (Unpublished – in-house positive control data)	Company # 3	SLS positive control Not coded	Concurrent testing with each trial	Defined limits	Generally GLP	Yes (both SM and CM)	MRD_{50} from each trial and summary of data over time
Company # 1[6 & 7] (Unpublished)	Corporate and contract laboratory	Materials tested as part of ongoing products/ingredient safety program / Coded for the testing laboratory	Reports to Company # 1 from the CRO. Analysis conducted in this BRD	Not Available	Generally GLP (*in vitro*)	Yes (both SM and CM)	Complete animal data (LVET) and summary MRD_{50} (derived from three trials)

[1] Performed using the silicon microphysiometer coverslip protocol with a 320-second exposures and normal human epidermal keratinocytes (see (Bruner, Miller et al. 1991)).
[2] Performed using the silicon microphysiometer coverslip protocol and the silicon microphysiometer transwell protocol with a 500-second exposure cycle and L929 cells.
[3] Performed using the silicon microphysiometer transwell protocol with the 400-second exposure cycle and L929 cells.
[4] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[5] Performed using the silicon microphysiometer coverslip protocol with a 300-second exposure and L929 cells.
[6] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[7] Performed using the silicon microphysiometer coverslip protocol with a 500-second exposure and L929 cells.

2. Test Definition (Module 1)

2.1 Rationale for the proposed test method

2.1.1 Intended uses / purpose

Currently the CM is used by industry early in the product development process to screen liquid ingredients for cosmetic, personal care, and household cleaning products. This is then often followed by evaluations of the final formulations for final in-house safety decisions. Data from the CM may be combined with information from other *in vitro* or in silico assays to provide a "weight of evidence" evaluation of the formulation. Information from this assay is generally not combined with animal data in making the final safety decision for the product.

At the time the CM technology was developed, a number of *in vitro* assays such as the Neutral Red Uptake assay were already proposed as potential replacements for the Draize eye irritation test. However, the great advantage of the CM or SM technology was that real time measurements could be made of the cytotoxic response of the target cells as opposed to the 2-3 days or longer time which was required of the existing cytotoxicity assays. Thus, the assay was mainly created not to reveal a completely new endpoint, but rather to provide data in a much shorter time period.

2.1.2 Regulatory rational and applicability

To the best of our knowledge, the CM assay is not currently included in the regulatory scheme of any country. Data are used primarily to evaluate raw materials and formulations where regulatory registration is not required. It has been reviewed informally by regulatory agencies in the US as part of the Interagency Regulatory Alternatives Group (IRAG) evaluation of alternative ocular irritation assays (Botham, Osborne et al. 1997), and it is expected to be included as one of the assays that will be evaluated by the US Environmental Protection Agency (through the Interagency Co-ordinating Committee on the Validation of Alternative Methods (ICCVAM)) as part of a larger initiative to replace the requirement for animal testing to determine the ocular irritation capacity of anti-microbial cleaning formulations. This evaluation is expected to begin in the 2007/2008 time frame.

2.1.3 Scientific basis for the test

Topical applications of chemicals can kill cells in several ways; among these are lysis of membranes, denaturation of proteins, saponification of lipids, and alkylation or other covalent interactions with macromolecules. The first three modes of action kill or damage very rapidly while the last may act rapidly but the evidence of the action may take some time to be manifested (Maurer, Parker et al. 2002). Certain chemical classes are

associated with these modes of action. Surfactants are primarily associated with membrane lysis although cationic surfactants may also act to precipitate proteins and other macromolecules. Organic solvents can act to delipidize and thus lyse membranes as well as denature (coagulate or precipitate) proteins. Acids tend to coagulate or precipitate proteins. Alkalis saponify lipids and denature proteins in a way that tends to allow them to penetrate into the cornea. Bleaches, peroxides, alkylators (e.g. mustards) bind to macromolecules (especially DNA) leading to cell death.

Damage to the eye is a function of the inherent cytotoxicity potential of the chemical or mixture, the effective concentration impacting the tissues and the residence time at that concentration on or in the tissues. The effective exposure is a combination of concentration and time of exposure (Figure 2.1.3.a). For example, a neat organic solvent may have a high cytotoxic potential but if it rapidly evaporates, the effective residence time will be less. Putting a large volume into a closed sac (e.g., lower conjunctival sac of the rabbit eye) will produce a very different effective exposure than a smaller amount placed (or accidentally splashed) onto the open surface of the cornea. Another solvent may have a longer residence time but have its cytotoxic potential rapidly reduced by dilution with tears. In this case, the irritation potential in a species with a low propensity to tear could show much more irritation than in a species with a high propensity to tear. The effective exposure to solids (powders) in the eye is a particular challenge. Powders placed into the conjunctival sac may have a residence time that ranges from minutes to a full day (and longer in some older studies) (Prinsen 2006). Traditional studies of eye irritation potential do not measure or control the effective exposure within or among studies. Thus, efforts to model exposure in alternative test systems are based on best estimates and approximations.

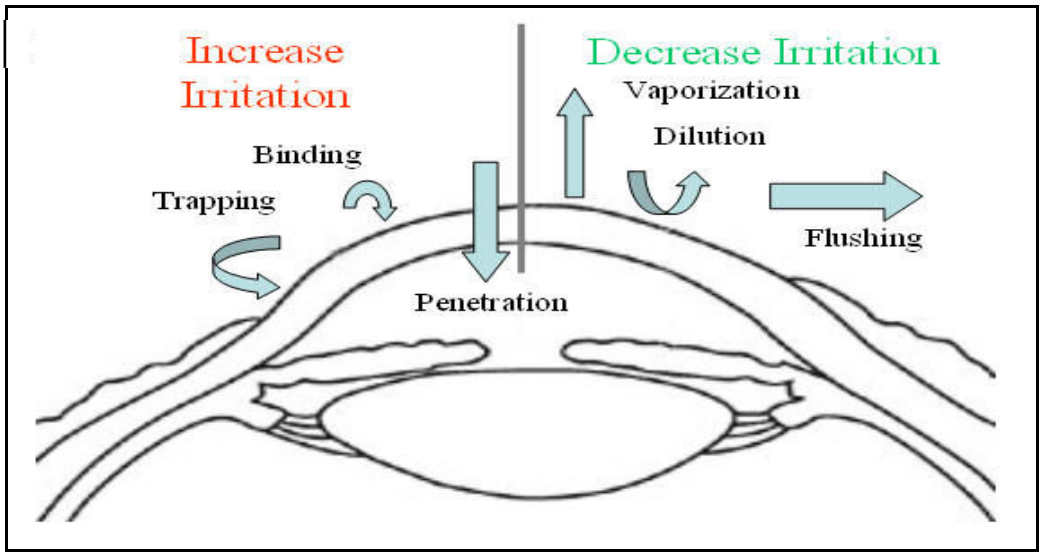

Figure 2.1.3.a Factors that impact exposure to the eye

Mechanistically, this cytotoxicity assay is intended to model the action of the surfactant on the cell membranes of the corneal and conjunctival epithelium where the test article would reside in an *in vivo* exposure. The potency of the surfactant (or surfactant

formulation) *in vivo* is related to the area and number of cell layers that can be lysed during the effective exposure period. More potent (and/or more substantive) surfactants will be more effective at a given concentration and exposure period. Potency can be a function of concentration (e.g., in a formulation) or chemical structure. Thus, a lower concentration of a more potent surfactant or more concentrated formulation would be required to lyse the membranes, and thus kill a given fraction of the cells in the epithelia (both corneal and conjunctival). Expressed another way, a given concentration of a more potent test material should lyse more cells (i.e., greater depth of penetration and injury). Initial depth of injury has been shown by Maurer, Jester, and collaborators (Jester, Petroll et al. 1998; Jester, Li et al. 2001; Maurer, Parker et al. 2002) to relate directly to the degree and duration of ocular injury (Figure 2.1.3.b). Their work has shown the relationship between cell initial killing and the resulting irritation. In the cytotoxicity assays with monolayer cells, a similar relationship between potency and effective concentration is expected for killing 50% of the target cell population (Harbell, Koontz et al. 1997).

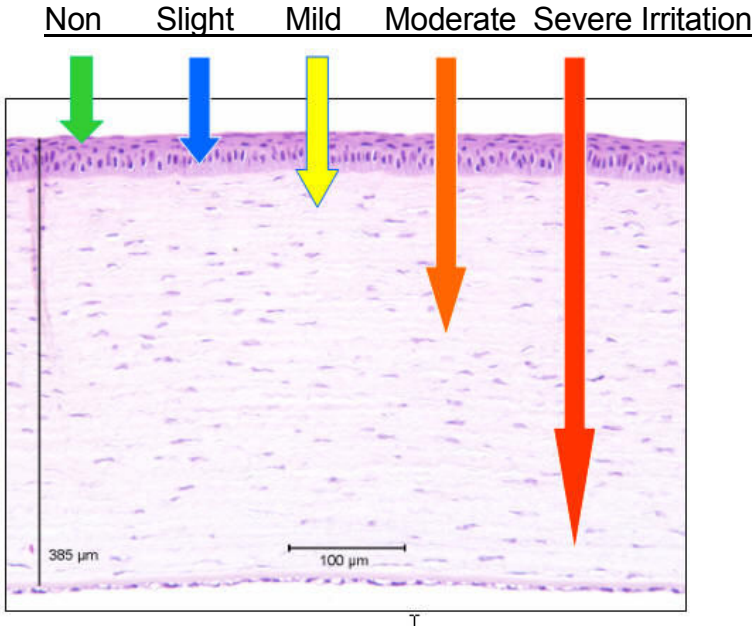

Figure 2.1.3.b Summary of the Depth of Injury Model

The CM estimates the metabolic rate (glucose utilization rate) of a population of cells by measuring the rate of excretion of acid by-products and resulting decrease in pH of the surrounding medium in an enclosed chamber. The rate of change in pH per unit time becomes the metabolic rate of the population. The basal metabolic rate and the ratio of glycolytic to aerobic metabolism (Krebs Cycle) may be different for different cell types. However, for the population of any one cell type, the ratio remains similar if the cells are handled in a consistent fashion. If a test material causes cytotoxicity to this population of cells it is assumed that the metabolic rate will fall. However, the metabolic rate may not fall immediately after exposure of the cells to a dilute concentration of toxicant. Populations of cells in culture are reported to metabolize glucose at only a fraction of their maximal metabolic rate (McConnell, Owicki et al. 1992). Thus, an up regulation of glucose metabolism can occur if the cells need energy to maintain their integrity in the face of a

mild biochemical insult. For example, exposure to a subcytotoxic concentration of surfactant can increase membrane leakage (to ions and water). This in turn can lead to an increase in the activity of ATP-dependent ion pumps and increased glucose metabolism. Thus early points in a killing curve can show increases in metabolic rate of 2- to 3-fold, but this metabolic rate then soon falls below 100% as higher concentrations of test material overwhelm the homeostatic controls within the cells (Figure 2.1.3.c).

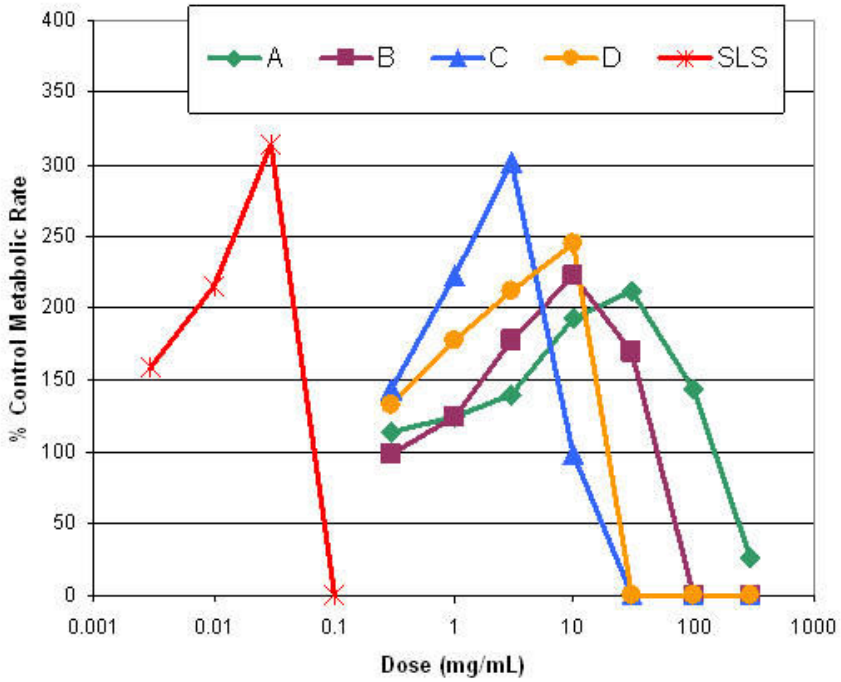

Figure 2.1.3.c Example of the metabolic rate data as a function of surfactant type and concentration

Although the metabolic rate is the physical parameter which is measured during the CM assay, the magnitude of metabolic rate itself is not directly related to eye irritation potential. Rather, the reduction of the metabolic rate to 50% of its basal rate is the parameter used to measure the impact of the test article on the test system (L929 cells in almost all cases). The CM assay exposes a population of cells to increasing concentrations of the test article (diluted in medium). The exposure follows a three step process where the first step is the exposure to the diluted test article, the second is the test article rinse-out and the third is the measurement of the metabolic activity. This means that the impact of the exposure is measured immediately and then a subsequent exposure is performed until the highest testable concentration has been used or the population of cells is severely damaged and the metabolic rate has declined to effectively zero. From the concentration response curve, the concentration that leads to a 50% decline in the metabolic rate of the population (the MRD_{50}) is calculated from the curve. The MRD_{50} values are used to compare test materials and provide a measure of ocular irritancy potential.

By current convention, the units of the MRD_{50} are mg/mL; however, many of the studies reviewed in this BRD presented data using related terminology, e.g. MRD_{50}

expressed in μg/ml or as the reciprocal of the MRD_{50}. For consistency we have converted all such values to MRD_{50} (mg/mL) and report them as such in this BRD.

For ease in understanding the mechanistic basis of the CM assay, a table (Table 2.1.3) has been compiled describing the events that are commonly considered to occur during eye irritation. Those events that are modeled (or are closely related) by the CM assay are indicated by a Y (yes) indication.

Table 2.1.3 Summary of events involved in chemical-induced eye irritation *in vivo.* Text in italics represents irreversible responses.

Events involved in chemical-induced eye irritation	Modeled by the CM assay?
Chemical interaction with tear film (Klyce and Beuerman 1988; Hackett and McDonald 1994)	N
Chemical binding to the conjunctival epithelium (Hogan and Zimmerman 1962; Hackett and McDonald 1994)	Y
Adhesion molecules compromised (Farquhar and Palade 1963; Van Meer, van Hof et al. 1992; Katahira, Sugiyama et al. 1997)	N
Corneal epithelium damage (Dua, Gomes et al. 1994)	Y
• Inhibition of receptor-mediated membrane transport (Dearman, Cumberbatch et al. 2003)	Y
• Compromise of cell membrane integrity of upper corneal epithelium (Dua, Gomes et al. 1994; Hackett and McDonald 1994; Maurer and Parker 1996)	Y
• Cell **membrane lysis** of all corneal epithelium layers (Hackett and McDonald 1994)	Y
Hydration of corneal stroma (Hackett and McDonald 1994)	N
Cross-linking *of proteins in corneal stroma (Butler and Hammond 1980; Eurell, Sinn et al. 1991; Chan and Hayes 1994)*	N
Erosion of corneal stroma (Baldwin, McDonald et al. 1973; Hackett and McDonald 1994; Maurer and Parker 1996)	N
Cell damage to corneal epithelium and limbus (Jacobs and Martens 1990; Wilhelmus 2001)	Partially
Dilation and increased lymphatic leakage from scleral vasculature (Hackett and McDonald 1994)	N
Stimulation of nerve endings, i.e. enhanced blinking, tearing (Chan and Hayes 1994)	N
Erosion of nerve endings in cornea and sclera (Butler and Hammond 1980; Klyce and Beuerman 1988; Araki, Ohahsi et al. 1994)	N
Duration of response, i.e. length of time cell responses deteriorate. Duration of response covers the effects of **reactive chemicals** which can cause **coagulation, saponification**, that are effects which develop and increase over time. (Hubert 1992; Maurer and Parker 1996)	N
Recovery from response, i.e. length of time for cell responses to return to control levels (Hubert 1992)	N

It can be seen that the CM assay most closely models some of the initial stages of interaction of an eye irritant with the cornea. The more distal occurrences in eye irritation such as gross tissue changes in the corneal stroma, and the recovery from the lesions, are not directly modeled. However, if the hypothesis of Jester, Mauer, and others that initial area and depth of injury is predictive of time to, and extent of, recovery, then the measurements made by the CM may have a relationship to recovery as well.

2.1.4 Similarities and differences of modes of action in the test method and the reference species

In the i*n vivo* rabbit eye test, the depth of damage to the tissue through subjective observations of tissue changes is being indirectly assessed. One might also postulate that a test substance is progressively diluted with time in the eye (tearing and blinking) and with increasing depth of penetration (in part by binding with the lipids themselves). A more potent surfactant irritant will penetrate (kill progressive layers of cells) more readily because its effective exposure (i.e., the concentration sufficient to lyse cells) will be maintained into the deeper layers (since a lower concentration of a potent surfactant will still effectively lyse cells). *In vitro*, the test substance is first diluted and then each dilution is assessed for its ability to kill the target population. The minimum concentration of test material capable of damaging a given fraction of the target cells is being determined. In some assays (e.g., neutral red release), each dilution is tested with a separate population of cells while in the Cytosensor, the cell population is exposed progressively to each increasing concentration. The test system (target cell population) is very uniform (cell number and distribution) and the period of exposure tightly controlled. Thus, it is possible to determine the concentration required to damage a given fraction of the cells and use that value to compare the “potency” of surfactant to that test system.

In vitro assays for eye irritation fall into three general categories based on the dynamic range of the test system and assay endpoint. Assays involving cells in monolayer or suspension culture generally have a relatively small dynamic range (range between all alive and all dead) since all the cells in the system are exposed at exactly the same time to exactly the same concentration of test material. Therefore, the test article is diluted over a series of concentrations and applied to one or more cultures of the chosen cell type for a set exposure period (dilution-based assays) in order to more easily differentiate between the toxicities of various test materials. The fluorescein leakage (transepithelial passage), neutral red uptake, neutral red release, CM, and red blood cell lysis assays are examples of dilution-based assays. In every case, the endpoint value that is used to predict ocular irritation potential is the concentration of test article that produces a measured change in the test system (e.g., concentration required to reduce the treated cell population viability to 50% of the negative control cell population viability).

Besides the general concepts mentioned above, the following background information should be considered in the overall design of any CM study:

Since the CM is a dilution-based assay, some common strengths and weaknesses of all dilution-based assays should be considered. Among the strengths of dilution-based assays are: 1) the ability to carefully control the test system (cultured cells), 2) a wide assay dynamic range (range of dilutions), 3) the possibility for machine scoring of the endpoint (often spectrophotometric or fluorometric readings), 4) generally good intra-assay consistency, and 5) a relatively low cost.

Among the potential weaknesses of dilution-based assays are: 1) the need to dilute the test article in a physiological aqueous medium, 2) the difficulty in modeling deep tissue penetration, and 3) the immediacy of the endpoint assessment (lack of time for delayed responses to be manifested).

Certain types of test articles are poor candidates for testing in dilution-based assays. Hydrophobic chemicals or formulations (creams, pastes, or lotions) may never really reach the test system. Use of intermediate solvents would change the normal distribution of the test substance and they are generally not used. Organic solvents would be subject to dilution in aqueous medium so that the test system would not necessarily be exposed to the delipidizing or dehydrating effect of the solvent. Acids and alkali materials would be partially or fully neutralized. In all cases, these effects of dilution would serve to alter the test material and potentially impact the prediction of irritation. In contrast, surfactants and surfactant formulations appear to be less impacted by dilution in aqueous medium, as long as there is the recognition that the critical micelle concentration (CMC) is important. Surfactants form micelles at higher concentrations which reduces the number of surfactant molecules available to react with the target tissue. If the in-use concentration of the surfactant is below the CMC and the *in vitro* test is conducted at dilutions above the CMC, then a possible underestimation of the toxicity of the material could occur since cytotoxicity would be reduced at the concentrations above the CMC. Reduction of the Draize score as higher concentrations of pure surfactants are placed in the rabbit eye is commonly seen when dose-response experiments are conducted *in vivo* (Dr. Edward Bueller, Hilltop Research, Personal Communication)

2.2 Test method protocols

Table 2.2 Major components of the protocols

Table 2.2.a Major components of the protocols

		(Bruner, Miller et al. 1991)	(Bruner, Kain et al. 1991)	Corporate Company # 2 (Catroux, Rougier et al. 1993))	Corporate (Company # 3, positive control data)
Critical protocol elements	Instrument	Silicon microphys-iometer	Silicon microphys-iometer	Silicon microphys-iometer	Cytosensor
	Cells	NHEK	NHEK	L929	L929
	Cell #/Confluency	90-95%	$1x10^3$/100nL vol	$3x10^5$/well	$6x10^5$/well
	Coverslip/transwell	Coverslip	Coverslip	Transwell	Transwell
	Duration of exposure	320 second	~320 second	400 seconds	810 seconds
Positive controls		Unknown	Unknown	SLS	SLS*
Negative controls		Medium	Medium	Medium	Medium
Vehicle controls		NA	NA	NA	NA
Benchmarks used		None	None	Yes, but coded	NA
Endpoint(s) measured		MRD_{50}	MRD_{50}	MRD_{50}	MRD_{50}
Prediction Model(s) applied		None	None	None	95% Confidence interval
Quality control criteria used		Unknown	Unknown	Unknown	Performance of the positive control
GLP compliance		No	No	No	Generally yes
Availability of a standardized SOP		No	No	Summary data	Yes
Limits of Use Described		Not stated	Not stated	Not stated	NA
Reference data available		Summary	Summary	Summary	NA

* A 10% solution of SLS in DI H_20 is diluted in low-buffered medium ($NaCO_3$-free DMEM with additional NaCl and 1 mM sodium pyruvate supplemented with 2 mM L-glutamine and 50 µg/mL gentamicin). Final concentrations in one-half log does from 0.003 mg/mL to 3.0 mg/mL are dosed to determine an MRD_{50} with a historical average of 0.0798 mg/mL.

Table 2.2.b Major components of the protocols

		(Bagley, Bruner et al. 1992)	(Bagley, Bruner et al. 1992)	Corporate (Company # 1)	Corporate (Company # 1)
Critical protocol elements	Instrument	Silicon Microphys-iometer	Silicon Microphys-iometer	Silicon Microphys-iometer	Cytosensor
	Cells	L929	L929	L929	L929
	Cell #/Confluency	90-100%	$3x10^5$/well	90-100%	$6x10^5$/well
	Coverslip/transwell	Coverslip	Transwell	Coverslip	Transwell
	Duration of exposure	500 second	500 second	500 seconds	810 seconds
Positive controls		SLS	SLS	SLS	SLS
Negative controls		Medium	Medium	Medium	Medium
Vehicle controls		NA	NA	NA	NA
Benchmarks used		None	None	Yes but coded	Yes but coded
Endpoint(s) measured		MRD_{50}	MRD_{50}	MRD_{50}	MRD_{50}
Prediction Model(s) applied		None	None	Corporate	Corporate
Quality control criteria used		Unknown	Unknown	Performance of the positive control	Performance of the positive control
GLP compliance		No	No	Generally yes	Generally yes
Availability of a standardized SOP		No	No	Yes	Yes
Limits of Use Described		Not stated	Not stated	Generally surfactant-based materials	Generally surfactant-based materials
Reference data available		Summary	Summary	Full	Full

Table 2.2.c Major components of the protocols

		CTFA Phase III (Gettings, Lordo et al. 1996)	EC/HO (Balls, Botham et al. 1995)	COLIPA (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999)	INVITTOX (IP-97)	INVITTOX (IP-102)
Critical protocol elements	Instrument	Silicon microphysiometer	Cytosensor	Cytosensor	Silicon microphysiometer	Cytosensor
	Cells	L929	L929	L929	L929	L929
	Cell #/Confluency	Confluent	~$6x10^5$/well	~$6x10^5$/well	90-100% confluency	5-$6x10^5$/well
	Coverslip/transwell	Coverslip	Transwell	Transwell	Coverslip	Transwell
	Duration of exposure	500 second	810 seconds	810 seconds	500 seconds	300 second
Positive controls		SLS	SLS	SLS	Unknown	None
Negative controls		Medium	Medium	Medium	Medium	Medium
Vehicle controls		NA	NA	NA	NA	NA
Benchmarks used		None	None	NA	None	None
Endpoint(s) measured		MRD_{50}	MRD_{50}	MRD_{50}	MRD_{50}	MRD_{50}
Prediction Model(s) applied		None	None	Yes	None	None
Quality control criteria used		Performance of the positive control	Performance of the positive control	Performance of the positive control; Audit of data by BIBRA QC Unit	Performance of the positive control	Unknown
GLP compliance		No	No	No	No	No
Availability of a standardized SOP		Yes	Yes	MA yes	Yes	Yes
Limits of Use Described		Surfactants only	Solubility in medium	Solubility in medium	pH ≤ 2.0 or pH ≥1 2.0	Solubility in medium
Reference data available		Yes (full data)	Yes (full data)	Yes (full data)		

2.2.1 Description of protocol components and rationale for differences

Overview of the methodology: The Cytosensor uses a low volume flow through chamber and a light-addressable potentiometer to measure the metabolic rate of a cell population. Metabolic rate is determined indirectly by the number of protons excreted into the low buffer medium (change in pH) per unit time. The light-addressable potentiometer forms the bottom of the flow through chamber and serves as a very sensitive and stable pH meter. While medium is flowing through the chamber, the pH is stable and governed by the medium. When the flow of medium is stopped, the pH begins to drop in a linear fashion over time. The actual change in pH during this measurement is generally less than 0.2 pH units.

Three common protocols developed by The Company # 1 were used in most of the studies reported in this BRD. The first two were based on the silicon microphysiometer (prototype instrument) and the third on the Cytosensor (commercial instrument). Most of the data in this BRD come from the third protocol. The two instruments differ primarily in the way the cells are introduced into the sensor chamber. The sensor chamber is composed of the light-addressable potentiometer sensor (sensor chip) on the bottom and ports for the medium (inlet and outlet). For the silicon microphysiometer, the cells were plated on metal coated glass coverslips and allowed to grow to confluence. The coverslip became the upper part of the low volume chamber with the cells on the downward side (facing the sensor chip). Medium then flowed between the coverslip and the sensor chip (Figures 2.2.1.a and 2.2.1.b). For the Cytosensor (and late modifications of the silicon microphysiometer), the cells are grown on a transwell membrane (discussed below). The whole transwell is placed into the sensor chamber and a plunger (with a spacer) pressed down on the membrane to seal it. There is a small medium-filled space between the sensor chip and the bottom of the transwell. The cells are attached to the top of the membrane so that the acid metabolites must pass through the membrane pores to reach the space in the lower part of the chamber. The medium is passed over the cells on the upper side of the membrane. The change from the coverslip system to the transwell system brought a change in the exposure, rinse, and read cycle as well. Figure 2.2.1.c shows the operating components of the instrument and Figure 2.2.1.d shows the low volume sensor chamber (transwell configuration). Based on the comparison of data generated in both the SM and CM, Company # 1 established a conversion algorithm so that all results generated initially from the SM can be compared to the results generated for the CM.

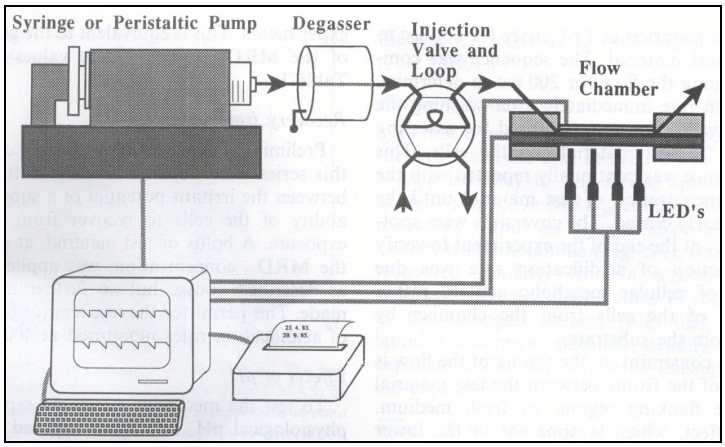

Figure 2.2.1.a Diagram of the operating components of the silicon microphysiometer (Bruner, Miller et al. 1991)

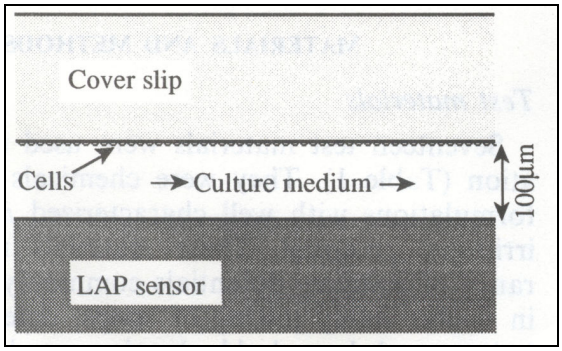

Figure 2.2.1.b The original silicon microphysiometer sensor chamber with the coverslip in place (Bruner, Miller et al. 1991)

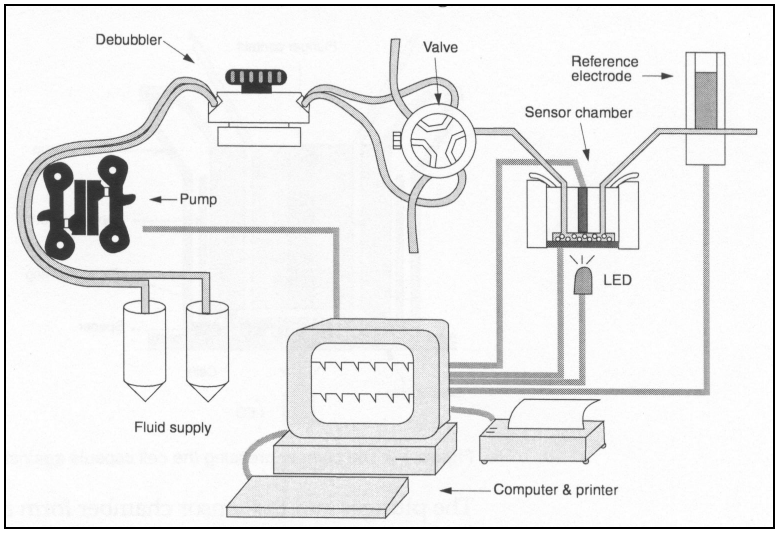

Figure 2.2.1.c Diagram of the operating components of the Cytosensor (Cytosensor Manual)

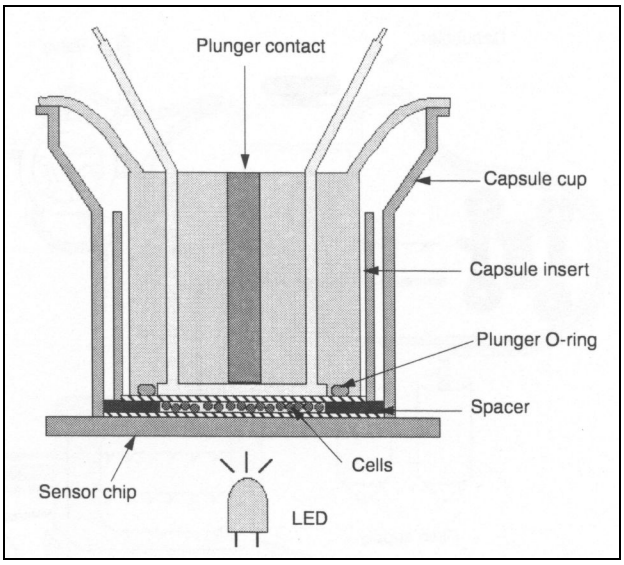

Figure 2.2.1.d The Cytosensor chamber with the transwell in place (Cytosensor Manual)

Originally, the silicon microphysiometer (coverslip chamber) used a 15-minute exposure, rinse, and read cycle. The cells were exposed to each concentration in two phases. In the first phase, the diluted test article was pumped (1.67 µL/sec) through the chamber for 120 seconds and then the flow halted for 200 seconds (total of 320 seconds of exposure). The chamber was then rinsed with fresh medium at the same rate for 380 seconds. The flow was then stopped for 200 seconds while the acidification rate was measured. This exposure protocol was used primarily on normal human epidermal keratinocytes (Bruner, Miller et al. 1991). Most of the studies in this BRD used L929 cells as the test system. The exposure protocol was altered so that the cells were exposed to the test article for a total of 500 seconds (300 seconds of flow and 200 seconds with the flow off), rinsed for 400 seconds, and the metabolic rate determined for 169 seconds. Flow was restarted with medium before the next dose was introduced. For the purposes of the BRD, this protocol will be used as the standard coverslip protocol. Because the valves were turned manually, the total cycle time was 1100 seconds. The Cytosensor (both the commercial instrument and the silicon microphysiometer with "Cytosensor-like" chambers used a 20-minute (1200-second) exposure, rinse, and read cycle. This is still the current protocol. The cells are exposed 810 seconds (100 µL per minute for one minute and 20 µL per minute for 12.5 minutes). The rinse cycle lasts for 6 minutes and the flow is 100 µL per minute. Finally, the flow is stopped for 25 seconds and the change in pH is measured. For the purposes of the BRD, this will be the standard transwell protocol (for either the converted silicon microphysiometer or the Cytosensor). In an early study, Company # 2 used the silicon microphysiometer with transwell protocol, but used a 400-second exposure cycle (Catroux, Rougier et al. 1993). In the study conducted by Bagley, Bruner et al. 1992 a 500-second exposure cycle was used when comparing the silicon microphysiometer (transwell) to the silicon microphysiometer (coverslip).

The bulk of the available data come from the transwell protocol using the 810-second exposure. These studies include the EC/HO, COLIPA, a large portion of the Company # 1 corporate database, and the positive control database from Company # 3.

The transwell was introduced by Company # 6 to allow more efficient introduction of the test system to the sensor chambers (including non-adherent cells in a gelatin matrix). However, this change limited the cell density and types of cells that could be used. The transwells have 3 micron pores that allow efficient communication between the upper surface of the membrane (with the cells) and the lower surface that faces the sensor itself. Confluent cell layers would interfere with this communication and so the cell density was reduced to a standard $6x10^5$ cells per well (seeded the day before use). The transwell uses a polycarbonate filter membrane that is less prone to interaction with test materials than other types of membranes but does not allow the human keratinocytes to attach. Thus, the L929 cells were selected because they would readily attach and were easy to grow in continuous culture. With the change to L929 cells, the SM exposure protocol was changed to 500 seconds. This is the protocol that was used in the Bagley, Bruner et al. 1992, and the Company # 1 corporate database. L929 cells were seeded to produce a confluent (or nearly confluent) cell layer on the coated coverslip. With the advent of the transwell sensor chamber, the 810-second exposure protocol was developed and this protocol was used in the EC/HO, COLIPA, the Company # 1 corporate database, and the Company # 3 positive control database.

2.2.1.1 Development of Conversion Algorithm between SM and CM

At the time that the SM was replaced with the CM by Company # 6, the Company # 1 sponsored a study to compare data obtained with the SM (coverslip protocol) for a set of 11 surfactant-containing materials with data obtained for the same materials with the CM (transwell protocol). The studies were carried out at a single laboratory (Company # 4). The raw data can be found in Annex F35. The testing protocol utilized a preliminary trial followed by at least three definitive trials. Data produced by the SM and CM are shown in Tables 2.2.1.1.a & b, respectively. It can be seen that the overall mean CV for each of the two methods is very similar (22.8% for the SM; 21.8% for the CM).

Following data collection from both instruments, the data were compared and the following equation was derived to translate SM coverslip data to CM transwell data:

Log_{10} (Cytosensor MRD_{50}) = 0.135 + 0.7753 x Log_{10} (Silicon Microphysiometer MRD_{50}).

A graph depicting the relationship between the SM and CM is given in Figure 2.2.1.1. The current standard Cytosensor protocol is attached in Annex A. Other protocols, including the INVITTOX protocols 97 and 102, are included in Annex A.

Table 2.2.1.1.a Silicon Microphysiometer data for 11 surfactant-containing materials from Company # 1 (See Annex F35)

Substance	Prelim*	Trial 1	Trial 2	Trial 3	Trial 4	Mean MRD_{50} (mg/mL)	SD	N	CV (%)
A	21.368	18.116	25.510	20.408		21.345	3.785	3	17.7
B	+	0.083	0.085	0.082		0.083	0.001	3	1.7
C	+	0.291	0.266	0.263		0.273	0.015	3	5.5
D	+	0.247	0.153	0.435	0.298	0.283	0.117	4	41.5
E	+	13.643	13.004	9.434		12.027	2.268	3	18.9
F	+	0.042	0.027	0.026		0.032	0.009	3	28.2
G	0.161	0.093	0.139	0.198		0.143	0.053	3	36.8
H	0.714	2.020	1.239	1.595		1.618	0.391	3	24.2
I	0.094	0.043	0.032	0.039		0.038	0.006	3	14.7
J	0.020	0.045	0.038	0.026		0.036	0.010	3	26.9
K	+	0.081	0.094	0.152		0.109	0.038	3	34.5
Mean									**22.8**
Median									**24.2**

* Not included in the mean calculation
+ Value not determined during assay

Table 2.2.1.1.b Cytosensor Microphysiometer data for 11 surfactant-containing materials from Company # 1 (See Annex F35)

Substance	Prelim*	Trial 1	Trial 2	Trial 3	Trial 4	Mean MRD_{50} (mg/mL)	SD	N	CV (%)
A	90.909	56.497	48.544	62.500		55.847	7.001	3	12.5
B	0.223	0.254	0.424	0.283		0.320	0.091	3	28.4
C	0.758	0.794	0.552	0.820		0.722	0.147	3	20.4
D	0.452	0.442	0.412	0.431		0.428	0.016	3	3.7
E	19.120	9.091	11.429	5.319		8.613	3.083	3	35.8
F	0.067	0.074	0.052	0.075		0.067	0.013	3	19.2
G	0.251	0.177	0.288	0.267		0.244	0.059	3	24.3
H	2.288	2.110	2.016	2.457		2.194	0.232	3	10.6
I	3.497	1.475	4.367	3.802		3.215	1.533	3	47.7
J	0.282	+	0.139	0.151	0.165	0.152	0.013	3	8.5
K	0.251	0.268	0.159	0.281		0.236	0.067	3	28.4
Mean									**21.8**
Median									**20.4**

* Not included in the mean calculation
+ Value not determined during assay

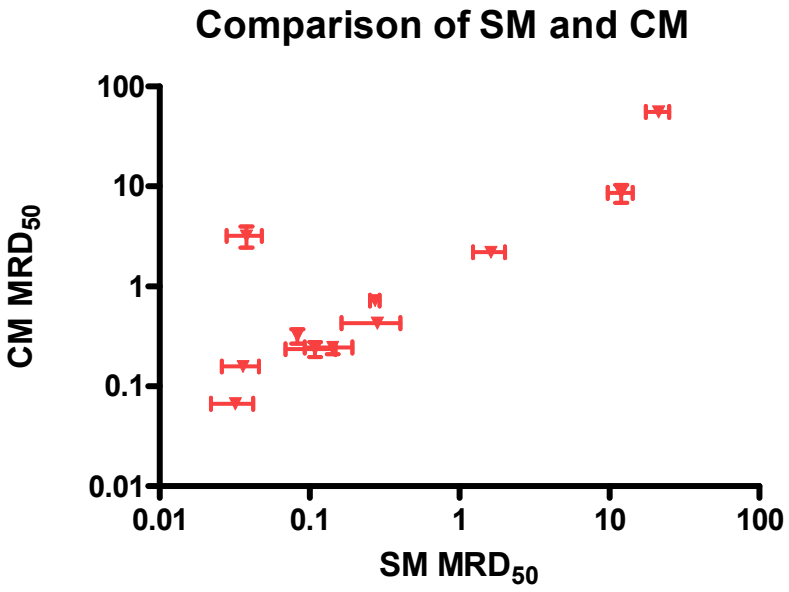

Figure 2.2.1.1 A comparison of data obtained from 11 surfactant-containing products with SM and CM.

2.2.2 Proposed critical components of the protocol that impact on reproducibility and/or predictive capacity of the assay

Because of the somewhat intricate nature of Cytosensor studies (due in a large part to the electro/mechanical complexity of the instrument), the initiator of any study conducted on the current CM machine according to protocols similar to the current protocol of Company # 3 (Annex A3) or INVITOX 97 or 102 (Annex A25 & A35) should be aware of the following Critical Protocol steps:

- Target Cells - The cell line of choice for ocular irritation assays is L929 cells. Although data from the use of other cell types, e.g. normal human keratinocytes, has been reported, virtually all safety studies conducted with the Cytosensor instrument (which uses a transwelll chamber rather than a coverslip) since the early 1990's have used L929 cells. Use of normal human keratinocytes with the transwell is not recommended.

- Instrument – The Cytosensor Microphysiometer manufactured by Company # 6, is the only instrument we are aware of that will produce the same type of data as are reported in this BRD. Results obtained with a predecessor instrument, the Silicon Microphysiometer, are contained in the BRD; however, the main prediction model described is only based on Cytosensor data.

- Capsule insert – This item must not be used in the assay. Although described in the user information which accompanies the Cytosensor, it has not been used for the vast majority of the studies reported here.

- Miscellaneous CM accessories – These should all be authorized by Company # 6. Because the Cytosensor is a complicated electronic instrument, the use of non-standard parts or accessories might result in completely aberrant readings.

- Medium – The medium should be low-buffered DMEM. The data provided by the Cytosensor are based on minute changes in pH which occur as a result of cellular metabolism. Use of fully buffered medium would essentially eliminate the ability to detect the level of pH changes necessary.

- Test material – A single phase solution/suspension should always be used. Non-dissolved material may clog the tubing or be blocked from entry into the chambers. In such an instance it would be difficult to show that the cells had been exposed to the desired concentration of test article.

- Exposure time – For standard safety assays the exposure time should be 810 seconds in order to match the experimental conditions for which the main prediction model was established. Longer or shorted exposure times will change the calculated MRD_{50} since toxicity is a function of exposure time.

- Baseline rates – It must always be ascertained that the machine and cells are stable before the experiment can begin because all subsequent data points are interpreted based on the baseline rate. In general this baseline should be between 50 and 200 microvolts/sec for 4 – 5 readings.

- Positive control – A positive control must be established in the user's laboratory so that there is assurance that Cytosensor is giving similar readings from day-to-day. Without this control it is impossible to compare the data for different test articles tested on different days.

- Cell density – Cells should be seeded at a density of ~6 X 10^5 cells/cup and then incubated for 16 – 32 hours under normal growth conditions before use. At the time of use the cells should be <80% confluent. Use of a fully confluent monolayer may interfere with communication between the upper and lower surfaces of the membrane, causing inaccurate readings.

- Confirmatory assay – It is strongly suggested that a confirmatory assay should always be conducted. Because the Cytosensor is an extremely sensitive instrument with a number of chambers, it has the potential to give spurious readings for some chambers which could skew the results. These anomalies might not be reflected in the performance of the positive control. Although this is rare, it is prudent to always confirm the response before reaching a safety conclusion.

2.2.3 List of studies with similar protocols

The Bagley, Bruner et al. 1992 study and the unpublished Silicon Microphysiometer study from Company # 1 were conducted with similar protocols. In Table 2.2.3.a, the critical protocol elements are listed. Company # 1 generally conducted their study according to GLP practices and with a positive control which was a quality control criterion for a valid test. Although Bagley, Bruner et al. 1992 did not conduct their study according to GLPs, the critical protocol elements remain the same as the Company # 1 study.

Table 2.2.3.a Silicon Microphysiometer studies with similar protocols

		Bagley, Bruner et al. 1992	Corporate (Company # 1)
Critical protocol elements	Instrument	Silicon Microphysiometer	Silicon Microphysiometer
	Cells	L929	L929
	Cell #/Confluency	90-100%	90-100%
	Coverslip/transwell	Coverslip	Coverslip
	Duration of exposure	500 second	500 seconds

The COLIPA (Brantom, Bruner et al. 1997), unpublished Cytosensor Microphysiometer study from Company # 1, positive control data from Company # 3, and EC/HO were conducted with similar protocols. In Table 2.2.3.b, the critical protocol elements are listed.

Table 2.2.3.b Cytosensor Microphysiometer studies with similar protocols

		COLIPA (Brantom, Bruner et al. 1997)	Corporate (Company # 1)	Corporate (Company # 3, positive control data)	EC/HO (Balls, Botham et al. 1995)
Critical protocol elements	Instrument	Cytosensor	Cytosensor	Cytosensor	Cytosensor
	Cells	L929	L929	L929	L929
	Cell #/Confluency	$6x10^5$/well	$6x10^5$/well	$6x10^5$/well	~$6x10^5$/well
	Coverslip/transwell	Transwell	Transwell	Transwell	Transwell
	Duration of exposure	810 seconds	810 seconds	810 seconds	810 seconds

2.2.4 Known applicability and limitations of the assay

The Cytosensor microphysiometer (CM) and its predecessor instrument the silicon microphysiometer (SM) have been in use with various toxicology protocols for over 15 years. Prediction of eye irritation was one of the first proposed uses (Parce, Owicki et al. 1989) and initial studies were conducted with a range of chemicals, e.g. solvents,

surfactants, alcohols, etc. As with any new technology, a range of chemical classes were evaluated to determine the applicability domain(s) and strengths and limitations of the assay (physical form, extremes of pH, etc). However, early on the major focus began to be placed on liquid soaps, detergent formulations, and household cleaning products (Bruner, Miller et al. 1991). Because of physical restrictions (the test material must be pumped into a chamber containing the test cells and then completely removed in the same fashion), the applicability domain has generally been considered restricted to test materials that are completely aqueous soluble. Currently, the assay is used primarily for evaluating the eye irritation potential of liquid surfactant-containing formulations/mixtures.

Because of the unique characteristics of the SM or CM instruments, the applicability domain is immediately limited to testing fully water soluble materials. The instrument functions by pumping the test material through a very small diameter hose onto the cells in the transwell. Any particulate matter that is present in the dosing solution could either clog the hoses - immediately ending the experiment - or might settle out on the transwell or on the cells themselves making it difficult, if not impossible, to determine the actual dose to which they were exposed. Table 2.2.4 describes the physiochemical properties of test materials and their compatibility with the CM assay for eye irritation.

Table 2.2.4 Physicochemical properties of test materials and their compatibility with the CM assay for eye irritation

Physicochemical Property	Is a material with this property compatible with the CM assay system? (based on solubility)
Fixative	No
Solvent	Yes, if aqueous soluble some solvents are testable
Extreme pH	No
Gases	No
Liquids	Yes, but must be aqueous soluble
Solid materials	Yes, if aqueous soluble, but cannot be tested in its solid form
Emulsions	No
Granular materials	No
Suspensions	No
Coloured materials	Yes
Toxicity affected by dilution	No
Highly viscous materials	No
Volatile materials	No
Reactive chemistries	No
Hydrophobic/lipophilic chemicals	No
Neat concentrations of chemicals	Yes, but a serial dilution will be performed with the neat concentration being the last dilution tested

2.2.5 Proposed prediction models

Historically, the only published prediction model for the CM is the one utilized in the COLIPA validation study ((Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999) which relates the MRD_{50} to the Draize Modified Maximum Average Score (MMAS) for

surfactant and surfactant-containing compounds. Expressed as an equation, the relationship is:

$$MMAS = A/1 + e^{B*(\log 10MRD50 - G)}$$

where A = 148.0, B = 1.813 and G = 2.329. This relationship was developed using information from the testing of 133 materials.

However, during the development of the current ECVAM-sponsored BRD, the management team requested that proposed prediction models based on classification criteria that are described by the Global Harmonized System, the EU classification system and the USEPA classification system be submitted prior to the final analysis of the predictive capacity of the data in the BRD.

Consequently we submitted three classification prediction models. The only existing PM that we were aware of that addressed categorical labeling was one under development for the USEPA as part of a project to develop non-animal labeling methods for anti-microbial cleaning products. That preliminary PM was:

MRD_{50} < 3 min = EPA I
80 min > MRD_{50} > 3 min = EPA III
MRD_{50} > 80 min = EPA IV

There were not sufficient data available to define a cut-off for EPA II materials, so by default these materials were included in the EPA I prediction interval and therefore overpredicted by one category.

The preliminary PM's for the GHS and EU system which were submitted were hypothesized using only a very early analysis of data and with knowledge of the only slightly better developed USEPA PM.

For the EU system the proposed PM was:

	$\underline{MRD_{50}}$
R41	<3 min
R36	<10 min; >3 min
Not Classified	>10 min

For the GHS system the proposed PM was:

	$\underline{MRD_{50}}$
1	<3 min
2A or 2B	<80 min; >3 min
No Category	>80 min

Following more detailed data analysis in the finalization of this BRD, *post hoc* evaluation of the data led to the following PM's on which the subsequent analysis of the predictive capacity of the CM assay for surfactants and surfactant containing formulations was based.

For the EU system the proposed PM is:

	MRD_{50}
R41	<2 min
R36	<10 min; >2 min
Not Classified	>10 min

For the GHS system the proposed PM was:

	MRD_{50}
1	<2 min
2A or 2B	<10 min; >2 min
No Label	>10 min

For the USEPA system the proposed PM was:

	MRD_{50}
I	<2 min
III	<80 min; >2 min
IV	>80 min

It can be seen that only slight modifications occurred when a more data were considered in the analysis. For all three classification systems the cut-off for the most severe labeling category was increased from <3 minutes to <2 minutes. This minor change appeared to fit the data distribution better without sacrificing sensitivity for the severe materials.

A second change was making the cut-off for GHS No Label materials >10 minutes rather than >80 minutes. This change was necessitated because we based our originally proposed classification on the data that had been developed for the USEPA classification work. However many materials which are EPA Category III materials fall into the lower GHS category (No Label) when evaluated by the GHS criteria. Thus the cut-off for the GHS No Label could be lowered to the less stringent >10 minutes.

It should be noted that the data analyzed to construct all three prediction models had fewer materials in the intermediate labeling categories (GHS 2A and 2B, EU R36, and EPA II and III) than in the severe or mild categories. In fact insufficient data were available to differentiate the two intermediate categories for either the GHS or USEPA system.

3. Within-laboratory reproducibility (Module 2)

The structure of almost all protocols analyzed in this BRD requires that SM or CM results for individual test materials be determined by averaging the results of at least three separate runs, where each run consists of testing multiple test article doses on a single cell culture. Replicate runs generally take place on a single day with a single operator, although they can occur on separate days with separate operators. Therefore, within-laboratory reproducibility for the SM or CM can be of two types: reproducibility in replicate cultures for a single test material done in multiple runs, generally by the same operator, and on the same day or within a few days (Type 1); and reproducibility of mean results for single test materials done in separate experimental settings and separated by longer time periods, perhaps of several years (Type 2). In the following analysis, the type of reproducibility being studied will be indicated.

3.1 Within-laboratory reproducibility for studies where raw data are available

Table 3.1 lists the various studies for which within-laboratory reproducibility information was available and describes the experimental parameters for each of the studies.

Table 3.1 Table presenting the results and relevant information pertinent to within-laboratory reproducibility for studies where raw data are available.

Studies	Results Reported			Replicates			Format of data (raw data, summary results, other)
	Intralaboratory Variability	Instrument	Number of test substances	No. of operators	No. of experiments	No. of replicates/ experiment	
The Company 1[1] (Bruner, Miller et al. 1991)	Yes; Mean, SEM and number of replicates	SM	17 of 17 tested	Unknown	1	N = 3-11	Summary; Mean ± SEM
Company # 2[2] (Catroux, Rougier et al. 1993)	Yes; Mean and SD – converted here to CV.	SM	18 of 21 tested	Unknown	1	3	Summary; Mean ± SD
EC/HO[3] (Balls, Botham et al. 1995)	No variability info. reported but calculated from raw data for this BRD	CM	35 of 60 tested	Unknown	1	≥ 3	Summary; Mean
CTFA[4] (Gettings, Lordo et al. 1996)	Yes; Mean, graphed SEM, and no. of replicates	SM	24 of 25 tested	Unknown	1	3 (20 materials), 4 (5 materials)	Summary; Mean
COLIPA[5] (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999)	No, but calculated for this BRD	CM	26 or 29 of 55 tested[6]	Unknown	1	≥ 3 for each of 2 labs	Raw data; Mean, SD and CV
Company # 4/Company # 3 positive control[5]	Not published	CM	1 of 1 tested	Multiple	629	1	Raw data; mean, SD and CV
20 chemicals from EC/HO and the COLIPA study run by same lab	No, but calculated for this BRD	CM	16 of 20 tested	Unknown	2	≥ 3	See above for individual studies
Company # 1/ Company # 3 comparison study for SM and CM	Results not published, but calculated for this BRD	SM & CM	11 of 11 tested	1	1	≥ 3	Raw data; Mean, SD & CV

[1] Performed using the silicon microphysiometer coverslip protocol with a 320-second exposure and normal human epidermal keratinocytes (see (Bruner, Miller et al. 1991)).
[2] Performed using the silicon microphysiometer transwell protocol with the 400-second exposure cycle and L929 cells.
[3] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[4] Performed using the silicon microphysiometer coverslip protocol with a 500-second exposure and L929 cells.
[5] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[6] One lab tested 26 materials while a second lab tested 29 materials.

3.2 Compilation of results

Eight studies or data sets (Table 3.1) were identified as providing information relevant to this analysis. Type 1 reproducibility (reproducibility of runs over a short time frame) information was obtained from seven data sets, although only summary information was obtained for the Bruner, Kain et al. 1991 study. Type 2 information (reproducibility of runs over long time frame) was obtained from the Company # 3 positive control data set which spanned over 12 years, and from a comparison between the mean results for 20 chemicals which were tested in both the EC/HO study and the COLIPA study. These two studies were conducted nearly two years apart.

3.2.1 Statistical approaches used: description & rationale for the approach used

Means, standard deviations (SD), and coefficients of variation (CV) were calculated (or in some cases transcribed directly from the manuscript) for the chemicals and formulations in all the studies except one (Bruner, Kain et al. 1991). For that study, the raw data could not be obtained, so only the mean and standard error of the mean (SEM) are reported as described in the manuscript.

CV values can sometimes be deceiving when used to analyze *in vitro* data which can range over several orders of magnitude, as is representative of the data in this BRD. Relatively small errors for small means can result in high CV's making the replicates appear to be variable. However, when the effect of these error ranges on the predicted value which may range over only two orders of magnitude (MAS for example) are examined, it can be seen that even an apparently large CV may have very little effect on the MAS value or hazard category that would be predicted.

3.2.2 Results and discussion

A presentation of the data and a discussion of the reproducibility it represents is contained in the following subsections. Where possible, the exact material tested is described to facilitate a more detailed analysis of the reasons for the lack of reproducibility where it occurs.

3.2.2.1 Company # 1study

The first Company #1 study reported (Bruner, Miller et al. 1991) compared a series of internal SM runs (3-11 per material) for 17 materials. The data were reported as average MRD_{50}; however, no SEM or SD were available. The data were also reported as average $pMRD_{50}$ ± SEM which cannot be easily converted to MRD_{50} ± SEM. The $pMRD_{50}$ is the negative logarithm of the average MRD_{50} in g/mL. The CV for these materials ranges from 0.6% to 16.9% which is considered to be quite low; however, it is based on log values so it cannot be directly compared with non-transformed data. The CV's of MRD_{50}'s expressed in log values will always be lower than the CV's of MRD_{50}'s expressed in normal concentration values.

Table 3.2.2.1 Within-laboratory reproducibility of SM from Bruner, Miller et al. 1991 using human keratinocytes and a 320 second exposure

Substance	MRD_{50} (mg/mL)	$pMRD_{50}$*	SD[1]	CV (%)[1]	Number of replicates
Surfactant Chemicals					
BAC (10%)	0.59	3.23	0.16	4.9	5
Sodium dodecylsulphate (40%)	0.55	3.26	0.03	0.9	3
Triethanolamine	17	1.77	0.3	16.9	11
Tween 20	17	1.77	0.22	12.4	5
Mean – Surfactant Chemicals				**8.8**	
Median – Surfactant Chemicals				**8.7**	
Surfactant Formulations					
Bar Soap 1	12	1.92	0.07	3.6	3
Bar Soap 2	7.8	2.11	0.22	10.4	6
Fabric Cleaner	18	1.74	0.21	12.1	3
Hand Soap	2.1	2.68	0.22	8.3	5
Hard Surface Cleaner 1	3.2	2.49	0.18	7.2	5
Hard Surface Cleaner 2	5.7	2.24	0.03	1.3	3
Heavy Duty Dishwashing Liquid	0.25	3.55	0.15	4.2	6
Heavy Duty Laundry Detergent	0.17	3.78	0.16	4.2	5
Light-Duty Dishwashing Liquid	0.69	3.16	0.02	0.6	4
Shampoo 1	0.77	3.11	0.09	2.9	5
Shampoo 2	0.61	3.21	0.02	0.6	3
Shampoo 3	0.39	3.41	0.1	2.9	3
Shampoo 4	1	3	0.3	10	11
Mean – Surfactant Formulations				**5.3**	
Median – Surfactant Formulations				**4.2**	
Mean – All Materials				**6.1**	
Median – All Materials				**4.2**	

* $pMRD_{50}$ is the negative log of the MRD_{50} in g/mL
[1] The SD and CV are based on the $pMRD_{50}$ values

3.2.2.2 Company # 2 study

Company # 2 reported a 53 chemical and product study (Catroux, Rougier et al. 1993) using the SM and L929 cells with three replicates for each material. Means and SD were available for 20 of the 21 surfactants. However, two of the surfactants were very non-toxic giving an average MRD_{50} of greater than 100 mg/ml. Because some of the individual data points for these two materials were censored (i.e. reported as “less than” or “greater than” a specified value), no accurate CV could be calculated. Therefore, only data for 18 of the 20 materials are reported in Table 3.2.2.2. The standard deviations were supplied to us by the authors from Company # 2. We did not calculate the SD’s, but we did calculate the CV’s using the values fromCompany # 2. The CV’s for the 18 materials ranged from 5.7% to 64.6%. The test materials were all pure surfactants or surfactant blends. The mean CV was 23.0%.

Table 3.2.2.2 Within-laboratory reproducibility from the Company # 2 study of Catroux, Rougier et al. 1993 which used the SM with a 400 sec exposure time. See Annex H3.

Substance	MRD_{50} (mg/ml)	SD	CV (%)	Number of replicates
Surfactant Chemicals				
1,2-dodecanol (etherified)	0.851	0.05	5.7	3
Acylamine polyglycol ethersulfate (genapol AMS)	5.357	1	18.7	3
Ammonium laurylsulphate	0.129	0.017	13.1	3
Blend of decanol and dodecanol (both etherified)	1.288	0.8	46.6	3
Blend of sodium and magnesium laurylethersulfate	0.871	0.15	17.6	3
Cocobetain derivative	0.263	0.17	64.6	3
Coprah amphoteric alky limidazolium dicarboxylate (miranol)	0.575	0.08	13.7	3
Dodecanol (etherified)	0.468	0.07	15.1	3
Hexadecyltrimethylammonium bromide (CTAB)	0.04	0.011	27.5	3
Industrial Tween 20	7.08	1.6	23.1	3
Octylphenoxy polyethoxy ethanol (Triton X100)	0.068	0.013	19.1	3
Polyoxyethylene sorbinate monolaurate (Tween 20)	20.89	7	34	3
Pyridinium cetyl bromide	0.105	0.02	19	3
Sodium lauryl sulphate (SDS)-A	0.079	0.022	27.8	3
Sodium laurylethersulphate	0.126	0.025	19.8	3
Sodium laurylsarcosinate	0.229	0.045	19.9	3
Tetradecyltrimethylammonium bromide (MTAB)	0.043	0.004	9.3	3
Triethanolamine laurylsulphate	0.056	0.011	19.7	3
Mean			**23.0**	
Median			**19.4**	

3.2.2.3 CTFA Phase III study

In 1992, CTFA began a study of the ability of a number of *in vitro* test methods to predict eye irritation potential of a series of surfactant-based personal care formulations. The SM was one method studied, and it was used in only one laboratory – Company # 4 Three replicates were conducted on each test material. The replicate assays may have occurred on the same day or within a matter of a few weeks at the most. All of the SM assays were conducted between 28 July 1992 and 17 September 1992 (see Annex F3) at Company # 4 the instrument used was the SM using L929 cells grown on a coverslip and a 300 second exposure time.

The results presented in Table 3.2.2.3.a show CV's ranging from 1.8% to 61.4% for the 25 products. There was no clear pattern as to which types of materials caused the greater variability. For example, four of the twelve shampoos (Shampoos 2, 5, 7, and 8) had four of the five lowest CV's, while two other shampoos (Shampoo 1 and Shampoo AntiD) had the two highest CV's. However, it was interesting to note that both of the more variable shampoos had the lowest water content (14% and 27%, respectively) of any of the 25 tested formulations. A graphical representation *in vitro* vs. *in vivo* data in Figure 3.2.2.3 showed that the SM CV's were of a similar magnitude to those of the *in vivo* test (Gettings, Lordo et al. 1996).

Data from multiple runs of the positive control (SLS starting concentration of 10% solution; Annex F2) conducted during this study provide another measure of within laboratory reproducibility. Table 3.2.2.3.b lists the eighteen results with the positive control (one positive control was run each day that a batch of test materials were run) over a nearly two month period. The CV for the 18 runs was 9.6%.

Table 3.2.2.3.a Within-laboratory reproducibility of SM from archived data originally obtained at Company # 4 and created for the CTFA Phase III study (Gettings, Lordo et al. 1996). The protocol utilized the SM and a 500 second exposure. Data from Annex F3.

Cytosensor Identification	Substance	Mean SM MRD_{50} (mg/mL)	SD	CV (%)	Number of replicates
Surfactant Formulations					
PGB-1	Baby Shampoo 1	2.13	0.52	24.5	3
PGH-1	Baby Shampoo 2	1.13	0.41	36	3
PGC-1	Bubble bath	0.65	0.24	36.9	3
PGQ-1	Cleansing Gel	6.54	2.05	31.3	3
PGF-1	Eye Makeup remover	32	9.79	30.6	3
PGT-1	Facial Cleaner	>500	0	Not Calculated	3
PGD-1	Facial Cleansing Foam	6.19	2.75	44.4	4
PGS-1	Foam Bath	0.75	0.14	18.2	3
PGZ-1	Gel Cleanser	2.99	0.95	31.7	3
PGI-1	Hand Soap	5.13	1.89	36.9	4
PGR-1	Liquid Soap 1	2.53	0.07	2.8	3
PGX-1	Liquid Soap 2	2.34	0.38	16.2	3
PGU-1	Mild Shampoo	7.31	1.14	15.6	3
PGN-1	Polishing Scrub	55.8	24.2	43.3	4
PGV-1	Shampoo 1	1.35	0.8	59.4	4
PGK-1	Shampoo 2	0.85	0.05	5.4	3
PGP-1	Shampoo 3	2.89	0.74	25.5	3
PGO-1	Shampoo 4	2	0.74	36.9	3
PGM-1	Shampoo 5	2.51	0.05	1.8	3
PGG-1	Shampoo 6	2.25	0.38	17.1	3
PGL-1	Shampoo 7	0.83	0.02	2.3	3
PGE-1	Shampoo 8	2.53	0.1	4	3
PGJ-1	Shampoo AntiD	0.79	0.48	61.4	4
PGW-1	Shower Gel	0.79	0.07	9	3
PGY-1	Skin Cleaner	0.75	0.1	13.3	3
Mean				**25.2**	
Median				**25.0**	

Table 3.2.2.3.b Within-laboratory reproducibility of the positive control at Company # 4 during the CTFA Phase III study (Gettings, Lordo et al. 1996). Data from Annex F3.

Date	SM MRD_{50} (mg/mL)
7/28/92	0.0650
7/29/92	0.0802
8/3/92	0.0861
8/4/92	0.0869
8/5/92	0.0806
8/10/92	0.0746
8/11/92	0.0881
8/12/92	0.0736
8/17/92	0.0863
8/18/92	0.0906
8/19/92	0.0904
8/24/92	0.0803
8/25/92	0.0838
8/26/92	0.0857
9/14/92	0.0921
9/15/92	0.102
9/16/92	0.0856
9/17/92	0.0845
Mean	**0.0842**
Median	**0.0857**
SD	**0.0081**
CV (%)	**9.60**

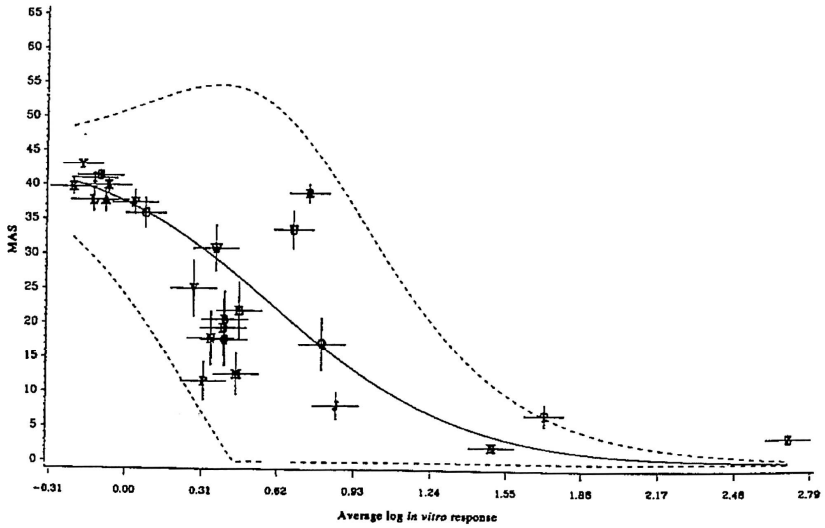

Figure 3.2.2.3 Regression plot of the relationship between *in vitro* endpoint and MAS for the SM; dashed curves represent 95% prediction bounds (Gettings, Lordo et al. 1996). Error bars represent one standard error.

3.2.2.4 EC/HO study

The EC/HO study of *in vitro* methods for eye irritation was designed to have three or more laboratories using each *in vitro* method. Four laboratories used the CM instrument with transwells and an 810 second exposure time. However, since only submission of mean CM values for each test material was requested by the Management Team, it has been difficult to obtain data from individual runs. At this point, raw individual run data for only 35 materials (the only ones determined by Company # 4 to be compatible with the CM) from one laboratory (Company # 4) have been located in the archives of the Institute For In Vitro Sciences. Only 32 of the 35 data points could be unequivocally linked to a specific test material. These data are presented in Table 3.2.2.4.

The raw data mean was compared against the published mean (SM31) to decode the individual raw data points (Annex G and H3). The mean MRD_{50} presented in the tables below was based on the average of transformed individual trial data ($rMRD_{50} \rightarrow MRD_{50}$), not the transformation of the mean log MRD_{50} which is presented in subsequent tables in this report. Values not used to calculate the reported mean are the preliminary (range finding) assay results which are used to narrow the dilution scheme and which in general practice is not used to calculate the mean. In addition, any other assays run (assays B - D) are assays in which an MRD_{50} could not be calculated because the cells did not die, or the assay was terminated early.

Table 3.2.2.4 Within-laboratory reproducibility of CM from archived data that was originally obtained at Company # 4 for the EC/HO study (Balls, Botham et al. 1995). The protocol utilized the CM using transwells and an 810 second exposure time. At least triplicate runs were performed. Raw data can be found in Annex H3. Surfactants are highlighted. N=35.

			Values used to calculate the reported Mean MRD_{50} (mg/mL)					Values not used to calculate the reported Mean MRD_{50} (mg/mL)					Results using the values used to calculate the reported mean				Results using all values other than > or < values to calculate mean[a]			
Sponsor Designation	Chemical	Formulation Type	Assay 1	Assay 2	Assay 3	Assay 4	Assay 5	Range Finding Assay	Assay A	Assay B	Assay C	Assay D	Average MRD_{50} (mg/mL)	N	SD	CV (%)	Average MRD_{50} (mg/mL)	N	SD	CV (%)
2376	2,5-Dimethylohexanediol	SO		151.36	151.36	165.96		>5.00					156.22	3	8.43	5.4	156.22	3	8.43	5.4
207	Acetone	SO		144.54	165.96	114.82		>5.00					141.77	3	25.68	18.1	141.77	3	25.68	18.1
2162	Ammonium nitrate	Other		169.82	125.89	144.54		>5.00					146.75	3	22.05	15.0	146.75	3	22.05	15.0
1373	Benzalkonium chloride 10%	SU			0.48	0.43	0.50	0.35	>3.50	<3.00			0.47	3	0.04	8.2	0.44	4	0.07	15.6
527	Benzalkonium chloride 5%	SU		0.95	0.98	1.38		3.01					1.10	3	0.24	21.7	1.58	4	0.97	61.5
3090	Benzalkonium chloride [1]/[2] 1	SU		5.01	3.89	7.08		3.02					5.33	3	1.62	30.4	4.75	4	1.75	36.9
1223	Cetylpyridinium bromide 10%	SU		0.98	1.15	0.93		3.09					1.02	3	0.11	11.1	1.54	4	1.04	67.6
1383	Cetylpyridinium bromide 6%	SU		1.26	0.72	2.69		1.20					1.56	3	1.02	65.3	1.47	4	0.85	57.8
2859	Cetylpyridinium bromide 0.1%	SU		128.82	91.20	91.20		>5.00					103.74	3	21.72	20.9	103.74	3	21.72	20.9
2653	Ethanol*	SO		162.18	67.61	147.91		>5.00					125.90	3	50.98	40.5	125.90	3	50.98	40.5
467	Ethyl acetate	SO				53.70		>4.70	>4.75	>4.75			53.70	1			53.70	1		
1719	Gammabutyrolactone	Other		93.33	144.54	112.20		>5.00					116.69	3	25.90	22.2	116.69	3	25.90	22.2
3111	Glycerol	SO		218.78		131.83	204.17	>5.00		>5.38			184.93	3	46.56	25.2	184.93	3	46.56	25.2
1666	Imidazole*	SU			23.44	22.39	23.44	24.55	>5.25				23.09	3	0.61	2.6	23.45	4	0.88	3.8
389	Isobutanol	SO		29.51	28.84	27.54		4.60					28.63	3	1.00	3.5	22.62	4	12.04	53.2
398	Isopropanol	SO		89.13	93.33	91.20		4.74					91.22	3	2.10	2.3	69.60	4	43.27	62.2
1581	L-Aspartic acid	AC		0.85	2.63	0.71		2.75					1.40	3	1.07	76.7	1.74	4	1.11	63.8
372	Methyl acetate	SO		89.13	93.33	93.33		4.98					91.93	3	2.43	2.6	70.19	4	43.52	62.0
408	Methyl ethyl ketone	SO		53.70	45.71	52.48		4.82					50.63	3	4.31	8.5	39.18	4	23.17	59.1
2582	Potassium cyanate	Other		50.12	30.90	30.20		26.92					37.07	3	11.30	30.5	34.53	4	10.53	30.5
1269	Promethazine HCl	Other		1.66	1.48	1.32		1.02					1.49	3	0.17	11.5	1.37	4	0.27	19.7
2380	Pyridine	Other		28.84	30.20	30.90		21.88					29.98	3	1.05	3.5	27.96	4	4.14	14.8
1016	Sodium hydroxide	AL		0.87	2.24	2.09		7.24					1.73	3	0.75	43.3	3.11	4	2.82	90.7
535	Sodium hydroxide	AL		14.13	26.92	8.32	21.88	4.47					17.81	4	8.23	46.2	15.14	4	9.29	61.4
2232	Sodium lauryl sulfate 3%	SU		3.02	2.88		3.24	3.09					3.05	3	0.18	5.8	3.06	4	0.15	4.8
363	Sodium lauryl sulfate 15%	SU		0.81	0.54	0.49		2.95					0.61	3	0.17	28.5	1.20	4	1.18	98.3
1721	Thiourea	Other		50.12	48.98	53.70		>5.00					50.93	3	2.47	4.8	50.93	3	2.47	4.8
1148	Trichloroacetic acid	AC		1.82	5.13	1.62		3.09					2.86	3	1.97	69.0	2.92	4	1.61	55.3
452	Trichloroacetic acid	AC		15.85	9.77	16.98		3.93					14.20	3	3.88	27.3	11.63	4	6.03	51.9
1037	Triton X-100 10%	SU		1.35	2.57	2.19		2.95					2.04	3	0.62	30.7	2.26	4	0.69	30.3
2482	Triton X-100 [1]/[2] 5%	SU		4.57	2.82	3.02		3.16					3.47	3	0.96	27.6	3.39	4	0.80	23.5
2767	Tween 20	SU		7.76	4.79	4.57		7.41					5.71	3	1.78	31.3	6.13	4	1.69	27.5
1384	[b]		not irr. at sol. (4.00)					>4.00	>4.00	>4.00	>4.00									
1769	[b]		not irr. at sol. (4.00)					>4.00	>4.00	>4.00	>4.00									
3274	[b]		not irr. at sol. (5.50)					>5.00	5.02	>5.50	>5.50	>5.50					5.02			
	Mean															**23.9**				**38.9**
	Median															**21.7**				**30.5**

[a] mean calculated using the values from assays 2-5, plus range finding assay if an actual value was obtained, i.e. < and > values were omitted.
[b] the identity of these chemicals is unknown
AC - Acid, AL - Alkaline, SO - Solvent, SU - Surfactant

3.2.2.5 COLIPA study

The COLIPA study used a set of 55 cosmetic formulations and ingredients - a large proportion of which were pure surfactants or surfactant based formulations - to assess the ability of *in vitro* methods to predict eye irritation potential. Two laboratories conducted the CM assay according to a standardized protocol (Annex A) which used an 810 second exposure time. Raw data from the studies conducted by Company # 4 and Company # 5 were obtained from the archives of Company # 3. These data are presented in Tables 3.2.2.5.a & 3.2.2.5.b (Company # 4) and Tables 3.2.2.5.c & 3.2.2.5.d (Company # 5). Raw data are found in Annexes F13 & H29.

CV's ranged from 1.3% to 55.9% (29 materials) for Company # 4 and from 1.0% to 59.4% (26 materials) for Company # 5. The mean CV for Company # 4 was 18.5% and for Company # 5 was 13.3%. The pump deodorant was the only material that seemed highly variable in both labs (CV of 48.5% [Company # 4] and 59.4% [Company # 5]). A second level of analysis was conducted to determine if there was a difference in reproducibility between surfactant and non-surfactant materials. Tables 3.2.2.5.a & 3.2.2.5.b (Company # 4) and Tables 3.2.2.5.c & 3.2.2.5.d (Company # 5) show that in both cases the variability of the surfactant materials (CV of 21.0% [Company # 4] and 14.9% [Company # 5]) was somewhat greater than the non-surfactants (CV of 13.8% [Company # 4] and 10.5% [Company # 5]).

The CV of each material for each lab was plotted against the mean MRD_{50} in Figure 3.2.2.5.a to determine if there was a relationship between the level of toxicity of a test material and its variability. The graph seems to show that the highest variability is associated with the mid-level of toxicity and that it tapers off to lower levels at the higher and lower end of the toxicity scale.

In addition, a comparison of the CV's for each material tested by both labs was made in Figure 3.2.2.5.b. If, in general, the composition of the test materials was directly related to the amount of variability found in the testing, then the within laboratory variability would be similar for each laboratory doing the testing. To test this hypothesis we graphed the CV's found in each lab against each other. It can be seen that there is a very low correlation (r^2= 0.3025) between the CV's of the 2 labs indicating that, as a general case, the composition of materials themselves did not significantly contribute to the variability. One specific exception might be the pump deodorant where between run variability was the highest (CV of 48.5% [Company # 4] and 59.4% [Company # 5]) of any of the test materials in both laboratories.

Table 3.2.2.5.a Surfactant Materials – COLIPA Intralaboratory variability of CM from archived Company # 4 data created for the COLIPA study for surfactant materials (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999). The protocol utilized L929 cells and an 810 second exposure. N=8 for surfactant formulations. N=12 for surfactant chemicals. Data from Annex F13.

Sample ID	Substance	MRD_{50} (mg/ml)	SD	CV (%)	Number of replicates
Surfactant Formulations					
2337	Eye make-up remover	87.77	1.17	1.3	3
3645	Gel cleaner	5.68	2.37	41.8	3
3343	Liquid soap #1	0.88	0.03	3.5	3
2429	Pump Deodorant	19.35	9.38	48.5	3
3105	Shampoo - baby	2.51	0.96	38.1	3
2092	Shampoo #1 normal	0.75	0.21	28.7	3
3213	Skin cleaner	0.63	0.1	16.3	3
Mean – Surfactant Formulations				**25.5**	
Median – Surfactant Formulations				**28.7**	
Surfactant Chemicals					
3399	Benzalkonium chloride 1%	4.11	0.89	21.6	3
2174	Benzalkonium chloride 10%	0.32	0.07	21	3
3770	Benzalkonium chloride 5%	0.81	0.1	12.7	3
3886	Cetylpyridinium bromide 6%	1.36	0.2	14.5	3
3589	Polyethylene glycol 400	296.5	34.17	11.5	3
2721	SLS 15%	0.52	0.02	3.5	3
2089	SLS 3%	3.23	0.65	20.2	3
2079	SLS 30%	0.31	0.02	5.8	3
3740	Triton X-100 1%	21.17	4.21	19.9	3
3244	Triton X-100 10%	2.47	0.57	23	3
3806	Triton X-100 5%	4.66	0.52	11.1	3
2171	Tween 20	9.50	5.31	55.9	3
Mean – Surfactant Chemicals				**18.4**	
Median – Surfactant Chemicals				**17.2**	
Mean – All Surfactant Materials				**21.0**	
Median – All Surfactant Materials				**19.9**	

Table 3.2.2.5.b Non-Surfactant Materials – COLIPA Intralaboratory variability of CM from archived Company # 4 data created for the COLIPA study for non-surfactant materials (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999). The protocol utilized L929 cells and an 810 second exposure. N=2 for non-surfactant formulations. N=8 for non-surfactant chemicals. Data from Annex F13.

Sample ID	Substance	MRD_{50} (mg/ml)	SD	CV (%)	Number of replicates
Non-Surfactant Formulations					
2115	Hair styling lotion	164.82	7.98	4.8	3
3525	Mouthwash	37.84	3.55	9.4	3
Mean – Non-Surfactant Formulations				**7.1**	
Median – Non-Surfactant Formulations				**7.1**	
Non-Surfactant Chemicals					
3453	Glycerol	214.83	25.35	11.8	3
2056	Imidazole	18.84	5.52	29.3	3
2356	Isopropanol	52.59	17.2	32.7	3
3870	Methyl ethyl ketone	54.18	3.16	5.8	3
3872	Propylene glycol	265.07	3.54	1.3	3
3524	Sodium hydroxide 1%	9.09	1	11	3
3631	Sodium hydroxide 10%	4.33	0.15	3.5	3
3148	Trichloroacetic acid 30%	1.12	0.31	28.1	3
Mean – All Non-Surfactant Chemicals				**15.4**	
Median – All Non-Surfactant Chemicals				**11.4**	
Mean – All Non-Surfactant Materials				**13.8**	
Median – All Non-Surfactant Materials				**10.2**	

Table 3.2.2.5.c Surfactant Materials – COLIPA Within-laboratory reproducibility of CM from archived Company # 5 data created for the COLIPA study for surfactant materials (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999). The protocol utilized L929 cells and an 810 second exposure. N=8 for surfactant formulations. N=10 for surfactant chemicals. Data from Annex H29.

Sample ID	Substance	MRD_{50} (mg/ml)	SD	CV (%)	Number of replicates
Surfactant Formulations					
2201	Eye make-up remover	99.31	1	1	3
3677	Gel cleaner	5.47	1.2	22	3
2254	Hair styling lotion	292.01	6.07	2.1	3
2003	Pump Deodorant	47.74	28.34	59.4	3
3306	Shampoo – baby	2.15	0.73	33.7	3
3440	Shampoo #1 normal	0.72	0.06	8.1	3
3328	Skin cleaner	0.76	0.05	6	3
2386	Liquid Soap #1	0.68	0.10	14.0	3
Mean – Surfactant Formulations				**18.3**	
Median – Surfactant Formulations				**11.1**	
Surfactant Chemicals					
3517	Benzalkonium chloride 1%	4.33	1.19	27.4	3
2901	Benzalkonium chloride 10%	0.31	0.05	16.4	3
2811	Benzalkonium chloride 5%	1.38	0.12	8.9	3
3825	Polyethylene glycol 400	>316.23	ND	ND	2
3191	SLS 15%	0.51	0.02	3.3	3
2136	SLS 3%	2.78	0.07	2.7	3
3207	Triton X-100 1%	16.79	0.73	4.3	3
3720	Triton X-100 10%	1.24	0.28	22.9	3
3500	Triton X-100 5%	2.42	0.07	2.7	3
3561	Tween 20	3.49	0.62	17.7	3
Mean – Surfactant Chemicals				**10.6**	
Median – Surfactant Chemicals				**6.6**	
Mean – All Surfactant Materials				**14.9**	
Median – All Surfactant Materials				**8.9**	

Table 3.2.2.5.d Non-Surfactant Materials – COLIPA Within-laboratory reproducibility of CM from archived Company # 5 data created for the COLIPA study for surfactant materials (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999). The protocol utilized L929 cells and an 810 second exposure. N=2 for non-surfactant formulations. N=7 for non-surfactant chemicals. Data from Annex H29.

Sample ID	Substance	MRD_{50} (mg/ml)	SD	CV (%)	Number of replicates
Non-Surfactant Formulations					
2254	Hair styling lotion	292.01	6.07	2.1	3
2563	Mouthwash	46.85	9.2	19.6	3
Mean – Non-Surfactant Formulations				10.9	
Median – Non-Surfactant Formulations				**10.9**	
Non-Surfactant Chemicals					
2479	Glycerol	208.7	3.06	1.5	3
3337	Imidazole	26.03	0.99	3.8	3
3789	Isopropanol	124.51	25.26	20.3	3
3556	Propylene glycol	218.86	7.59	3.5	3
3357	Sodium hydroxide 1%	13.59	5.11	37.6	3
3434	Sodium hydroxide 10%	0.6	0.01	1.9	3
3864	Trichloroacetic acid 30%	1.24	0.05	4.2	3
Mean – Non-Surfactant Chemicals				10.4	
Median – Non-Surfactant Chemicals				**3.8**	
Mean – All Non-Surfactant Materials				10.5	
Median – All Non-Surfactant Materials				**3.8**	

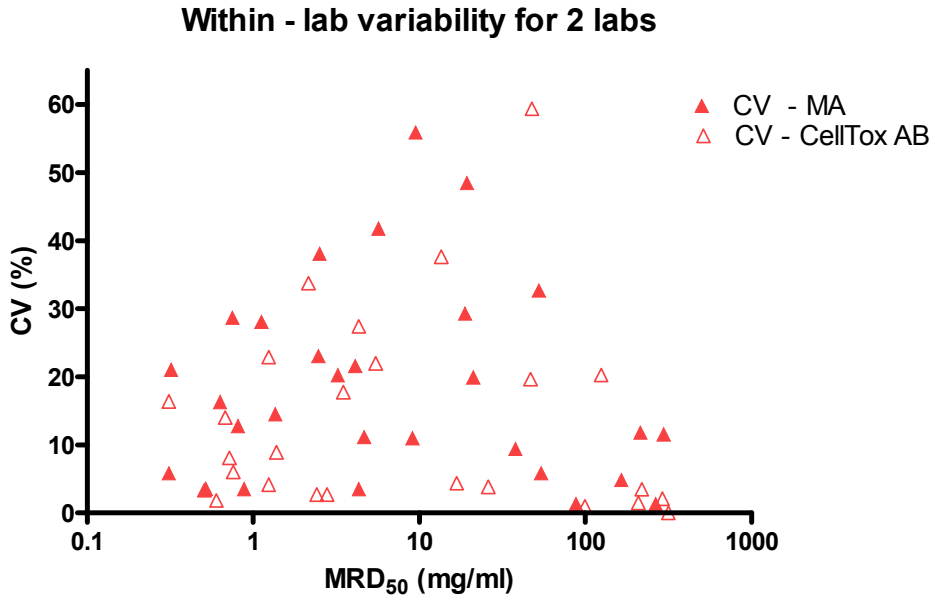

Figure 3.2.2.5.a **Graph of the variability of CM data for the two laboratories participating in the COLIPA study. The CV for each test material measurement is plotted against the mean MRD_{50} value for that material. Twenty-nine materials are plotted for Company # 4 and 26 materials for Company # 5.**

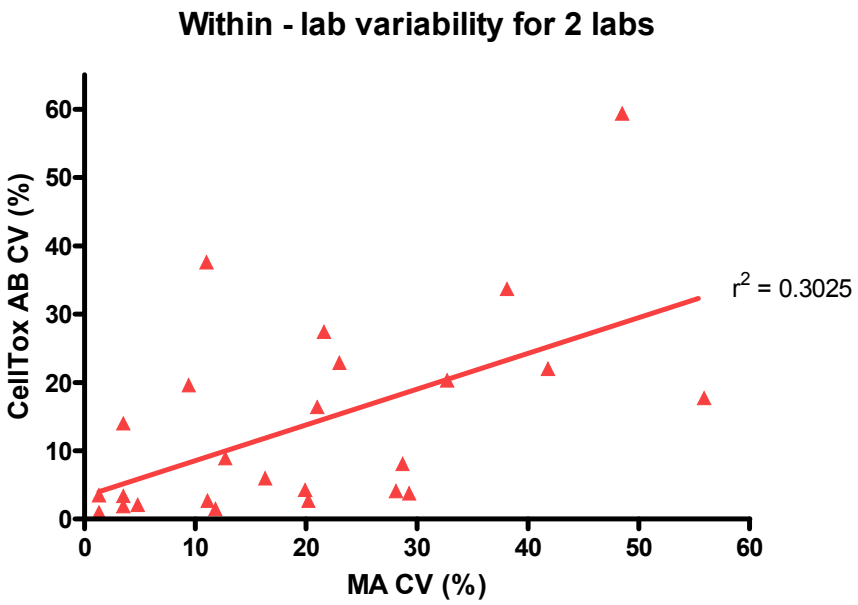

Figure 3.2.2.5.b Graph of the coefficient of variation of CM data for the two laboratories participating in the COLIPA study. The CV obtained for each test material at Company # 5 is plotted against the CV obtained for each test material at Company # 4

3.2.2.6 SM and CM comparison study

At the time that the new CM became available from Company # 6, the Company # 1 conducted a study to determine the relationship between data obtained with the CM (transwell) and an 810 second exposure with data obtained from the prior 500 second protocol using the SM (coverslip). Three replicates were conducted for each of eleven proprietary substances with both the CM and the SM. These unpublished data were obtained from the archives of the Company # 1 and are found in Tables 3.2.2.6.a and 3.2.2.6.b. The mean MRD_{50} presented in the tables below was based on the average of transformed individual trial data ($rMRD_{50} \rightarrow MRD_{50}$), not the transformed mean $rMRD_{50}$.

For the SM, the CV's for the 11 surfactant containing test materials ranged from 1.7% to 41.5% (mean CV 22.8%), and for the CM the CV's ranged from 3.7% to 47.7% (mean CV 21.8%). It appears that there are no major differences for the within-laboratory variability between the two instruments and protocols. The identity of the materials could not be determined.

Table 3.2.2.6.a SM data from a study to determine a conversion factor for changing SM (500 sec exposure) values to CM (810 sec exposure) values. N = 11 materials. Data from Annex F35.

Substance	MRD_{50} (mg/mL)	SD	CV (%)	Number of replicates
A	21.3	3.79	17.7	3
B	0.083	0.00	1.7	3
C	0.273	0.02	5.5	3
D	0.283	0.12	41.5	4
E	12.0	2.27	18.9	3
F	0.032	0.01	28.2	3
G	0.143	0.05	36.8	3
H	1.62	0.39	24.2	3
I	0.038	0.01	14.7	3
J	0.036	0.01	26.9	3
K	0.109	0.04	34.5	3
Mean			**22.8**	
Median			**24.2**	

Table 3.2.2.6.b CM data from a study to determine a conversion factor for changing SM (500 sec exposure) values to CM (810 sec exposure) values. N = 11 materials. Data from Annex F35.

Substance	MRD_{50} (mg/mL)	SD	CV (%)	Number of replicates
A	55.8	7.00	12.5	3
B	0.320	0.09	28.4	3
C	0.722	0.15	20.4	3
D	0.428	0.02	3.7	3
E	8.61	3.08	35.8	3
F	0.067	0.01	19.2	3
G	0.244	0.06	24.3	3
H	2.19	0.23	10.6	3
I	3.21	1.53	47.7	3
J	0.158	0.01	10.6	3
K	0.236	0.07	28.4	3
Mean			**21.8**	
Median			**20.4**	

3.2.2.7 Materials common to the EC/HO and the COLIPA study (Type II reproducibility)

Sixteen of the twenty materials which were common to both the EC/HO study and the COLIPA study provide the opportunity to investigate longer-term within lab reproducibility (or type II reproducibility) for the laboratory of Company # 4 Nearly two years elapsed between the initiation of the EC/HO study (April 1993) and the initiation of the COLIPA study (January 1995) at Company # 4 Although the materials had similar identities, it should be noted that they were purchased at different times (approximately 2 years apart) and could have had different purities or compositions (for the materials that are only broadly described mixtures of different sized polymers). Tables 3.2.2.7.a and 3.2.2.7.b show the results for 11 surfactant materials and 9 non-surfactants, respectively. Two materials were determined to be compatible with testing with the microphysiometer instrument in use for the EC/HO study but not for the COLIPA study, while one material was determined to be testable in the COLIPA study, but not the EC/HO study. One of the common materials was judged not testable for both studies. Thus there was fairly good correlation on the testing decisions for both studies even though the criteria for “testability” was more carefully defined in the COLIPA protocol than it was in the EC/HO protocol.

For the 9 surfactants and 7 non-surfactants that were actually tested in both studies, the CV’s ranged from 0.6% to 37.7% (mean CV 17.4%) for surfactants and 5.0% to 65.1% (mean CV 32.5%) for the non-surfactants. The reproducibility for the surfactants was very similar for the Type I and Type II situations (mean CV of 19.5% (Company # 4); COLIPA study] and 26.11% [SM31 (Company # 4); EC/HO study], respectively). The higher variability for the non-surfactant materials was primarily due to the variability of the three highly acidic or basic materials – 1% sodium hydroxide, 10% sodium hydroxide, and 30% trichloroacetic acid.

Tables 3.2.2.7.a & 3.2.2.7.b show MRD_{50} results for materials of similar identity. It appears from these data that long term within lab reproducibility is good for the surfactant materials; however, this conclusion is based on CVs determined from only two measurements per test material (mean COLIPA value vs. mean EC/HO value) and therefore should not be over-interpreted. The mean MRD_{50} presented in the tables below was based on the transformed mean log MRD_{50} given in the Bibra Preliminary Report on the EC/HO study, not the average of transformed individual trial data ($rMRD_{50} \rightarrow MRD_{50}$).

Table 3.2.2.7.a Surfactant materials - Comparison of the MRD_{50} values for testing conducted approximately 21 months apart. N = 11 surfactant materials. Data from Annex H7 & F13.

Substance	COLIPA Mean MRD_{50} (mg/mL) [CV%] MA	EC/HO Mean MRD_{50} (mg/mL) SM 31	Mean MRD_{50} (mg/mL)	SD	CV (%)
Surfactant Chemicals					
Benzalkonium chloride 1%	4.11	5.16	4.62	0.72	15.6
Benzalkonium chloride 10%	0.321	0.47	0.39	0.1	26.3
Benzalkonium chloride 5%	0.811	1.09	0.96	0.2	21.4
Cetylpyridinium bromide 10%	*	1.02	1.02	*	*
Cetylpyridinium bromide 6%	1.36	1.35	1.35	0.01	0.6
Polyethylene glycol 400	296.5	*	296.5	*	*
Sodium lauryl sulphate 15%	0.517	0.60	0.56	0.06	10.9
Sodium lauryl sulphate 3%	3.23	3.04	3.13	0.15	4.8
Triton X-100 10%	2.47	1.96	2.21	0.37	16.6
Triton X-100 5%	4.66	3.39	4.03	0.9	22.3
Tween 20	9.50	5.53	7.5	2.83	37.7
Mean – All Surfactant Chemicals					**17.4**
Median – All Surfactant Chemicals					**16.6**

* - Material determined to be unsuitable for testing

Table 3.2.2.7.b Non-surfactant materials - Comparison of the MRD_{50} values for testing conducted approximately 21 months apart. N = 9 non-surfactant chemicals. Data from Annex H7 & F13.

Substance	COLIPA Mean MRD_{50} (mg/mL) [CV%] MA	EC/HO Mean MRD_{50} (mg/mL) SM 31	Mean MRD_{50} (mg/mL)	SD	CV (%)
Non-Surfactant Chemicals					
Ethyl acetate	*	53.7	53.7	*	*
Glycerol	214.8	180.7	197.75	24.11	12.2
Imidazole	18.8	23.1	20.95	3.04	14.5
Isopropanol	52.6	91.2	71.9	27.29	38
Methyl ethyl ketone	54.2	50.5	52.35	2.62	5
n-Butyl acetate	*	*	*	*	*
Sodium hydroxide 1%	9.09	16.2	12.65	5.03	39.8
Sodium hydroxide 10%	4.33	1.60	2.97	1.93	65.1
Trichloroacetic acid 30%	1.12	2.47	1.8	0.95	53.2
Mean – All Non-Surfactant Chemicals					**32.5**
Median – All Non-Surfactant Chemicals					**38.0**

* - Material determined to be unsuitable for testing

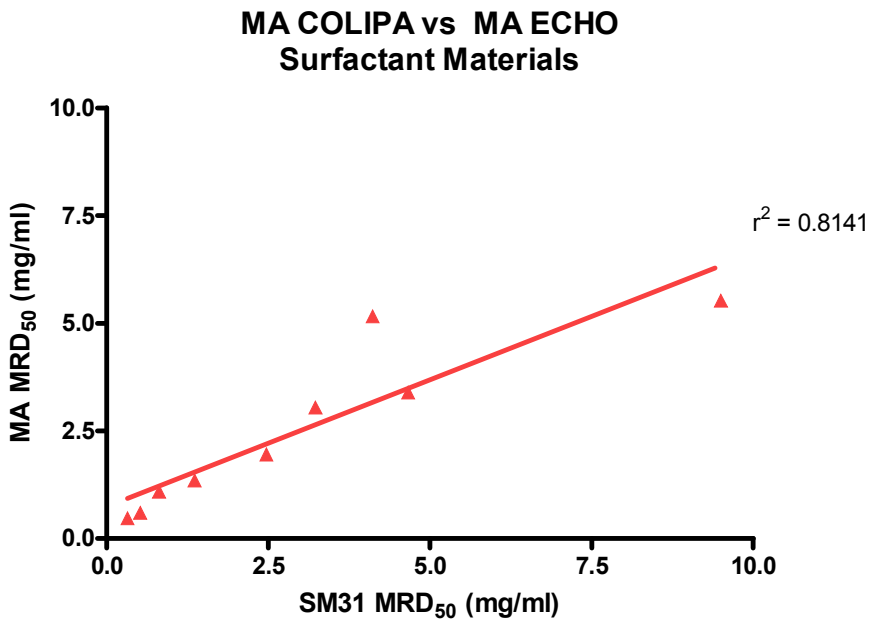

Figure 3.2.2.7.a Surfactant materials for comparison between inter-laboratory data of COLIPA and EC/HO studies.

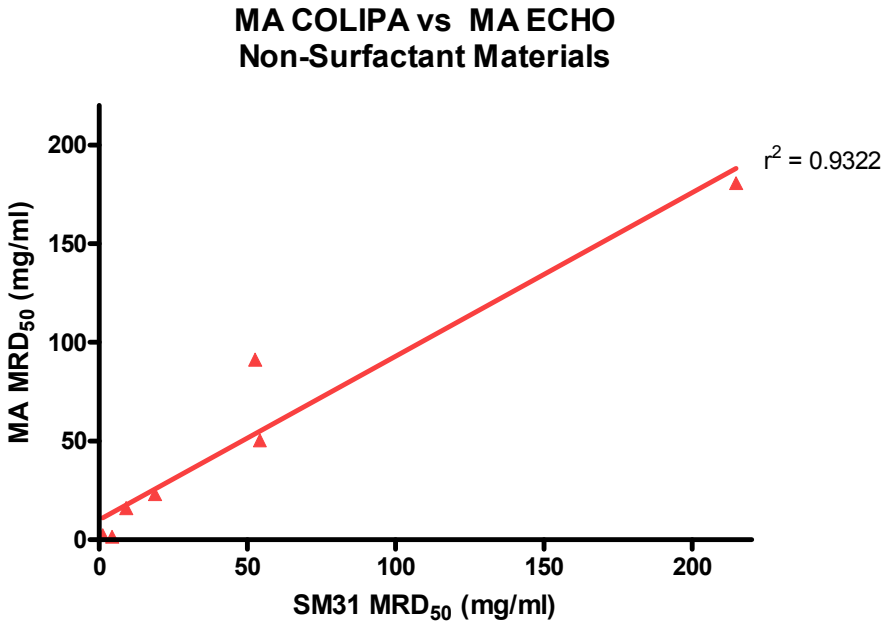

Figure 3.2.2.7.b Non-surfactant materials for comparison between inter-laboratory data of COLIPA and EC/HO studies.

3.2.2.8 Historic positive control data from Company # 3

The CM instrument was first used by the *in vitro* toxicology staff at Company # 4 in 1994. At that time the practice of maintaining a graphical record of the results of the positive control material – 10% SLS in sterile, deionized water – was begun (Figure 3.2.2.8). This practice has continued through the transfer of the instrument and staff to Company # 3 in 1997, and continues to this day. Table 3.2.2.8 presents a summary of the results for 629 assays conducted over a 12 plus year period as well as the results from the last 94 assays conducted over the last two years. That little change has occurred in the absolute MRD_{50} in the last 12 years can be inferred from the 12 year average of 0.0799 mg/mL versus the last two year's average of 0.0775 mg/mL. The average CV calculated over the last 12 years is 14.3%. Over the last approximately 2 years the average CV has increased to 18.9%.

Table 3.2.2.8 Positive Control Data of SLS completed at Company # 3. Data from Annex H17.

Substance	Dates	No. of Assays	Mean MRD_{50} (mg/mL)	SD	CV (%)
SLS	April, 14 1994 – June 30, 2006	629	0.0799	0.011	14.3
SLS	March 2, 2004 - June 30, 2006	94	0.0775	0.015	18.9

SLS MRD_{50} values are plotted on a control graph with upper and lower cut-off ranges graphed at two SD of all data (March 2004 – June 2006). Assays performed on days when the MRD_{50} fell outside of the two SD range (5 points on this graph) were repeated. Because on some days more than one SLS control was run, some points may overlap such that it may appear that fewer than 94 values are plotted.

It appears from these data that there is good long term with-in lab reproducibility for a single material.

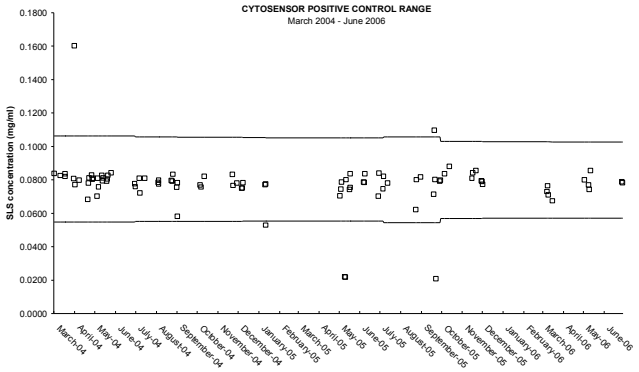

Figure 3.2.2.8 Graph of 10% SLS (positive control) MRD_{50} values obtained at Company # 3 over a 28-month period.

3.2.3 Additional information/evaluation – Operator Variability

Operator variability was assessed using the positive control data collected from technicians performing the Cytosensor Microphysiometer assay at Company # 3. The average MRD_{50} (mg/mL), standard deviation, and coefficient of variance is presented in Table 3.2.3 for each of the eight technicians. Raw data are in Annex H25. The mean MRD_{50} collected for the eight technicians was 0.0795 mg/mL, with a mean coefficient of variation of 12.6%. Although different batches of SLS were used between the operators, the CV remains low which indicates a low degree of operator variability for the Cytosensor Microphysiometer when using the same protocol.

Table 3.2.3 Operator variability assessed using the positive control data at Company # 3. N = 1 material. See Annex H25.

Operator Number	Number of Experiments	Average MRD_{50} (mg/mL)	Standard Deviation	CV (%)
1	49	0.0830	0.0050	6.03
2	110	0.0792	0.0082	10.31
3	32	0.0760	0.0047	6.14
4	76	0.0814	0.0228	28.06
5	44	0.0791	0.0073	9.18
6	80	0.0781	0.0102	13.07
7	18	0.0779	0.0166	21.25
8	7	0.0773	0.0055	7.15
Total for all operators	416	0.0795	0.0125	12.6

3.3 Additional studies where raw data are not available: attempt to combine the data using weigh-of-evidence approaches

There were no other studies which we could find which would allow any inferences about within laboratory reproducibility. The one other important SM study (Bagley, Bruner et al. 1992) described elsewhere in this BRD reports only average MRD_{50}'s and provides no information on either number of trials or the reproducibility of those trials.

3.3.1 Protocols and long term intralaboratory reproducibility

Sixteen of the twenty materials which were common to both the EC/HO study and the COLIPA study provide the opportunity to investigate longer-term within lab reproducibility (or type II reproducibility) for Company # 4 laboratory. Nearly two years elapsed between the initiation of the EC/HO study (April 1993) and the initiation of the COLIPA study (January 1995) at Company # 4 Thus there was fairly good correlation on the testing decisions for both studies even though the criteria for "testability" was more carefully defined in the COLIPA protocol than it was in the EC/HO protocol. The

correlation coefficient between surfactants tested in both the COLIPA and EC/HO studies was 0.8141. The correlation coefficient between non-surfactants tested in both the COLIPA and EC/HO studies was 0.9322.

There were no major discrepancies for the within-laboratory variability between the SM and CM and their respective protocols (Table 3.2.2.6.a and 3.2.2.6.b). For the SM, the CV's for the 11 test materials ranged from 1.7% to 41.5% (mean CV 22.8%), and for the CM the CV's ranged from 3.7% to 47.7% (mean CV 21.8%).

Company # 4 was included in three different studies, the CTFA, EC/HO, and COLIPA. Although all three studies had different protocols, the CV was approximately the same. Therefore, the difference in protocol did not have a significant effect on the MRD_{50} values (mean CV 27.5%).

3.3.2 Test materials

For the similar chemicals tested in the COLIPA and EC/HO study at Company # 4 21 months apart, the non-surfactant chemicals had a slightly higher CV (32.5%) than the surfactant and surfactant containing test materials (17.4%). However, during the COLIPA study, the surfactants had a slightly higher CV for both laboratories (19.5% and 26.11%) than the non-surfactants (17.5% and 28.9%). The study from Company # 2 tested only surfactants for a CV of 23.0%. The surfactants and surfactant containing materials had a consistent CV of approximately 20% between three different protocols (Company # 2, COLIPA, and EC/HO); however, the non-surfactants had a varying CV ranging from 65.1% to 5.0% between the two different protocols (COLIPA and EC/HO).

When looking at the MRD_{50} of the COLIPA study versus the MRD_{50} of the EC/HO study (Figure 3.2.2.7.a and b), the correlation coefficient for the surfactants was slightly lower at 0.8141 than for the non-surfactants 0.9322. However, the number of materials for each set was small (9 and 7 for the surfactants and non-surfactants, respectively).

3.3.3 Classifications

To determine if reproducibility was related to the degree of irritation (by MRD_{50} rank), we plotted the percent CV against the MRD_{50} for the COLIPA, EC/HO, CTFA, and Bagley studies in Figure 3.3.3.a. There appeared to be no strong correlation between MRD_{50} and CV, however there were a few more of the extremely high CV values (>70) for materials in the moderate to non-irritating (MRD_{50} = 1 – 300 mg/ml) range.

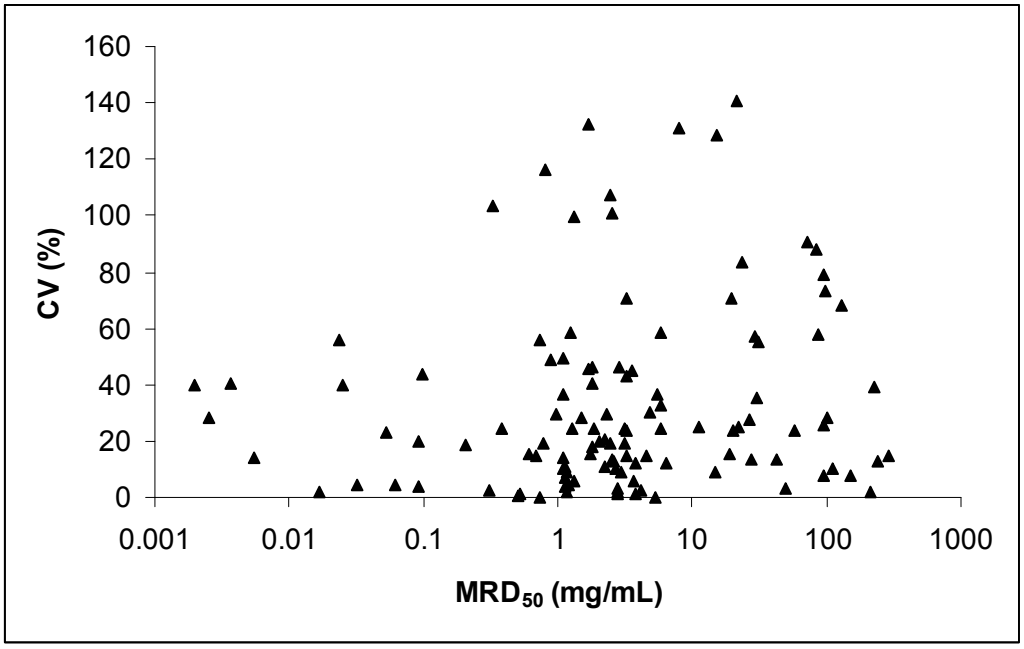

Figure 3.3.3.a Comparison of CV and MRD_{50} for the COLIPA, EC/HO, CTFA and the Bagley studies.

4. Transferability (Module 3)

4.1 Brief description of study results on transferability and availability of Standardised Operating Procedures (SOPs)

ECVAM has recently discussed transferability of tests in their manuscript describing a modular approach to validation (Hartung, Bremer et al. 2004). They state that transferability "should demonstrate that the test can be successfully repeated in a laboratory different from the one which has developed or which was involved in the optimization of the test." They further say that the description should provide an estimation of the amount of training that will be necessary to successfully transfer the test to a naïve laboratory as well as to identify possible sources of within-laboratory and between laboratory variability.

The most notable data assessing interlaboratory transferability of the CM protocol were generated by Company # 4 and Company # 5 during the COLIPA study (Brantom, Bruner et al. 1997). The data from this study indicate that a thorough and precise protocol can be transferred between laboratories with success. Before the study began, both laboratories were given a common protocol which addressed decision criteria for which materials could be tested, as well as detailed procedures for actual testing. The protocol was discussed in detail between the study directors for each laboratory, sources of potential variability were identified, and methods to approach potential problems were addressed. However, we were unable to find any written records of these discussions. No formal laboratory training sessions were conducted. The protocol for these studies can be found in Annex A.

During the conduct of the actual COLIPA study, communications between the two laboratories were kept to a minimum to ensure that the results of the study clearly reflected the reproducibility that could be expected from two completely independent laboratories conducting any set of studies. Test materials were coded in a double blind fashion, i.e. both laboratories had different code designations for identical test materials, by a third party so that results for a test material could not be easily be discussed between laboratories.

As can be seen in Table 4.1.a, between-laboratory reproducibility (the CV for the mean MRD_{50}s for each laboratory) ranged from 1.0% - 106.9% with a mean CV of 24.7%. It should be noted that the highest CV (106.9%) was found for a high concentration of base (10% NaOH) which might be expected in a dilution-based assay conducted in a medium with minimal buffering capacity. However, 1% NaOH and 30% trichloroacetic acid were much more reproducible. Section 3 (Within-laboratory reproducibility) of this BRD also addressed reproducibility difficulties for materials that were strongly acidic or basic.

The mean between-laboratory CV for these studies (24.7%; 21.4% if the results for 10% NaOH are removed) was somewhat higher than the individual within-laboratory CV (18.5% for MA, 13.1% for CT) for the chemicals and formulations that were tested by both

laboratories. These data would seem to indicate that a second laboratory (Company # 5) can successfully repeat data from a laboratory that was involved in the optimization of the test (Company # 4).

Table 4.1.a Results from Company # 4 and Company # 5 laboratories for the COLIPA study demonstrating the transferability of the Cytosensor Microphysiometer protocol. N = 10 surfactant chemicals, N = 7 surfactant formulations, N = 7 Non-surfactant chemicals, and N = 2 non-surfactant formulations. Data from Annex F13 & H29.

Chemical	Conc. tested	MRD_{50} Values (mg/mL)		Number of Replicates		Mean MRD_{50}	SD MRD_{50}	CV (%)
		Company # 4	Company # 5	Company # 4	Company # 5			
Surfactant Chemicals								
Benzalkonium chloride 1%	1%	4.11	4.33	3	3	4.22	0.16	3.7
Benzalkonium chloride 10%	10%	0.32	0.31	3	3	0.31	0.01	3.2
Benzalkonium chloride 5%	5%	0.81	1.38	3	3	1.1	0.4	36.7
Polyethylene glycol 400	100%	296.5	316.23	3	3	306.36	13.95	4.6
SLS 15%	15%	0.52	0.51	3	3	0.51	0.01	1
SLS 3%	3%	3.23	2.78	3	3	3	0.32	10.6
Triton X-100 1%	1%	21.17	16.79	3	3	18.98	3.1	16.3
Triton X-100 10%	10%	2.47	1.24	3	3	1.85	0.87	46.8
Triton X-100 5%	5%	4.66	2.42	3	3	3.54	1.58	44.7
Tween 20	100%	9.5	3.49	3	3	6.5	4.25	65.4
Mean – Surfactant Chemicals								**23.3**
Median – Surfactant Chemicals								**13.5**
Surfactant Formulations								
Eye make-up remover	100%	87.77	99.31	3	3	93.54	8.16	8.7
Gel cleaner	100%	5.68	5.47	3	3	5.58	0.15	2.6
Liquid soap #1	100%	0.88	0.68	3	3	0.78	0.14	18.5
Pump Deodorant	5%	19.35	47.74	3	3	33.54	20.08	59.9
Shampoo – baby	100%	2.51	2.15	3	3	2.33	0.25	10.8
Shampoo #1 normal	100%	0.75	0.72	3	3	0.74	0.02	2.2
Skin cleaner	100%	0.63	0.76	3	3	0.7	0.09	13
Mean – Surfactant Formulations								**16.5**
Median – Surfactant Formulations								**10.8**
Non-Surfactant Chemicals								
Glycerol	100%	214.83	208.7	3	2	211.77	4.34	2
Imidazole	100%	18.84	26.03	3	3	22.43	5.09	22.7
Isopropanol	100%	52.59	124.51	3	3	88.55	50.86	57.4
Propylene glycol	100%	265.07	218.86	3	3	241.97	32.67	13.5
Sodium hydroxide 1%	1%	9.09	13.59	3	3	11.34	3.19	28.1
Sodium hydroxide 10%	10%	4.33	0.6	3	3	2.47	2.64	106.9
Trichloroacetic acid 30%	30%	1.12	1.24	3	3	1.18	0.09	7.3
Mean – Non-Surfactant Chemicals								**34.0**
Median – Non-Surfactant Chemicals								**22.7**
Non-Surfactant Formulations								
Hair styling lotion	100%	164.82	292.01	3	3	228.41	89.94	39.4
Mouthwash	100%	37.84	46.85	3	3	42.35	6.37	15
Mean – Non-Surfactant Formulations								**27.2**
Median – Non-Surfactant Formulations								**27.2**
Mean – All Materials								**24.7**
Median – All Materials								**14.3**

Two Standard Operating Procedures (SOPs) are available for this method from ECVAM SIS, INVITOX #102 "The Silicon Microphysiometer Toxicity Test - Procter and Gamble" written by Dr. Rosemarie Osborne and INVITOX #97 "The Silicon Microphysiometer Toxicity Test - Microbiological Associates" written by Dr. John Harbell. There is also a protocol outline available for the Institute For In Vitro Sciences Cytosensor Microphysiometer Bioassay; however, this document does not describe the test method in as much detail as the SOPs. The INVITOX SOPs describe the method in detail, but were submitted in 1996 so a number of modifications are necessary to both documents. A list of the major revisions that we suggest are necessary to be made to the Company # 4 SOP are listed in the table below (Table 4.1.b). The same revisions are applicable to the Procter and Gamble SOP also.

Table 4.1.b Major revisions to the Company # 4 SOP

SOP Section	Modification
Test Status – first six sentences	"The Cytosensor Microphysiometer, manufactured by Company # 6, measures alterations in the acidification rate of cells. The microphysiometer consists of a variety of components which include (1) Cytosensor Microphysiometer Unit(s), (2) "Cytosoft" – the computer program which runs the microphysiometer and collects the data, (3) sensor chambers, and (4) a printer. Various adherent cell types can be grown on a capsule cup, which is a disposable cup with a polycarbonate membrane with a 3 µm pore size. A spacer (circular disk) is added to the capsule cup and this assembly (cell capsule) is added to the sensor chamber. The silicon chip within the sensor chamber is capable of detecting very small changes in pH."
Procedure Details – Starvation Medium	Change to Seeding medium which contains 1% Fetal Bovine Serum
Procedure Details – Route of Administration	The test article is exposed for 810 seconds, followed by an approximate 300 second rinse, and an approximate 20 second rate measure
Procedure Details – Growth of Cells	The L929 cells are seeded in capsule cups using the seeding medium the day before the dosing is to be initiated. The cell capsules and spacer are placed into the microphysiometer flow chambers and exposed to MDMEM at 37 $\pm$1°C prior to dosing.
Procedure Details – Dose selection – second and third paragraphs	The exposure time for each test material dilution is 810 seconds. The test article doses for the definitive assay are chosen so that at least seven treatments are available for the determination of the MRD_{50}.

4.2 Facilities and major fixed equipment needed

In order to perform the Cytosensor Microphysiometer assay, equipment for both cell culture (Table 4.2.a) and specialized assay equipment (Table 4.2.b) are necessary. Company # 6 no longer supports the Cytosensor (i.e. provides repair services and replacement parts) as of June 2007. Sale of replacement parts will continue after this time only until the supply of parts is exhausted. Company # 6 will not procure any more parts or disposables after this date. They have also informed us that the names of their current suppliers for replacement parts will not be revealed to any current user.

These changes will obviously have a severe impact on anyone who is just now anticipating taking up the assay. Company # 3 laboratory has sufficient spare parts and disposables to last for several years, but after this time it is likely that the machine cannot be supported.

Table 4.2.a Equipment for cell culture

Equipment	Use
Laminar flow hood	Cell culture manipulations
Incubator, 37°, 5% CO_2, 90% humidity	Cell culture incubation
Cell Counter or Hemacytometer	Performing cell counts
Inverted Microscope	Observing the confluence of cell cultures
Water bath	Warming cell culture materials
Aspirator	Removal of media during routine cell culture and passaging
Refrigerator	Storing chemicals and reagents
Freezer, liquid nitrogen container	Storage of medium components and cell banks
Pipettes	Adding test materials and cell culture media

Table 4.2.b Specialized assay equipment

Equipment	Use
Company # 6 Cytosensor Unit	Automated dosing, rinsing, and data collection for the assay
Sensor Chamber	The capsule kit is added to this part of the Cytosensor Unit which includes the pH sensor device that collects data for the assay
Cytosoft computer program (see the paragraph below for additional information)	Software used to run the Cytosensor Unit and collect and calculate data
Debubbler membranes	Warmed to 6°C above the temperature of the sensor chamber to allow dissolved gases to leave the fluid and escape through the membrane

Reference Electrode Maintain Kit (KCl)	Stabilizes pH readings from the sensor chamber
Spacers	Defines the height and diameter of the compartment in which the cells are confined
Sterilant Solution Kit	To sterilize the Cytosensor Unit after experimentation is completed
Capsule Cups, 12 mm, 3 µm pore size	Immobilize living cells in the sensor chamber while allowing fluid movement through the cell layer
Balance	Weighing test materials for accurate dosing solution preparation

The Cytosoft program runs on a Macintosh IIsi. At least 5 megabytes of RAM (a minimum of which 2.5 megabytes must be allocated for the Cytosoft program) are needed along with a floating-point processor and 32-bit QuickDraw System 7.0 or higher to run this program. Two and a half megabytes of RAM will allow Cytosoft to gather and save data for 13 hours. It is recommended that all other software be closed and that Cytosoft is the only program running on the computer while data are being collected. If other programs are running, there is a risk of causing printing errors and/or software crashes.

The newest version of the software is Cytosoft 2.03 which can run on Mac OS 9, but has not been tested with OS X (Personal communication; Company # 6). Company # 6 states that both the Mac G3 and G4 computers can be used to operate the CM.

4.3 Required level of training, expertise, and demonstrated proficiency needed

The following discussion is based primarily on the 12-plus years of experience that staff at Company # 3 have had with the machine, as well as past conversation with other users.

Basic training in sterile cell culture technique is essential for this assay. Individuals wishing to perform this assay must demonstrate proficiency in the propagation of the L929 cells used in this assay. These procedures include training on utilization of a hemacytometer or Coulter Counter for calculation of the proper seeding density. A possible source of within and between laboratory variability is controlling the cell seeding so that the protocol requirement of a 50 to 200 µvolt/sec acidification rate can be routinely achieved. Also, individuals wishing to perform the Cytosensor Microphysiometer assay must be familiar with the preparation of cell culture media and must follow the proper protocol or SOP and construct a batch record to record the components of and steps taken to manufacture this media.

In addition to this basic sterile technique and cell culture training, individuals should demonstrate proficiency in performing serial dilutions and must be trained in the use of the Cytosoft software. Special attention should be paid to proper training in the test article dilution techniques to ensure that accurate dilutions are prepared, as this is another area

of possible variation within and between laboratories. In the experience of Company # 3, the training is usually comprised of watching an experienced technician set up the assay, create dilutions of the test articles and controls, and run the proper protocol once or twice followed by hands-on performance of the techniques for 2 to 3 assays under the direction of an experienced technician. Where appropriate, additional training in the analysis of the data via an Excel spreadsheet to determine the MRD_{50} value is also necessary. A working knowledge of Good Laboratory practices (GLPs) may also be helpful. Proper training and competency in the performance of this assay can be demonstrated by successfully running a positive control (e.g. SLS) and obtaining results within a predetermined historical mean or within the margin of error for the data generated during the COLIPA study mentioned previously.

The table below (Table 4.3) represents data obtained during training for various technicians performing the Cytosensor Microphysiometer assay at Company # 3. The SOPs and protocol remained essentially the same over time and between technicians. This data demonstrates the ability to teach the method to naïve individuals and obtain similar results after successful completion of a monitored training period. The MRD_{50} value obtained during these technician training runs is presented along with the batch of SLS used and the date on which the training was performed.

Table 4.3 Example of the MRD_{50} values for SLS obtained during the training of technicians at Company # 3. Data from internal records at Company # 3.

Date	Technician	Batch #	MRD_{50} (mg/mL)
02/26/99	Technician 1		0.0754
02/26/99	Technician 1		0.0967
04/23/03	Technician 2	40	0.0807
04/23/03	Technician 2	43	0.0953
04/24/03	Technician 2	40	0.0811
04/24/03	Technician 2	40	0.0831
04/24/03	Technician 2	43	0.0789
04/24/03	Technician 2	43	0.0846
08/28/03	Technician 3	43	0.0806
08/29/03	Technician 3	43	0.0799
08/24/05	Technician 4	59	0.0775
08/24/05	Technician 4	59	0.0782
08/24/05	Technician 4	61	0.0915
08/24/05	Technician 4	61	0.0780
08/25/05	Technician 4	61	0.0778
08/25/05	Technician 4	61	0.0622
03/02/06	Technician 4	62	0.0955
03/02/06	Technician 4	62	0.0870
03/02/06	Technician 5	62	0.0879
03/16/06	Technician 4	63	0.0711
03/16/06	Technician 5	63	0.0724
		Average	**0.0817**
		SD	0.0086
		CV (%)	10.5

5. Between-laboratory reproducibility (Module 4)

The SM and CM assays have been used by more than one laboratory in two major international studies and one more restricted study (see Table 5.1) which allowed between-laboratory reproducibility to be assessed. The EC/HO study had four laboratories (Department of Chemistry, Stanford University [Palo Alto, CA, USA]; Company # 2; Company # 4 and Company # 1) using the CM with the 810 second transwell protocol. The COLIPA study had two laboratories (Company # 5 and Company # 4) using the CM which also had an 810 second protocol. The restricted study of Bagley, Bruner et al. 1992 would not normally technically qualify for analysis in this section since the machines used by the participating laboratories (Company # 4 and Company # 2) were different. One laboratory used the SM fitted with a transwell cell chamber, and one used the SM with the glass coverslips. However, the other important factors in the protocols, e.g treatment time, cell line, general operation, were virtually identical making it useful to evaluate the reproducibility of the general method within this BRD.

5.1 Relevant results and information for studies where raw data is available

Table 5.1 Table presenting the relevant results and information for studies where raw data are available

Studies	Number of chemicals tested	Number of products tested	Instrument	Coded ?	Results reported (Range)	Replicates		Data format	Chemical classes tested	Ranges of toxicity covered	Physico-chemical properties covered
						n. of labs	n. of reps				
Several companies[1] (Bagley, Bruner et al. 1992)	12 of 12 tested	20 of 20 tested	SM (2 Protocols)	Yes	MRD_{50} (mg/mL) 0.002-15.364	2	3	Summary; MAS and log MRD_{50} only	Cleaning product ingredients-almost all surfactants	MAS 0.3 – 43 for analysis; 0 – 44.7 total	See Annex B, C, & D
EC/HO[2] (Balls, Botham et al. 1995)	31 – 38 (depending on lab) of 60 were compatible	0	CM	Yes	MRD_{50} (mg/mL) 0.39-363.08	4	3	Individual log MRD_{50} data; raw animal data	Alcohols, surfactants, ketones, bases, acids, esters	MMAS 0 – 108 EU: R36, R41, NC GHS : 1, 2A, 2B, NC EPA: I-IV	See Annex B, C, & D
COLIPA[3] (Brantom, Bruner et al. 1997)	17-20 (depending on lab) of 23 were compatible	9 of 32 were compatible	CM	Yes	MRD_{50} (mg/mL) 0.31-306.36	2	3	Individual log MRD_{50} data; raw animal data	Alcohols, surfactants, ketones, bases, acids, esters	MAS 0 – 106 EU: R36, R41, NC GHS : 1, 2A, 2B, NC EPA: I-IV	See Annex B, C, & D

[1] Performed using the silicon microphysiometer coverslip protocol and silicon microphysiometer transwell protocol with a 500-second exposures and L929 cells.
[2] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[3] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.

5.2 Discussions, e.g., were sources of variability taken into consideration

There was not sufficient information available from the three studies to investigate such items as time frame of the assay, different media suppliers, and level of adherence to the protocol. However, for both the COLIPA and EC/HO studies a detailed protocol was available and all participating laboratories agreed to conduct the study according to this document. Variables which are known are discussed in the individual study sections below.

5.2.1 Bagley, Bruner et al. 1992 study

Although the main purpose of this study was to determine the utility of the SM assay to predict the ocular irritation potential of 32 coded test materials, a secondary goal was to compare data from the traditional SM device with a prototype machine similar to what would later be the CM. Most of the protocol parameters were conserved between the two laboratories, e.g. L929 cells and a 500 second exposure period were used in both. However, the cells were plated on glass cover slips for exposure in one laboratory while the cells were plated on the permeable membrane of a transwell in the other laboratory. Although the medium and serum were purchased from the same company – GIBCO – one batch was produced in the US and the other in France.

A correlation analysis of the results from both laboratories was conducted by the study authors, and a very good correlation (Pearson/Spearman correlation coefficients 0.93/0.90 and all points lying close to a line of unity) were reported. As expected from their close correlation with each other, both assays also gave similar relatively high correlation coefficients with the Draize scores of the 32 test materials (Pearson/Spearman of 0.82/0.75 for the SM and 0.86/0.81 for the CM).

Table 5.2.1.a and 5.2.1.b display the mean, SD, and CV for results from each report for each of the 32 test materials. Fifty-six percent of the interlaboratory CV's were <25%. The results of this study indicate that when testing chemical ingredients, household cleaning products, personal care products, and cosmetics, the protocol used was very reproducible even when two different machine configurations (SM with glass coverslips and SM with transwells) were used. This indicates that the two different machine configurations provide similar data.

A comparison of the reproducibility of surfactant ingredients, non-surfactant ingredients, and surfactant based formulations and mixtures was conducted (Table 5.2.1.a, 5.2.1.b, and 5.2.1.c), and it could be seen that variability for the non-surfactant appeared slightly higher (CV non-surfactant 52.8% [two wildly disparate values only; one of 103.5% and one of 2.5%], surfactant ingredients 29.8%, and surfactant based formulations and mixtures 36.7%).

Table 5.2.1.a Surfactant based formulations and mixtures - Interlaboratory reproducibility of SM results from Bagley, Bruner et al. 1992 study. The number of replicates for each lab is unknown. N = 21 surfactant formulations.

Chemical	Conc. tested	MRD_{50} values (mg/mL)		Mean MRD_{50} (mg/mL)	SD	CV (%)
		Lab 1	Lab 2			
Surfactant Formulations						
Baby shampoo	100%	0.129	0.068	0.098	0.043	43.8
Bar soap, 10%	10%	0.063	0.059	0.061	0.003	4.4
Bath foam	100%	1.905	3.767	2.836	1.316	46.4
Bath gel/bath foam	100%	3.451	4.121	3.786	0.473	12.5
Dishwashing liquid	100%	1.528	2.158	1.843	0.446	24.2
Dishwashing liquid	100%	7.228	4.519	5.873	1.916	32.6
Ethyl acetate, 10% in Tween 80	10%	0.018	0.032	0.025	0.01	39.9
Foaming bath, 10%	10%	0.096	0.09	0.093	0.004	4.1
Hair conditioner	100%	0.003	0.002	0.002	0.001	28.3
Hair gel	100%	0.014	0.032	0.023	0.013	56.2
Hair gel	100%	0.005	0.003	0.004	0.001	40.7
Laundry detergent liquid	100%	29.309	1.419	15.364	19.721	128.4
Liquid soap	100%	1.919	1.535	1.727	0.272	15.7
Shampoo	100%	3.532	3.819	3.676	0.203	5.5
Shampoo	100%	2.716	3.811	3.264	0.774	23.7
Shampoo	100%	2.286	0.394	1.34	1.338	99.9
Shampoo	100%	4.406	0.731	2.568	2.598	101.2
Shower gel	100%	3.864	3.793	3.828	0.05	1.3
Shower gel with baby oil, 10%	10%	0.062	0.044	0.053	0.012	23.1
Shower gel, 10%	10%	0.104	0.078	0.091	0.018	20.2
Skin cleanser	100%	0.232	0.178	0.205	0.038	18.6
Mean – All Surfactant Materials						**36.7**
Median – All Surfactant Materials						**24.2**

Table 5.2.1.b Surfactant ingredients - Interlaboratory reproducibility of SM results from Bagley, Bruner et al. 1992 study. The number of replicates for each lab is unknown. N = 9 surfactant chemicals.

Chemical	Conc. tested	MRD_{50} values (mg/mL)		Mean MRD_{50}	SD MRD_{50}	CV (%)
		Lab 1	Lab 2			
Surfactant Chemicals						
CTAB, 10%	10%	2.877	2.495	2.686	0.271	10.1
CTAC, 1%	1%	0.519	0.509	0.514	0.007	1.3
MTAB, 10%	10%	1.75	2.323	2.036	0.405	19.9
Sodium lauryl sulphate, 10%	10%	1.374	1.262	1.318	0.079	6
Sodium lauryl sulphate, 5%	5%	0.448	1.033	0.74	0.414	55.9
Triton X-100, 10%	10%	0.144	1.489	0.817	0.951	116.5
Tween 20	100%	0.031	0.033	0.032	0.001	4.2
Tween 20, 10%	10%	0.006	0.005	0.005	0.001	14.3
Tween 80, 10%	10%	0.001	0.003	0.002	0.001	39.9
Mean – All Surfactant Chemicals						**29.8**
Median – All Surfactant Chemicals						**14.3**

Table 5.2.1.c Non-surfactant ingredients – Interlaboratory reproducibility of SM results from Bagley, Bruner et al. 1992 study. The number of replicates for each lab is unknown. N = 2 non-surfactant chemicals.

Chemical	Conc. tested	MRD_{50} values (mg/mL)		Mean MRD_{50}	SD MRD_{50}	CV (%)
		Lab 1	Lab 2			
Non-Surfactant Chemicals						
Citric acid, 18%	100%	0.087	0.564	0.325	0.337	103.5
Triethanolamine	100%	0.017	0.017	0.017	0.00	2.1
Mean – All Non-Surfactant Chemicals						**52.8**
Median – All Non-Surfactant Chemicals						**52.8**

5.2.2 EC/HO study

The EC/HO study was conducted with the CM fitted with a transwell. L929 cells were used as the target with an exposure time of 810 seconds. Four laboratories participated using a single protocol, but not enough information is available to establish how closely the participants in each laboratory adhered to the protocol. It is known that no formal training sessions were held between the laboratories. Sixty coded test materials which ranged across a wide variety of chemicals were tested. Raw data can be found in Annex H7 & H13 (Lovell, BIBRA Project No. 1367/1, Vol.1).

5.2.2.1 Reproducibility of "testable" decision for EC/HO study

One measure of reproducibility is a comparison of the somewhat subjective decision process involved in determining whether a material actually meets the criteria for compatibility with the test system. This is relatively easy to investigate for the EC/HO study since each laboratory made its own decision whether or not the chemical was compatible with the test system, i.e. was aqueous soluble and relatively non-viscous. The table below shows the results of this criterion being applied independently by each laboratory.

Table 5.2.2.1 Decisions of the laboratories involved in the EC/HO study as to whether individual test materials were compatible with testing in the CM. Data from Annex H13.

Number of labs agreeing material was "testable"	Number of test materials (% of total)
4 of 4	28 (46.6%)
3 of 4	2 (3.3%)
2 of 4	4 (6.7%)
1 of 4	10 (16.7%)
0 of 4	16 (26.7%)

It can be seen that in most of the cases (73.3%) all four labs came to the same decision, i.e. either all of the labs decided the material was "testable" or all of the labs decided it was not "testable". However, in 26.7% of the cases (16 chemicals), there was disagreement; one or more of the labs made a different decision than the remaining labs. It appears that there is some subjectivity in the "testable" or "not testable" decision.

A second question that can be asked about the reproducibility of the "testable" decision is whether one laboratory was constantly an outlier in the testable decision. In this study for the ten cases where only one laboratory concluded that the material was testable Laboratories 30 and 32 were each responsible for four of the decisions, with Laboratories 31 and 33 responsible for one each of the remaining two cases. There were two cases where three of four laboratories decided to test and one laboratory did not, in one case the outlier was Laboratory 32 and in the other case Laboratory 33. Thus it appears that Laboratory 32 was responsible for more of the "outlier" decisions than any other laboratory.

A third question that can be asked is: for situations where one or more laboratories determined that a material was not testable, did the remaining laboratories that did test the material show good or poor reproducibility? For the two chemicals (thiourea and ammonium nitrate) for which data were reported by three of the four labs, the results were split with thiourea being reproducible and ammonium nitrate variable (Table 5.2.2.2.b). For the four chemicals (cyclohexanol, l-aspartic acid, sodium perborate, and methylcyanoacetate) which were tested in only two laboratories, the results showed great variability. Although the results for L-aspartic acid were very similar between the laboratories, the results for the other three chemicals were quite different between the laboratories. These results suggest that as the disagreement among the labs as to the

testability of the chemical increases, the results that they produce are more disparate. This is good evidence that when two or more labs agreed that the chemical was not compatible, that this was the correct decision since there did appear to be problems in the testing of those chemicals.

It should be noted that none of the surfactants were determined to be untestable by any of the four laboratories.

5.2.2.2 Reproducibility of MRD_{50} values for the EC/HO study

A comparison of the MRD_{50} values for each of the 44 EC/HO test materials (16 of the 60 total materials could not be tested in the CM) which one or more of the participating laboratories tested is given in Table 5.2.2.2.a (12 surfactants) and Table 5.2.2.2.b (32 non-surfactants). These data come from the Bibra report (Lovell, BIBRA Project No. 1367/1, Vol.1). The mean, SD and CV values are based on the MRD_{50} values expressed in mg/ml. SD's and CV's are listed for the 23 non-surfactant chemicals where at least two of the four laboratories provided data. The overall average CV was 46.2% (37.0% for surfactants; 50.6% for non-surfactants). The largest discrepancies were found for situations where only two laboratories tested the materials; in three of these cases the CV's were greater than 100%. If only the 28 chemicals (11 surfactants, 17 non-surfactants) for which all four laboratories provided data are considered, the average CV drops to 38.2%.

It is likely that the relatively high variability was due to the fact that a general training session for all the labs was not held before the study was conducted. It also appears that the data produced by laboratory 32 often appeared to be at variance with the other three laboratories. Although it is tempting to reevaluate the variability only using data from the other three laboratories, this may give an erroneous impression of the reproducibility of the assay in the absence of formal interlaboratory training sessions.

Table 5.2.2.2.a Surfactant Materials - Between-laboratories reproducibility of CM results from EC/HO study. The number of replicates for each lab is unknown. N = 12 surfactant chemicals. Data from Annex H13.

Chemical	Conc. tested	MRD_{50} Values (mg/mL)				Mean MRD_{50} (mg/mL)	SD	CV (%)
		CM 30	CM 31	CM 32	CM 33			
Surfactant Chemicals								
Benzalkonium chloride	5%	1.15	1.09	0.98	1.28	1.13	0.12	11.1
Benzalkonium chloride	10%	0.26	0.47	0.38	0.44	0.39	0.09	24.2
Benzalkonium chloride [1]/[2]	1%	4.71	5.16	4.65	3.58	4.53	0.67	14.8
Cetylpyridinium bromide	10%	0.78	1.02	2.34	0.89	1.26	0.73	58.2
Cetylpyridinium bromide	6%	0.6	1.35	0.44	1.11	0.87	0.43	48.8
Cetylpyridinium bromide	0.1%	48.19	102.33	7.76	180.3	84.65	74.62	88.1
Polyethylene glycol 400	100%	*	*	*	363.92	*	*	*
Sodium lauryl sulfate	15%	0.62	0.6	0.51	0.74	0.62	0.1	15.5
Sodium lauryl sulfate	3%	2.71	3.04	3.74	3.64	3.28	0.49	15
Triton X-100	10%	1.61	1.96	1.5	2.22	1.82	0.33	18
Triton X-100 [1]/[2]	5%	1.9	3.39	5.09	2.53	3.23	1.39	43
Tween 20	100%	1.52	5.53	4.98	1.06	3.27	2.31	70.5
Mean – Surfactant Chemicals								**37.0**
Median – Surfactant Chemicals								**24.2**

* Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

Table 5.2.2.2.b Non-surfactant materials - Between-laboratories reproducibility of CM results from EC/HO study. The number of replicates for each lab is unknown. N = 48 non-surfactant chemicals. Data from Annex H13.

Chemical	Conc. tested	MRD_{50} Values (mg/mL)				Mean MRD_{50} (mg/mL)	SD	CV (%)
		CM 30	CM 31	CM 32	CM 33			
Non-Surfactant Chemicals								
1-Naphthalene acetic acid	100%	12.11	*	*	*	*	*	*
1-Naphthalene acetic acid	100%	*	*	*	*	*	*	*
2,2-Dimethylbutanoic acid	100%	*	*	*	*	*	*	*
2,5-Dimethylohexanediol	100%	76.21	155.96	6.21	156.31	98.67	72.25	73.2
2,6-Dichlorobenzoyl chloride	100%	*	*	*	*	*	*	*
2-Ethyl-1-hexanol	100%	*	*	*	*	*	*	*
4-Carboxybenzaldehyde	100%	*	*	*	*	*	*	*
Acetone	100%	153.82	140.28	139	162.18	148.82	11.15	7.5
Ammonium nitrate	100%	40.27	145.55	27.99	*	71.27	64.62	90.7
Benzoyl-L-tartaric acid	100%	0.81	*	*	*	*	*	*
Captan 90 concentrate	100%	*	*	*	*	*	*	*
Chlorhexidine	100%	*	*	*	*	*	*	*
Cyclohexanol	100%	15.49	*	0.58	*	8.03	10.5	131.3
Dibenzyl phosphate	100%	0.75	*	*	*	*	*	*
Ethanol	100%	97.05	117.49	123.03	110.41	111.99	11.22	10

Ethyl acetate	100%	*	53.7	*	*	*	*	*
Ethyl trimethyl acetate	100%	*	*	*	*	*	*	*
Ethyl-2-methylacetoacetate	100%	*	*	0.4	*	*	*	*
Fomesafen	100%	*	*	*	*	*	*	*
Gammabutyrolactone	100%	79.98	114.82	0.91	179.47	93.79	74.39	79.3
Glycerol	100%	121.62	180.72	8.26	208.93	129.88	88.87	68.4
Imidazole	100%	22.75	23.07	0.18	48.75	23.69	19.85	83.8
Isobutanol	100%	28.84	28.64	22.54	31.62	27.91	3.83	13.7
Isopropanol	100%	83.18	91.2	87.1	143.55	101.26	28.39	28
L-Aspartic acid	100%	1.11	1.17	*	*	1.14	0.04	3.6
Maneb	100%	*	*	*	*	*	*	*
Methyl acetate	100%	61.09	91.83	116.14	109.65	94.68	24.64	26
Methyl cyanoacetate	100%	42.95	*	0.13	*	21.54	30.28	140.5
Methyl ethyl ketone	100%	55.72	50.47	78.16	47.97	58.08	13.77	23.7
Methyl isobutyl ketone	100%	*	*	0.81	*	*	*	*
Methylcyclopentane	100%	*	*	*	*	*	*	*
n-Butyl acetate	100%	*	*	*	*	*	*	*
n-Hexanol	100%	*	*	*	*	*	*	*
n-Octanol	100%	*	*	*	*	*	*	*
Parafluoraniline	100%	*	*	3.48	*	*	*	*
Potassium cyanate	100%	28.18	36.06	9.4	50.82	31.11	17.25	55.4
Promethazine HCl	100%	1.35	1.48	0.81	1.45	1.27	0.31	24.4
Pyridine	100%	1.54	29.99	15.92	31.48	19.73	14.01	71
Quniacrine	100%	*	*	1.08	*	1.08	*	*
Sodium hydroxide	10%	2.28	1.6	2.67	2.49	2.26	0.47	20.8
Sodium hydroxide	1%	28.18	16.22	32.36	31.41	27.04	7.48	27.5
Sodium oxalate	100%	*	*	*	*	*	*	*
Sodium perborate, 4H20	100%	0.11	*	*	3.27	1.69	2.24	132.6
Tetraaminopyrimidine sulfate	100%	1.05	*	*	*	*	*	*
Thiourea	100%	50.12	50.93	*	47.97	49.68	1.53	3.1
Toluene	100%	*	*	*	*	*	*	*
Trichloroacetic acid	30%	1.69	2.47	0.81	2.2	1.79	0.73	40.7
Trichloroacetic acid	3%	13.9	13.8	16.29	16.11	15.03	1.36	9
Mean – Non-Surfactant Chemicals								**50.6**
Median – Non-Surfactant Chemicals								**28.0**
Mean when all four labs tested material – All Materials								**38.2**

* Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

In addition to the above table, an extensive analysis of the reproducibility and predictivity of the assays in the EC/HO study, not presented in the Balls, Botham et al. 1995 manuscript on the EC/HO study, was prepared by BIBRA International as BIBRA

Project No. 1367. This analysis consisted of graphical representations of the log transformed scores for each lab for each test material (Figure 5.2.2.2.a), a plot of residuals for each lab for each chemical (Figure 5.2.2.2.b), and a correlation matrix for all four laboratories with each other and with Draize MMAS scores (Table 5.2.2.2.c).

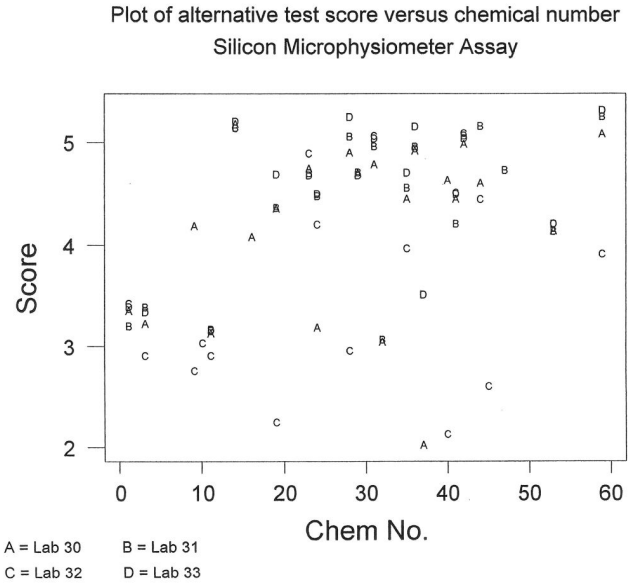

Figure 5.2.2.2.a Plot of CM test scores (log MRD_{50}) versus individual chemicals for the EC/HO study. From BIBRA project report 1367.

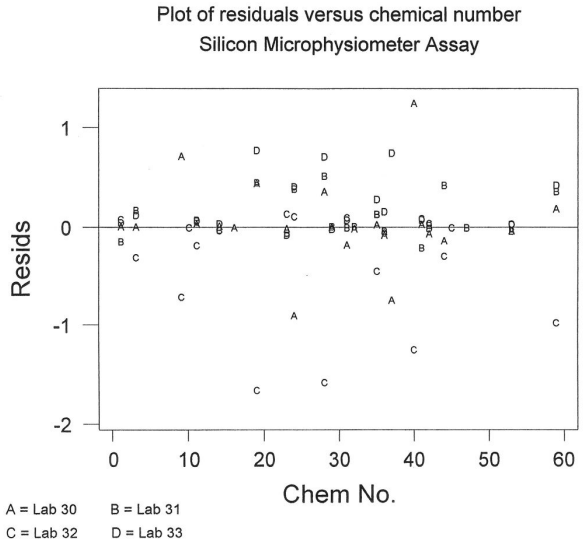

Figure 5.2.2.2.b Plot of residuals of CM test scores versus individual chemicals for the EC/HO study. From BIBRA project report 1367.

Table 5.2.2.2.c Correlation matrix of alternative CM test scores from individual laboratories for the EC/HO study. From BIBRA project report 1367.

Correlation Matrix of alternative test scores and MMAS
Silicon Microphysiometer Assay

	MMAS	Lab 30	Lab 31	Lab 32	Lab 33
MMAS	1.000				
Lab 30	-0.357	1.000			
Lab 31	-0.591	0.901	1.000		
Lab 32	-0.328	0.427	0.569	1.000	
Lab 33	-0.577	0.875	0.986	0.507	1.000

The Figures 5.2.2.2.a and 5.2.2.2.b, and Table 5.2.2.2.c indicate that the results from three of the labs (30, 31, and 33) correlate quite well, while the results from lab 32 often appear as outliers. This is similar to the conclusions from the "testability" decisions where lab 32 appeared to be responsible for more of the outlier decisions than any other laboratory. There is not enough information available at this time to determine what the reasons for variability of the MRD_{50} results are. Table 5.2.2.2.c should be viewed with the understanding that even though the results between two laboratories have a high linear correlation, the actual values obtained can be quite different since one laboratory may be consistently higher or lower than the other laboratory.

5.2.2.3 Reproducibility of predicted hazard classifications for the EC/HO study

A comparison of the between laboratories reproducibility of the prediction of hazard classifications is given in this section. Since none of the formal studies of the CM reported on in this BRD had predetermined prediction models for hazard classifications (although several did for Draize scores), the following analyses are based on prediction models derived during the construction of this BRD and presented in Chapter 6 – Predictive Capacity. Specifically these analyses of the EC/HO study are based on the prediction models proposed in Section 6.1.3.1.

Tables 5.2.2.3.a and 5.2.2.3.b. present the predicted EU, GHS and EPA classifications predicted for the surfactant and non-surfactant materials, respectively from the MRD_{50} values produced by each of the four participating laboratories. These

predictions were then consolidated into summary tables which are Tables 5.2.2.3.c and d for the surfactants and non-surfactant materials, respectively.

Table 5.2.2.3.c shows that for the surfactant materials where all four laboratories tested the materials (all but one of the cases) that 6 of the 11 materials were predicted to be the same classification, 3 of the 11 materials were predicted identically by 3 of the 4 labs, and 2 of the materials were had similar predictions for between less than three of the labs.

Table 5.2.2.3.d shows that for the non-surfactant materials where all four laboratories tested the materials that 9 of 17 materials were predicted the same by all four labs. Five materials had agreement between only 3 of the 4 labs and 3 of the 17 materials had agreement between less than 3 of the labs.

For the two non-surfactant materials where only three of the labs tested the materials three labs agreed on one and only two labs agreed on the other. If only two labs tested the materials, then both agreed for one material and both disagreed for the remaining three materials.

It appears from the above data that as fewer labs decided that a material was not testable under the constraints of the protocol, the reproducibility of the hazard predictions became worse.

Table 5.2.2.3.a Surfactant Materials - EU, GHS, and EPA classifications based on Cytosensor MRD_{50} values from EC/HO study. Cut-off values from Figures 6.1.3.1.a, 6.1.3.1.b, and 6.1.3.1.c were used. The number of replicates for each lab is unknown. N = 12 surfactant materials.

Chemical	Conc. tested	EU				GHS				EPA			
		CM 30	CM 31	CM 32	CM 33	CM 30	CM 31	CM 32	CM 33	CM 30	CM 31	CM 32	CM 33
Benzalkonium chloride	5%	R41	R41	R41	R41	1	1	1	1	I	I	I	I
Benzalkonium chloride	10%	R41	R41	R41	R41	1	1	1	1	I	I	I	I
Benzalkonium chloride [1]/[2]	1%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III
Cetylpyridinium bromide	10%	R41	R41	R36	R41	1	1	2A or 2B	1	I	I	II or III	I
Cetylpyridinium bromide	6%	R41	R41	R41	R41	1	1	1	1	I	I	I	I
Cetylpyridinium bromide	0%	R36	NL	R36	NL	2A or 2B	NL	2A or 2B	NL	II or III	IV	II or III	IV
Polyethylene glycol 400	100%	*	*	*	NL	*	*	*	NL	*	*	*	IV
Sodium lauryl sulfate	15%	R41	R41	R41	R41	1	1	1	1	I	I	I	I
Sodium lauryl sulfate	3%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III
Triton X-100	10%	R41	R41	R41	R36	1	1	1	2A or 2B	I	I	I	II or III
Triton X-100 [1]/[2]	5%	R41	R36	R36	R36	1	2A or 2B	2A or 2B	2A or 2B	I	II or III	II or III	II or III
Tween 20	100%	R41	R36	R36	R41	1	2A or 2B	2A or 2B	1	I	II or III	II or III	I

* = not tested

Table 5.2.2.3.b Non-surfactant Materials – EU, GHS, and EPA classifications based on Cytosensor MRD_{50} values from EC/HO study. Cut-off values from Figures 6.1.3.1.a, 6.1.3.1.b, and 6.1.3.1.c were used. The number of replicates for each lab is unknown. N = 48 non-surfactant materials.

Chemical	Conc. tested	EU				GHS				EPA			
		CM 30	CM 31	CM 32	CM 33	CM 30	CM 31	CM 32	CM 33	CM 30	CM 31	CM 32	CM 33
1-Naphthalene acetic acid	100%	R36	*	*	*	2A or 2B	*	*	*	II or III	*	*	*
1-Naphthalene acetic acid	100%	*	*	*	*	*	*	*	*	*	*	*	*
2,2-Dimethylbutanoic acid	100%	*	*	*	*	*	*	*	*	*	*	*	*
2,5-Dimethylohexanediol	100%	R36	NL	R36	NL	2A or 2B	NL	2A or 2B	NL	II or III	IV	II or III	IV
2,6-Dichlorobenzoyl chloride	100%	*	*	*	*	*	*	*	*	*	*	*	*
2-Ethyl-1-hexanol	100%	*	*	*	*	*	*	*	*	*	*	*	*
4-Carboxybenzaldehyde	100%	*	*	*	*	*	*	*	*	*	*	*	*
Acetone	100%	NL	NL	NL	NL	NL	NL	NL	NL	IV	IV	IV	IV
Ammonium nitrate	100%	R36	NL	R36	*	2A or 2B	NL	2A or 2B	*	II or III	IV	II or III	*
Benzoyl-L-tartaric acid	100%	R41	*	*	*	1	*	*	*	I	*	*	*
Captan 90 concentrate	100%	*	*	*	*	*	*	*	*	*	*	*	*
Chlorhexidine	100%	*	*	*	*	*	*	*	*	*	*	*	*
Cyclohexanol	100%	R36	*	R41	*	2A or 2B	*	1	*	II or III	*	I	*
Dibenzyl phosphate	100%	R41	*	*	*	1	*	*	*	I	*	*	*
Ethanol	100%	NL	NL	NL	NL	NL	NL	NL	NL	IV	IV	IV	IV
Ethyl acetate	100%	*	R36	*	*	*	2A or 2B	*	*	*	II or III	*	*
Ethyl trimethyl acetate	100%	*	*	*	*	*	*	*	*	*	*	*	*
Ethyl-2-methylacetoacetate	100%	*	*	R41	*	*	*	1	*	*	*	I	*
Fomesafen	100%	*	*	*	*	*	*	*	*	*	*	*	*
Gammabutyrolactone	100%	R36	NL	R41	NL	2A or 2B	NL	1	NL	II or III	IV	I	IV
Glycerol	100%	NL	NL	R36	NL	NL	NL	2A or 2B	NL	IV	IV	II or III	IV
Imidazole	100%	R36	R36	R41	R36	2A or 2B	2A or 2B	1	2A or 2B	II or III	II or III	I	II or III
Isobutanol	100%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III
Isopropanol	100%	NL	NL	NL	NL	NL	NL	NL	NL	IV	IV	IV	IV
L-Aspartic acid	100%	R41	R41	*	*	1	1	*	*	I	I	*	*
Maneb	100%	*	*	*	*	*	*	*	*	*	*	*	*
Methyl acetate	100%	R36	NL	NL	NL	2A or 2B	NL	NL	NL	II or III	IV	IV	IV
Methyl cyanoacetate	100%	R36	*	R41	*	2A or 2B	*	1	*	II or III	*	I	*
Methyl ethyl ketone	100%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III
Methyl isobutyl ketone	100%	*	*	R41	*	*	*	1	*	*	*	I	*
Methylcyclopentane	100%	*	*	*	*	*	*	*	*	*	*	*	*

n-Butyl acetate	100%	*	*	*	*	*	*	*	*	*	*	*	*
n-Hexanol	100%	*	*	*	*	*	*	*	*	*	*	*	*
n-Octanol	100%	*	*	*	*	*	*	*	*	*	*	*	*
Parafluoraniline	100%	*	*	R36	*	*	*	2A or 2B	*	*	*	II or III	*
Potassium cyanate	100%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III
Promethazine HCl	100%	R41	R41	R41	R41	1	1	1	1	I	I	I	I
Pyridine	100%	R41	R36	R36	R36	1	2A or 2B	2A or 2B	2A or 2B	I	II or III	II or III	II or III
Quniacrine	100%	*	*	R41	*	*	*	1	*	*	*	I	*
Sodium hydroxide	10%	R36	R41	R36	R36	2A or 2B	1	2A or 2B	2A or 2B	II or III	I	II or III	II or III
Sodium hydroxide	1%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III
Sodium oxalate	100%	*	*	*	*	*	*	*	*	*	*	*	*
Sodium perborate, 4H20	100%	R41	*	*	R36	1	*	*	2A or 2B	I	*	*	II or III
Tetraaminopyrimidine sulfate	100%	R41	*	*	*	1	*	*	*	I	*	*	*
Thiourea	100%	R36	R36	*	R36	2A or 2B	2A or 2B	*	2A or 2B	II or III	II or III	*	II or III
Toluene	100%	*	*	*	*	*	*	*	*	*	*	*	*
Trichloroacetic acid	30%	R41	R36	R41	R36	1	2A or 2B	1	2A or 2B	I	II or III	I	II or III
Trichloroacetic acid	3%	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III

* = not tested

Table 5.2.2.3.c Surfactant Materials – Agreement table for EU, GHS, and EPA classifications based on Cytosensor MRD_{50} values for the EC/HO study.

Where 4 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
4 labs	6	6	6
3 labs	3	3	3
<3 labs	2	2	2

Table 5.2.2.3.d Non-Surfactant Materials – Agreement table for EU, GHS, and EPA classifications based on Cytosensor MRD_{50} values for the EC/HO study.

Where 4 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
4 labs	9	9	9
3 labs	5	5	5
<3 labs	3	3	3
Where 3 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
3 labs	1	1	1
2 labs	1	1	1
<2 labs	0	0	0
Where 2 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
Both agree	1	1	1
Both disagree	3	3	3

5.2.3 COLIPA study

The COLIPA study (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999) was conducted with the CM which utilizes a transwell. L929 cells were used as the target with an exposure time of 810 seconds. The two instruments used in the study were essentially identical; differing only in that one instrument had two 4-channel modules and the other one 4-channel module. Both laboratories participated using a single protocol (Annex A), but not enough information is available to establish how closely the participants in each laboratory adhered to the protocol. Each laboratory was asked to conduct the study according to the "spirit of GLP". The results from each laboratory were subjected to a quality audit by the QAU from BIBRA, but the individual laboratories did not document the extent of QA that was applied within their own laboratories. It is known that the protocol and potential problem areas were discussed between the study directors before the start of the testing phase. Fifty-five coded test materials (23 chemicals, 32 products) which were representative of substances commonly used in the cosmetics industry were tested.

5.2.3.1 Reproducibility of “testable” decision for COLIPA study

Each laboratory made its own decision whether or not the chemical was compatible with the test system, i.e. formed a true solution in the culture medium. It was reported (Harbell, Osborne et al. 1999) that six of the 55 samples formed microsuspensions which were not acceptable for testing under the strict requirements of the protocol for this study; however, the authors stated that microsuspensions were normally tested under current practice. The table below shows the results of this criterion being applied independently by each laboratory.

Table 5.2.3.1 Decisions of the laboratories involved in the COLIPA study as to whether individual test materials were compatible with testing in the CM. Data from Annexes F13 & H29.

Number of labs agreeing material was “testable”	Number of test materials (% of total)
2 of 2	26 (47.2%)
1 of 2	3 (5.5%)
0 of 2	26 (47.2%)

Similar to the results of the EC/HO study, in most of the cases (94.4%) both labs came to the same decision, i.e. either both of the labs decided the material was “testable” or both of the labs decided it was not “testable”. In only 5.5% of the cases (3 chemicals out of 55) was there disagreement. This study showed less subjectivity in the “testable” or ”not testable” decision then the EC/HO study, perhaps because of the clear protocol prohibition against testing any material which was not in a single phase at the highest tested concentration. The three discrepant chemicals were: 1) cetylpyridinium bromide 6% which was tested by Company # 4 but not by Company # 5 (cetylpyridinum bromide 10% was considered unsuitable for testing by both labs), 2) sodium lauryl sulphate 30% which was tested by Company # 4 but not by Company # 5 (sodium lauryl sulfate 10% was tested by both labs), and 3) methyl ethyl ketone which again was tested by Company # 4 and not by Company # 5.

5.2.3.2 Reproducibility of MRD_{50} values for the COLIPA study

A comparison of the MRD_{50} values for each of the 26 COLIPA test materials which one or more of the participating laboratories tested is given in Tables 5.2.3.2.a, 5.2.3.2.b, and 5.2.3.2.c. The mean, SD and CV values are based on the MRD_{50} values expressed in mg/ml. SD’s and CV’s are only listed for the 26 chemicals where both participating laboratories provided data.

The mean CV for the surfactant materials was 23.3%, for the surfactant based formulations and mixtures was 16.5%, and for the non-surfactant materials was 32.5%. The mean CV for the non-surfactants was highly influenced by the 106.9% CV of 10% NaOH, as might be expected from the dilution protocol which is used to test the materials. The overall average CV was 24.7%.

Table 5.2.3.2.a Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer results from COLIPA study. N = 13 surfactant chemicals. Data from Annexes F13 & H29.

Chemical	Conc. tested	MRD_{50} Values (mg/mL)		Number of Replicates		Mean MRD_{50} (mg/mL)	SD	CV (%)
		Company # 4	Company # 5	Company # 4	Company # 5			
Surfactant Chemicals								
Benzalkonium chloride 1%	1%	4.11	4.33	3	3	4.22	0.16	3.7
Benzalkonium chloride 10%	10%	0.32	0.31	3	3	0.31	0.01	3.2
Benzalkonium chloride 5%	5%	0.81	1.38	3	3	1.1	0.4	36.7
Cetylpyridinium bromide 10%	10%	*	*	-	-			
Cetylpyridinium bromide 6%	6%	1.36	*	3	-	1.36		
Polyethylene glycol 400	100%	296.5	316.23	3	2	306.36	13.95	4.6
SLS 15%	15%	0.52	0.51	3	3	0.51	0.01	1
SLS 3%	3%	3.23	2.78	3	3	3	0.32	10.6
SLS 30%	30%	0.31	*	3	-	0.31		
Triton X-100 1%	1%	21.17	16.79	3	3	18.98	3.1	16.3
Triton X-100 10%	10%	2.47	1.24	3	3	1.85	0.87	46.8
Triton X-100 5%	5%	4.66	2.42	3	3	3.54	1.58	44.7
Tween 20	100%	9.5	3.49	3	3	6.5	4.25	65.4
Mean – Surfactant Chemicals								**23.3**
Median – Surfactant Chemicals								**13.5**

* - Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

Table 5.2.3.2.b Surfactant based formulations and mixtures - Between-laboratories reproducibility of Cytosensor Microphysiometer results from COLIPA study. N = 22 surfactant formulations. Data from Annexes F13 & H29.

Chemical	Conc. tested	MRD_{50} Values (mg/mL)		Number of Replicates		Mean MRD_{50} (mg/mL)	SD	CV (%)
		Company # 4	Company # 5	Company # 4	Company # 5			
Surfactant Formulations								
Cleansing foam III	100%	*	*	-	-	-	-	-
Emulsion antiperspirant	100%	*	*	-	-	-	-	-
Eye make-up remover	100%	87.77	99.31	3	3	93.54	8.16	8.7
Gel cleaner	100%	5.68	5.47	3	3	5.58	0.15	2.6
Hair conditioner	100%	*	*	-	-	-	-	-
Hair dye base F#1	100%	*	*	-	-	-	-	-
Hair dye base form #3	100%	*	*	-	-	-	-	-
Hand cleaner	100%	*	*	-	-	-	-	-
Hand soap	100%	*	*	-	-	-	-	-
Hydrophilic ointment	100%	*	*	-	-	-	-	-
Liquid soap #1	100%	0.88	0.68	3	3	0.78	0.14	18.5
Moisturiser with sunscreen	100%	*	*	-	-	-	-	-
Perfumed skin lotion	100%	*	*	-	-	-	-	-
Polishing scrub	100%	*	*	-	-	-	-	-
Pump Deodorant	5%	19.35	47.74	3	3	33.54	20.08	59.9
Shampoo – baby	100%	2.51	2.15	3	3	2.33	0.25	10.8
Shampoo #1 normal	100%	0.75	0.72	3	3	0.74	0.02	2.2
Shampoo 2-in-1	100%	*	*	-	-	-	-	-
Shampoo antidandruff	100%	*	*	-	-	-	-	-
Shower gel	100%	*	*	-	-	-	-	-
Skin cleaner	100%	0.63	0.76	3	3	0.7	0.09	13
Sunscreen SPF 15	100%	*	*	-	-	-	-	-
Mean – Surfactant Formulations							**16.5**	
Median – Surfactant Formulations							**10.8**	

* - Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

Table 5.2.3.2.c Non-Surfactants, ingredients, and mixtures - Between-laboratories reproducibility of Cytosensor Microphysiometer results from COLIPA study. N = 10 non-surfactant chemicals and N = 10 non-surfactant formulations. Data from Annexes F13 & H29.

Chemical	Conc. tested	MRD_{50} Values (mg/mL)		Number of Replicates		Mean MRD_{50} (mg/mL)	SD	CV (%)
		MA	CT AB	MA	CT AB			
Non-Surfactant Chemicals								
Ethyl acetate	100%	*	*	-	-			
Glycerol	100%	214.83	208.7	3	2	211.77	4.34	2
Imidazole	100%	18.84	26.03	3	3	22.43	5.09	22.7
Isopropanol	100%	52.59	124.51	3	3	88.55	50.86	57.4
Methyl ethyl ketone	1%	54.18	*	3	-	54.18		
n-Butyl acetate	100%	*	*	-	-			
Propylene glycol	100%	265.07	218.86	3	3	241.97	32.67	13.5
Sodium hydroxide 1%	1%	9.09	13.59	3	3	11.34	3.19	28.1
Sodium hydroxide 10%	10%	4.33	0.6	3	3	2.47	2.64	106.9
Trichloroacetic acid 30%	30%	1.12	1.24	3	3	1.18	0.09	7.3
Mean – Non-Surfactant Chemicals								**34.0**
Median – Non-Surfactant Chemicals								**22.7**
Non-Surfactant Formulations								
Blush	100%	*	*	-	-			
Cologne	100%	*	*	-	-			
Eye liner	100%	*	*	-	-			
Eye shadow	100%	*	*	-	-			
Hair dye base form #2	100%	*	*	-	-			
Hair styling lotion	100%	164.82	292.01	3	3	228.41	89.94	39.4
Mascara	100%	*	*	-	-			
Mouthwash	100%	37.84	46.85	3	3	42.35	6.37	15
Sunscreen lotion	10%	*	*	-	-			
Toothpaste	100%	*	*	-	-			
Mean – Non-Surfactant Formulations								**27.2**
Median – Non-Surfactant Formulations								**27.2**
Mean – All Non-Surfactant Materials								**32.5**
Median – All Non-Surfactant Materials								**22.7**

* - Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

5.2.3.3 Reproducibility of predicted hazard classifications for the COLIPA study

A comparison of the between laboratories reproducibility of the prediction of hazard classifications is given in this section. Since none of the formal studies of the CM reported on in this BRD had predetermined prediction models for hazard classifications (several did for Draize scores), the following analyses are based on prediction models derived during the construction of this BRD and presented in Chapter 6 – Predictive Capacity. Specifically these analyses of the COLIPA study are based on the prediction models proposed in Section 6.1.3.3.

Tables 5.2.3.3.a, b and c. present the predicted EU, GHS and EPA classifications predicted for the surfactant materials, surfactant-based formulations and non-surfactant materials, respectively from the MRD_{50} values produced by each of the two participating laboratories. These predictions were then consolidated into summary tables which are Tables 5.2.3.3.d, 5.2.3.3.e, and 5.2.3.3.f for the surfactant materials, surfactant-based formulations and non-surfactant materials, respectively.

Table 5.2.3.3.d shows that for the surfactant materials where both laboratories tested the materials that 90% (9 of the 10) materials were predicted to be the same classification for all three classification systems.

Table 5.2.3.3.e shows that for the surfactant-based formulations where both laboratories tested the materials that 100% (7 of 7) materials were predicted the same both labs.

For the 9 non-surfactant materials where both of the labs tested the materials (Table 5.2.3.3.f), 76.7% (7of 9 materials) were predicted to be the same hazard classification.

Table 5.2.3.3.a Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications from COLIPA study. Cut-off values are based on Figures 6.1.3.3.d, 6.1.3.3.e, and 6.1.3.3.f. N = 14 surfactant materials.

Chemical	Conc. tested	EU		GHS		EPA	
		Company # 4	Company # 5	Company # 4	Company # 5	Company # 4	Company # 5
Benzalkonium chloride 1%	1%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Benzalkonium chloride 10%	10%	R41	R41	1	1	I	I
Benzalkonium chloride 5%	5%	R41	R41	1	1	I	I
Cetylpyridinium bromide 10%	10%	*	*	*	*	*	*
Cetylpyridinium bromide 6%	6%	R41	*	1	*	I	*
Polyethylene glycol 400	100%	NL	NL	NL	NL	IV	IV
SLS 15%	15%	R41	R41	1	1	I	I
SLS 3%	3%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
SLS 30%	30%	R41	*	1	*	I	*
Triton X-100 1%	1%	NL	NL	NL	NL	II or III	II or III
Triton X-100 10%	10%	R36	R41	2A or 2B	1	II or III	I
Triton X-100 5%	5%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Tween 20	100%	R36	R36	2A or 2B	2A or 2B	II or III	II or III

* - Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

Table 5.2.3.3.b Surfactant based formulations and mixtures - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications from COLIPA study. Cut-off values are based on Figures 6.1.3.3.d, 6.1.3.3.e, and 6.1.3.3.f. N = 22 surfactant based formulations.

Chemical	Conc. tested	EU		GHS		EPA	
		MA	CT AB	MA	CT AB	MA	CT AB
Cleansing foam III	100%	*	*	*	*	*	*
Emulsion antiperspirant	100%	*	*	*	*	*	*
Eye make-up remover	100%	NL	NL	NL	NL	IV	IV
Gel cleaner	100%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Hair conditioner	100%	*	*	*	*	*	*
Hair dye base F#1	100%	*	*	*	*	*	*
Hair dye base form #3	100%	*	*	*	*	*	*
Hand cleaner	100%	*	*	*	*	*	*
Hand soap	100%	*	*	*	*	*	*
Hydrophilic ointment	100%	*	*	*	*	*	*
Liquid soap #1	100%	R41	R41	1	1	I	I
Moisturiser with sunscreen	100%	*	*	*	*	*	*
Perfumed skin lotion	100%	*	*	*	*	*	*
Polishing scrub	100%	*	*				
Pump Deodorant	5%	NL	NL	NL	NL	II or III	II or III
Shampoo – baby	100%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Shampoo #1 normal	100%	R41	R41	1	1	I	I
Shampoo 2-in-1	100%	*	*	*	*	*	*
Shampoo antidandruff	100%	*	*	*	*	*	*
Shower gel	100%	*	*	*	*	*	*
Skin cleaner	100%	R41	R41	1	1	I	I
Sunscreen SPF 15	100%	*	*	*	*	*	*

* - Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

Table 5.2.3.3.c Non-Surfactants, ingredients, and mixtures - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications from COLIPA study. Cut-off values are based on Figures 6.1.3.3.a, 6.1.3.3.b, and 6.1.3.3.c. N = 10 surfactant chemicals and 10 surfactant materials.

Chemical	Conc. tested	EU		GHS		EPA	
		Company # 4	Company # 5	Company # 4	Company # 5	Company # 4	Company # 5
Surfactant Chemicals							
Ethyl acetate	100%	*	*	*	*	*	*
Glycerol	100%	NL	NL	NL	NL	IV	IV
Imidazole	100%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Isopropanol	100%	R36	NL	2A or 2B	NL	II or III	IV
Methyl ethyl ketone	1%	R36	*	2A or 2B	*	II or III	*
n-Butyl acetate	100%	*	*	*	*	*	*
Propylene glycol	100%	NL	NL	NL	NL	IV	IV
Sodium hydroxide 1%	1%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Sodium hydroxide 10%	10%	R36	R41	2A or 2B	1	II or III	I
Trichloroacetic acid 30%	30%	R41	R41	1	1	I	I
Surfactant Materials							
Blush	100%	*	*	*	*	*	*
Cologne	100%	*	*	*	*	*	*
Eye liner	100%	*	*	*	*	*	*
Eye shadow	100%	*	*	*	*	*	*
Hair dye base form #2	100%	*	*	*	*	*	*
Hair styling lotion	100%	NL	NL	NL	NL	IV	IV
Mascara	100%	*	*	*	*	*	*
Mouthwash	100%	R36	R36	2A or 2B	2A or 2B	II or III	II or III
Sunscreen lotion	10%	*	*	*	*	*	*
Toothpaste	100%	*	*	*	*	*	*

Table 5.2.3.3.d Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications agreement from COLIPA study.

Where 2 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
Both agree	9	9	9
Both disagree	1	1	1

Table 5.2.3.3.e Surfactant based formulations and mixtures - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications agreement from COLIPA study.

Where 2 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
Both agree	7	7	7
Both disagree	0	0	0

Table 5.2.3.3.f Non-Surfactants, ingredients, and mixtures - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications agreement from COLIPA study.

Where 2 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
Both agree	7	7	7
Both disagree	2	2	2

5.2.4 Analysis of materials common to both the EC/HO and COLIPA studies

The EC/HO and the COLIPA studies shared 20 test materials for which data were produced by all six (5 unique) laboratories (4 EC/HO laboratories and 2 COLIPA laboratories with one laboratory which participated in both studies). This allows between-laboratory reproducibility over a longer time period (~21 months). The general reproducibility of these studies was presented by Harbell, Osborne et al. 1999 in a tabular format shown here as Table 5.2.4.a. The correlation coefficients presented only represent results derived from fourteen materials where data were obtained from all 6 laboratories (see Tables 5.2.4.b and 5.2.4.c).

Table 5.2.4.a Between laboratories reproducibility correlation coefficients for 14 common test materials tested in the EC/HO (30-33) and COLIPA (27-28) studies. Lab 27 and 31 are the same laboratory (Harbell, Osborne et al. 1999).

Lab	27	28	30	31	32	33	
27		0.96	0.93	0.97	0.68	0.93	
28	0.94		0.96	0.99	0.59	0.95	
30	0.88	0.86		0.97	0.68	0.99	Pearson correlation
31	0.95	0.96	0.91		0.57	0.96	
32	0.71	0.62	0.71	0.69		0.61	
33	0.88	0.85	0.97	0.89	0.60		
Spearman correlation							

It can be seen that in general there is a very good correlation (around 0.9 or greater) between at least four of the laboratories. Only one lab (32) appears to be a slight outlier with correlation coefficients with the other labs that range between 0.57 and 0.71.

Tables 5.2.4.b and 5.2.4.c show a comparison of results between the six laboratories (five unique) for the 20 materials common to the COLIPA and EC/HO studies. It can be seen that there were only 14 materials where data were obtained from all six laboratories. For the surfactants (Table 5.2.4.b), one difference between the studies was that neither lab in the COLIPA study determined that 10% cetylpyridinium bromide met the criteria for testing, whereas all four labs did in the EC/HO study. This is probably due to the fact that the protocol for the COLIPA study was more descriptive about the qualifying criteria used. With a slightly more dilute mixture of cetylpyridinium bromide (6%), all but

one laboratory determined that it could be tested. The CV's for between-laboratory reproducibility ranged from 1.2% to 46.6%.

For the non-surfactant materials (Table 5.2.4.c) only seven of the nine materials had the same determination for testability by all labs in each study. Six of the materials were testable by all six labs and one material was determined to be not testable by all 6 laboratories. The CV's for between-laboratory reproducibility ranged from 3.9% to 57.8%.

The between laboratory reproducibility of the CM for materials common to the EC/HO study and the COLIPA study were also analyzed with respect to the predicted hazard classifications. These prediction models utilized to make these predictions are discussed in sections 5.2.2.3 and 5.2.3.3 for the EC/HO study and the COLIPA study, respectively.

Table 5.2.4.d illustrates the reproducibility for the surfactant materials, and Table 5.2.4.e illustrates the reproducibility of the non-surfactant materials. These results are summarized in Table 5.2.4.f for the surfactant materials and in Table 5.2.4.g for the non-surfactant materials.

The surfactant summary (Table 5.2.4.d) shows that when all 6 labs tested the material, the labs were in complete agreement for 5 out of 7 (71.4%) materials. For one material only 5 labs agreed and for 2 materials only 4 labs agreed. There was only one case where 5 labs tested the material, and in this case all 5 labs agreed.

For the non-surfactant materials agreement was not as good. Table 5.2.4.g shows that when all 6 labs tested the material, there was complete agreement for only 1 of 5 (20%) of the materials. Five labs agreed on three materials and four labs agreed on 2 materials. In the one case where only 5 labs tested the material, all five labs agreed on the hazard prediction.

Table 5.2.4.b Surfactant materials - Between laboratories reproducibility for 11 test materials in common for the EC/HO (Labs 30-33) and COLIPA (Company # 4 and Company # 5) studies. Company # 4 (Lab 27) and SM31 are the same laboratory. N = 11 surfactant materials. Annexes H13, F13 & H29.

Substance	COLIPA MRD_{50} (mg/mL)			EC/HO MRD_{50} (mg/mL)					Between Study Mean	Between Study
	Company # 4 (27)	Company # 5 (28)	Mean	SM 30	SM 31	SM 32	SM 33	Mean	MRD_{50} (mg/mL)	CV (%)
Benzalkonium chloride 1%	4.11	4.33	4.22	4.71	5.16	4.65	3.58	4.53	4.37	4.9
Benzalkonium chloride 10%	0.32	0.31	0.32	0.26	0.47	0.38	0.44	0.39	0.35	14.3
Benzalkonium chloride 5%	0.81	1.38	1.1	1.15	1.09	0.98	1.28	1.13	1.11	1.9
Cetylpyridinium bromide 10%	*	*	*	0.78	1.02	2.34	0.89	1.26	1.26	*
Cetylpyridinium bromide 6%	1.36	*	1.36	0.6	1.35	0.44	1.11	0.87	1.12	30.7
Polythylene glycol 400	296.50	316.23	306.37				363.92	363.92	335.14	12.1
Sodium lauryl sulphate 15%	0.52	0.51	0.51	0.62	0.6	0.51	0.74	0.62	0.57	12.9
Sodium lauryl sulphate 3%	3.23	2.78	3.01	2.71	3.04	3.74	3.64	3.28	3.14	6.2
Triton X-100 10%	2.47	1.24	1.86	1.61	1.96	1.5	2.22	1.82	1.84	1.2
Triton X-100 5%	4.66	2.42	3.54	1.9	3.39	5.09	2.53	3.23	3.38	6.5
Tween 20	9.50	3.49	6.5	1.52	5.53	4.98	1.06	3.27	4.88	46.6
Mean										**13.8**
Median										**9.3**

Table 5.2.4.c Non-surfactant materials - Between laboratories reproducibility for 9 test materials in common for the EC/HO (Labs 30-33) and COLIPA (Company # 4 and Company # 5) studies. Company # 4 (Lab 27) and SM31 are the same laboratory. N = 9 non-surfactant materials. Annexes H13, F13 & H29.

Substance	COLIPA MRD_{50} (mg/mL)			EC/HO MRD_{50} (mg/mL)					Between Study Mean	Between Study
	Company # 4 (27)	Company # 5 (28)	Mean	SM 30	SM 31	SM 32	SM 33	Mean	MRD_{50} (mg/mL)	CV (%)
Ethyl acetate	*	*	*	*	53.7	*	*	53.7	53.7	*
Glycerol	214.80	208.7	211.75	121.62	180.72	8.26	208.93	129.88	170.82	33.9
Imidazole	18.84	26.03	22.4	22.75	23.07	0.18	48.75	23.69	23.04	3.9
Isopropanol	52.60	124.5	88.55	83.18	91.2	87.1	143.55	101.26	94.9	9.5
Methyl ethyl ketone	54.20	*	54.2	55.72	50.47	78.16	47.97	58.08	56.14	4.9
n-Butyl acetate	*	*	*	*	*	*	*	*	*	*
Sodium hydroxide 1%	9.09	13.6	11.35	28.18	16.22	32.36	31.62	27.1	19.19	57.8
Sodium hydroxide 10%	4.33	0.6	2.47	2.28	1.6	2.67	2.49	2.26	2.36	6.2

Trichloroacetic acid 30%	1.12	1.24	1.18	1.69	2.47	0.81	2.2	1.79	1.49	29.2
Mean										**20.8**
Median										**9.5**

* - Participating laboratory did not test the chemical because it determined that chemical was not compatible with the test system.

Table 5.2.4.d Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications for test materials in common for the EC/HO (Labs 30-33) and COLIPA (Company # 4) studies. Company # 4 (Lab 27) and SM31 are the same laboratory. Cut-off values from Figures 6.1.3.4.1.b, 6.1.3.4.1.c, 6.1.3.4.2.b, 6.1.3.4.2.c, 6.1.3.4.3.b, and 6.1.3.4.3.c were applied. N = 11 surfactant materials. Annexes H13, F13 & H29.

Substance	EU						GHS						EPA					
	COLIPA		EC/HO				COLIPA		EC/HO				COLIPA		EC/HO			
	Company # 4 (27)	Company # 5 (28)	SM 30	SM 31	SM 32	SM 33	Company # 4 (27)	Company # 5 (28)	SM 30	SM 31	SM 32	SM 33	Company # 4 (27)	Company # 5 (28)	SM 30	SM 31	SM 32	SM 33
Benzalkonium chloride 1%	R36	R36	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III	II or III	II or III
Benzalkonium chloride 10%	R41	R41	R41	R41	R41	R41	1	1	1	1	1	1	I	I	I	I	I	I
Benzalkonium chloride 5%	R41	R41	R41	R41	R41	R41	1	1	1	1	1	1	I	I	I	I	I	I
Cetylpyridinium bromide 10%	*	*	R41	R41	R36	R41	*	*	1	1	2A or 2B	1	*	*	I	I	II or III	I
Cetylpyridinium bromide 6%	R41	*	R41	R41	R41	R41	1	*	1	1	1	1	I	*	I	I	I	I
Polyethylene glycol 400	NL	NL	*	*	*	NL	NL	NL	*	*	*	NL	IV	IV	*	*	*	IV
Sodium lauryl sulphate 15%	R41	R41	R41	R41	R41	R41	1	1	1	1	1	1	I	I	I	I	I	I
Sodium lauryl sulphate 3%	R36	R36	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III	II or III	II or III
Triton X-100 10%	R36	R41	R41	R41	R41	R36	2A or 2B	1	1	1	1	2A or 2B	II or III	I	I	I	I	II or III

Triton X-100 5%	R36	R36	R41	R36	R36	R36	2A or 2B	2A or 2B	1	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	I	II or III	II or III	II or III
Tween 20	R36	R36	R41	R36	R36	R41	2A or 2B	2A or 2B	1	2A or 2B	2A or 2B	1	II or III	II or III	I	II or III	II or III	I

Table 5.2.4.e Non-Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications for test materials in common for the EC/HO (Labs 30-33) and COLIPA (Company # 4 and Company # 5) studies. Company # 4 (Lab 27) and SM31 are the same laboratory. Cut-off values from Figures 6.1.3.4.a, 6.1.3.4.b, and 6.1.3.4.c were applied. N = 9 non-surfactant materials. Annexes H13, F13 & H29.

Substance	EU						GHS						EPA					
	COLIPA		EC/HO				COLIPA		EC/HO				COLIPA		EC/HO			
	Company # 4 (27)	Company # 5 (28)	SM 30	SM 31	SM 32	SM 33	Company # 4 (27)	Company # 5 (28)	SM 30	SM 31	SM 32	SM 33	Company # 4 (27)	Company # 5 (28)	SM 30	SM 31	SM 32	SM 33
Ethyl acetate	*	*	*	R36	*	*	*	*	*	2A or 2B	*	*	*	*	*	III	*	*
Glycerol	NL	NL	NL	NL	R36	NL	NL	NL	NL	NL	2A or 2B	NL	IV	IV	IV	IV	II or III	IV
Imidazole	R36	R36	R36	R36	R41	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	1	2A or 2B	II or III	II or III	II or III	II or III	I	II or III
Isopropanol	R36	NL	NL	NL	NL	NL	2A or 2B	NL	NL	NL	NL	NL	II or III	IV	IV	IV	IV	IV
Methyl ethyl ketone	R36	*	R36	R36	R36	R36	2A or 2B	*	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	*	II or III	II or III	II or III	II or III
n-Butyl acetate	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Sodium hydroxide 1%	R36	R36	R36	R36	R36	R36	2A or 2B	2A or 2B	2A or 2B	2A or 2B	2A or 2B	2A or 2B	II or III	II or III	II or III	II or III	II or III	II or III

Sodium hydroxide 10%	R36	R41	R41	R41	R41	R41	2A or 2B	1	1	1	1	1	II or III	I	I	I	I	I
Trichloroacetic acid 30%	R41	R41	R41	R41	R41	R41	1	1	1	1	1	1	I	I	I	I	I	I

Table 5.2.4.f Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications agreement for test materials in common for the EC/HO (Labs 30-33) and COLIPA (Company # 4 and Company # 5) studies. Company # 4 (Lab 27) and SM31 are the same laboratory.

Where all 6 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
6 labs	5	5	5
5 labs	1	1	1
4 labs	2	2	2
3 labs	0	0	0
<3 labs	0	0	0
Where 5 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
5 labs	1	1	1
4 labs	0	0	0
3 labs	0	0	0
<3 labs	0	0	0
Where 4 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
4 labs	0	0	0
3 labs	1	1	1
<3 labs	0	0	0

Table 5.2.4.g Non-Surfactant Materials - Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications agreement for test materials in common for the EC/HO (Labs 30-33) and COLIPA (Company # 4 and Company # 5) studies. Company # 4 (Lab 27) and SM31 are the same laboratory.

Where all 6 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
6 labs	2	2	2
5 labs	4	4	4
4 labs	0	0	0
3 labs	0	0	0
<3 labs	0	0	0
Where 5 labs tested the material			
Agreement	**EU**	**GHS**	**EPA**
5 labs	1	1	1
4 labs	0	0	0
3 labs	0	0	0
<3 labs	0	0	0

5.2.5 Analysis of materials common to both the CTFA and COLIPA studies

Since there were also surfactant-based formulations which were common to both the CTFA and COLIPA studies it seemed reasonable to use these materials to help assess the between laboratory reproducibility of the CM assay. This seemed as if it would be especially useful since one lab (Company # 4) participated in both studies. However, it is known that there were several differences in the studies, one of which is that the common formulations had to be remade for the COLIPA study using the same formulas as the CTFA study. This is problematic because it is possible, if not likely, that due to the several years difference between the studies that the composition and/or purity of many of the ingredients would be different.

To determine if this was the case we first looked at the rabbit scores for the 6 materials that were common to the two studies. Table 5.2.5 shows that while three of the materials (Baby shampoo 2, Eye Makeup Remover, and Skin Cleanser) had identical hazard classifications between the two studies, the other three materials (Liquid Soap 1, Shampoo 1, and Gel Cleanser) had widely disparate scores. Considering just the EU classifications, Liquid Soap 1 and Shampoo 1 both went from a CTFA study Not Classified to a COLIPA study R41. The Gel Cleanser hazard categories changed in the opposite direction – from an R41 in the CTFA study to a Not Classified in the COLIPA study.

Thus we concluded that it was likely that the composition of some of the materials differed between the studies (although the Draize test also differed since local anesthesia was used in the CTFA study but not in the COLIPA study), and therefore it would not make sense to attempt to estimate reproducibility of the *in vitro* test using data of such uncertainty.

Table 5.2.5 Between-laboratories reproducibility of Cytosensor Microphysiometer hazard classifications for test materials in common for the CTFA and COLIPA studies. Hazard classifications are based on *in vivo* Draize data.

CTFA							COLIPA					
Substance Designation	Conc. Tested	n. of animals	EU	GHS	EPA	CM converted value MRD_{50} (mg/mL)	Conc. Tested	n. of animals	EU	GHS	EPA	Average MRD_{50} (mg/mL) (avg MA & CT)
Liquid Soap 1	25%	6	Not classified	No category	Category III	2.80	100%	3	R41	Category 1	Category I	0.78
Gel Cleanser*	100%	6	R41	No category	Category I	3.19	100%	3	Not classified	No Category	Category III	5.58
Baby Shampoo 2	100%	6	R41	Category 1	Category I	1.50	100%	3	R41	Category 1	Category I	2.33
Shampoo 1*	25%	6	Not classified	No category	Category III	1.72	100%	3	R41	Category 1	Category I	0.735
Eye Makeup Remover	100%	6	Not classified	No category	Category IV	20.0	100%	3	Not classified	No Category	Category IV	93.5
Skin Cleaner*	100%	6	R41	Category 1	Category I	1.09	100%	3	R41	Category 1	Category I	0.696

* - COLIPA formulation slightly different than the CTFA formulation

5.3 Compilation of results, statistical approaches used: description & rationale for the approach used to determine between-laboratory reproducibility

Reproducibility was assessed in the forgoing sections mainly by calculating mean MRD_{50} values from individual CM or SM runs for a specific test material and comparing these values to the mean MRD_{50} values for the same material when conducted in another laboratory or in several other laboratories. The mean MRD_{50} values from the individual laboratories were compared by calculating their mean and SD and then calculating the CV. We feel that this is a reasonable method of comparison when the results from at least three laboratories are being compared, but in some cases data from only two laboratories are available, e.g. the COLIPA study. Because of the small sample size on which it is based, we feel that in these cases the CV is probably not an exceptionally strong statistic.

In some cases it was possible to present correlation coefficients to compare groups of laboratories. These statistics were valuable in such instances as the EC/HO study where 4 laboratories were involved. By conducting a lab vs. lab comparison, we were able to determine if one of the groups of laboratories appeared to be an “outlier” from the others. However, correlation coefficients do not address how closely actual values match each other, only the trends between a set of values. Two laboratories which have good correlation coefficients could actually have dissimilar values as long as one lab was always consistently higher or lower than the other lab.

We also investigated whether the hazard predictions produced by the participating laboratories were reproducible. We presented this information in a detailed tabular form as well as in summary tables. The only statistic that we used was the percentage of materials for which all (or a stated portion) of the laboratories predicted identical hazard classifications. We were unable to find a more quantitative measure of reproducibility for the hazard classifications.

5.4 Additional studies where raw data are not available: attempt to combine the data using weight-of-evidence approaches

The two major studies (COLIPA and EC/HO) (plus an analysis of the overlapping test materials between the studies) and the one smaller study (Bagley, Bruner et al. 1992) reported here give the best indication of the between laboratory reproducibility of the assay. These studies were performed with blinded test materials, in multiple laboratories, in different countries, and with different sources of media. However, with one exception, the results appeared to be quite comparable across a range of materials with the caveat that only materials that are water–soluble can be adequately tested. These were the only studies where identical test materials were tested in at least two different laboratories. We are unaware of other studies which have overlapping materials that could be used to combine with these studies in a weight of evidence approach.

6. Predictive Capacity (Module 5)

As discussed in Module 1, the CM assay monitors the real time action of increasing test article dilutions on a monolayer culture of cells. The cytotoxicity of the L929 cells is observed by the CM as a decrease in the pH surrounding the cells. It has been proposed that similar cytotoxic actions of test materials on the epithelium, stroma, and endothelial cells of the human eye are a major factor in causing ocular irritation (Jester, Petroll et al. 1998; Jester, Li et al. 2001; Maurer, Parker et al. 2002). Historically, the materials tested in the CM assay have been surfactant-based household cleaning and personal care products and ingredients. Because of the restrictions of this dilution based assay, it is currently used primarily for evaluating the eye irritation potential of liquid surfactant containing formulations and mixtures.

Although data from human experience (e.g. accidental exposure) or approved clinical studies would be the best way to assess the predictive capacity of the CM, almost all studies cited in this BRD have used the Draize rabbit eye test MAS or MMAS values as the standard for a quantitative measure of eye irritation. One exception was the CTFA Phase III study which also evaluated the CM by its ability to predict the US Federal Hazardous Substance Act categories or the Kay-Calandra hazard categories. None of the studies addressed the ability to address the EU, GHS or EPA hazard classifications as we have in this BRD.

Regardless of the fact that summarized Draize MAS or MMAS scores were routinely used in the cited manuscripts, we were able to obtain raw data from the animal tests of approximately half of the studies evaluated here. This allowed us to calculate the EU, GHS or EPA irritation categories based on published criteria. For this subset of studies we then used the EU, GHS, and EPA categories as the standard against which we could judge the predictive capacity of the CM test. For studies where only summary data were available we report the predictive capacity based on the ability to predict Draize MAS or MMAS scores.

A significant problem in analyzing how well any *in vitro* test predicts the outcome of an *in vivo* test is that a single value is generally associated with the animal score and this single value is treated as a “gold standard”. In reality, there is no single eye irritation value that characterizes a test material; the value that is obtained will generally vary each time the material is tested. Thus it is extremely unlikely that an *in vitro* score and an *in vivo* score will match exactly, no matter how perfect the *in vitro* test is performed. This fact is often overlooked in most validation studies. Generally the animal score is treated as a single fixed value (since the animal test is generally conducted only once), and the *in vitro* test is then assessed for its “accuracy” based on how well its data match that of the animal test. Only a few studies, e.g the CTFA Phase III study, have taken the animal test variability into account. The CTFA study used bootstrap resampling to estimate with-in group variability for each test material so that Draize scores could be represented more realistically with their variability.

As mentioned above, one reason that Draize MAS scores are usually treated as unvarying values is that both ethical and financial considerations generally demand that a rabbit eye test only be conducted a single time. Thus for many materials there is no information about what score might occur in a repeat test, and without the results of multiple tests it is difficult to address variability.

However, there is one approach which can supply some quantitative insight into this problem. Because over the years the Draize test protocol has evolved from a six rabbit test to a three rabbit test, there is one way of estimating variability for materials which were tested with the six rabbit protocol. It is possible to analyze the ocular response of the six rabbits by placing them into smaller groups. For example, the results for each of the six individual rabbits can be recombined into multiple unique groups of three rabbits (matching the number of rabbits used in today's standard protocol). In fact, all rabbits (designated A – F in the following example) in a six rabbit test can be recombined into 20 unique three rabbit groups, e.g. ABC, ABD, ABE, ABF, etc. This is an approach already used by others in studies to determine the necessary sample size for a rabbit ocular irritation test (DeSousa, Rouse et al. 1984). Each three rabbit group can then be given a hazard classification according to the published guidelines from specific regulatory bodies. The number of subgroups in each hazard classification can then be viewed as a measure of the variability of the test. If all 20 subgroups are classified as R36, for example, then the R36 classification for that material can be considered not very variable. However, if 10 subgroups are rated as No Label and the other 10 are rated as R41, then the results for that material would be considered quite variable. In essence the above results mean that if the material were tested in multiple three rabbit tests, half of the tests would rate it as a very severe R41 material, and the other half of the tests would rate it as a mild No Label material. Therefore, an *in vitro* test of the same material should not necessarily be expected to always make a prediction of R41, which would be the overall prediction of the six rabbit test.

To demonstrate the level of Draize test variability which occurs in the real world, we have examined the animal data from the CTFA Phase III study. This study had arguably one of the best controlled animal studies because it was conducted under GLP's and utilized a randomized block design (3 males and 3 females) with each animal's dosing initiated on a separate day.

Table 6.0 shows for the CTFA Phase III study the number of three rabbit subgroups which fall into each of the hazard categories for the three regulatory classification schemes (GHS, EU, and EPA). Data which support these classifications can be found in spreadsheets contained in Annex I) CTFA Animal Data) It can be seen that in some cases all of the three rabbit subgroups give the same hazard classification as the six rabbit study, e.g. the EU classification for HZB, HZC and HZD is No Label, and each of the 20 three rabbit subgroups for each test material is also No Label. However, for those same three test materials classified by GHS criteria there is considerable difference between the subgroups and the original six rabbit study. For example, HZC is No Label by the six rabbit test, but only half (10) of the three rabbit groups are No Label; seven are 2B and 3 are category 1. This means if the test were repeated 20 times using the current three rabbit

protocol there would be an equal chance of having a higher than No Label score (10 out of 20 times) as there would be of having the No Label score (10 out of 20 times). Similar results can be seen for many of the materials in this study.

Even more dramatic examples can be found in the CTFA Phase III study. HZE, for example, is classified R41 by the six rabbit test, but only 10 of the subgroups have R41 classifications, the other 10 are No Label! Thus if the three rabbit test were run only once, there would be a 50% chance of having the lowest classification (No Label) and an equal chance of having the highest label (R41). HZP is another interesting example. Although it has a 6-rabbit GHS classification of No Label, 6 out of 20 tests (30% of the time) give a Category 1 result – three categories higher than that determined by the 6 rabbit test! Other interesting examples are highlighted in bold in the table.

Table 6.0 Recombination of each 6 rabbit test result into 20 three rabbit test subgroups. Each subgroup was classified separately according to the rules for each of the three classification systems, and the number of subgroups falling into each hazard category is indicated. Numbers in bold, shaded areas represent results from test materials where the subgroups differed in their hazard classification from the overall six rabbit classification. Data from the CTFA Phase III study. N = 25 materials. Raw animal data from Annexes I55 & I3.

		6 animal study score			GHS Counts				EU Counts			EPA Counts			
		GHS	EU	EPA	1	2A	2B	NL	R41	R36	NL	I	II	III	IV
Shampoo 7	HZA	1	R41	1	**16**	**4**	0	0	**16**	**3**	**1**	**16**	**4**	0	0
Liquid Soap 1	HZB*	NL	NL	3	0	0	**4**	**16**	0	0	20	0	0	20	0
Shampoo 1	HZC*	NL	NL	3	**0**	**0**	**10**	**10**	0	0	20	0	0	20	0
Shampoo 5	HZD*	NL	NL	3	0	0	0	20	0	0	20	0	0	20	0
Gel Cleaner	HZE	NL	R41	1	**10**	**0**	**0**	**10**	**10**	**0**	**10**	**10**	**0**	**10**	0
Baby Shampoo 2	HZF	1	R41	1	**16**	**4**	0	0	**16**	**3**	**1**	**16**	**4**	0	0
Shampoo 8	HZG*	NL	NL	3	**0**	**0**	**0**	**20**	0	0	20	0	0	20	0
Eye Makeup re.	HZH	NL	NL	4	0	0	0	20	0	0	20	0	0	0	20
Skin Cleaner	HZI	1	R41	1	**19**	**1**	0	0	**19**	**1**	0	**19**	**1**	0	0
Mild Shampoo	HZJ	NL	NL	4	0	0	0	20	0	0	20	0	0	0	20
Bubble bath	HZK	1	R41	1	20	0	0	0	20	0	0	20	0	0	0
Foam Bath	HZL	1	R41	1	**19**	**0**	**1**	0	**19**	**0**	**1**	**19**	**0**	**1**	0
Shampoo 3	HZM*	NL	NL	3	0	0	0	20	0	0	20	**0**	**0**	**10**	**10**
Shampoo 6	HZN*	NL	NL	3	0	0	0	20	0	0	20	0	0	20	0
Baby Shampoo 1	HZP	NL	NL	3	**0**	**0**	**0**	**20**	0	0	20	0	0	**19**	**1**
Cleaning Gel	HZQ	NL	NL	3	**0**	**0**	**0**	**20**	0	0	20	0	0	20	0
Facial Cleaning Foar	HZR*	NL	R41	1	**10**	**0**	**3**	**7**	**10**	**0**	**10**	**10**	**0**	**10**	0
Shower Gel	HZS	1	R41	1	**19**	**1**	**0**	0	**19**	**1**	0	**19**	**1**	0	0
Polishing Scrub	HZT	NL	NL	4	0	0	0	20	0	0	20	0	0	0	20
Hand Soap	HZU*	NL	NL	3	**0**	**0**	**4**	**16**	0	0	20	0	0	20	0
Shampoo 4	HZV*	NL	NL	3	0	0	0	20	0	0	20	0	0	20	0
Liquid Soap 2	HZW*	2B	NL	3	**0**	**0**	**16**	**4**	0	0	20	0	0	20	0
Shampoo 2	HZX	1	R41	1	**19**	**1**	0	0	**19**	**0**	**1**	**16**	**4**	0	0
Shampoo AntiD	HZY	1	R41	1	**16**	**4**	0	0	**16**	**4**	0	**16**	**4**	0	0
Facial Cleaner	HZZ	NL	NL	4	0	0	0	20	0	0	20	0	0	0	20

* tested at 25% (w/v) in vivo and in vitro (starting material)

The conclusion from studying this example is that neither a Draize MAS score nor a hazard classification is an unvarying physical constant for the test material. Therefore, an *in vitro* test should not be expected to exactly match a hazard category determined *in vivo*

because the next time the animal test is run it might also fail to match the hazard classification of the first animal test.

6.1 Studies with available raw data

There were 4 studies where sufficient raw data was available to attempt to determine predictive capacity for regulatory classification schemes. For the vast majority of the test materials in the EC/HO study, the CTFA Phase III study, and the COLIPA study, individual animal and tissue data from a traditional Draize test were available and sufficient to allow the unequivocal determination of the EU, GHS, or EPA eye irritation category. However, some of the animal tests were conducted in such a way that all the appropriate data for the determination of EU, GHS, or EPA category were not available. In these cases, the test material was left out of the analysis.

It should also be noted that the Draize test for the CTFA Phase III study was conducted using ocular anesthesia, whereas the Draize tests for the EC/HO study and the COLIPA study did not use anesthesia. It is expected that the eye irritation results for the same compounds may differ because of this difference in protocol.

For the Company # 1 unpublished data studies, the animal ocular irritation tests were conducted using the Low Volume Eye Test (LVET). The LVET uses the same scoring scale as the Draize test, but is conducted with one-tenth the volume (100 µL) applied to the center of the cornea as opposed to the conjunctival sac. The LVET was used since it has been proposed to be more predictive (yet still over predictive) of the human response than the traditional Draize test (Griffith, Nixon et al. 1980; Freeberg, Nixon et al. 1986).

Method of analysis

In general, all of the following analyses were conducted in a similar fashion. First the raw animal data, which consisted of tissue scores for individual animals taken at designated time points until the lesions had cleared or until 21 days had elapsed, were inserted into spreadsheets constructed to apply the rules established by the GHS, EU, and EPA scoring systems. These spreadsheets then returned the classification for that material. In some cases, the appropriate data from the animal test had not been recorded or supplied for the chemical or formulation in question. In cases where the hazard classification of the test material could not be unambiguously categorized, it was dropped from the analysis. Thus some of the graphs in the following sections will have fewer data points than the number of test materials listed in some of the previous tables.

Next, the assigned categories were given a scale ranking from 1 (highest) to 3 (lowest for the EU scale) or 4 (lowest for the GHS and EPA scale) and then matched with the appropriate *in vitro* score. These paired numbers were then graphed (GraphPad Prism®, GraphPad Software, Inc., San Diego, CA) on an XY graph with the abscissa being the numerical hazard category. Additional axis labels were then added to indicate the hazard class, and at the same time the now duplicative numerical rankings were removed.

Specific breakpoints between hazard categories could then be visualized as places where *in vitro* cut-off lines could be drawn.

These cut-off values, i.e. prediction models, determined post-hoc were sometimes slightly different from the proposed prediction models which had been submitted to the ECVAM management team in September 2006. The changes occurred because the originally proposed cut-offs were based on only preliminary analysis of hazard category data. Prior to this time the only published PM for the CM had been for the prediction of Draize scores. See Section 2.2.5 of this BRD for a more detailed description of the determination of the prediction models.

For studies where insufficient animal data were supplied to determine the hazard classification, an XY graph was constructed relating the *in vitro* scores to the Draize score. Although these graphs will not be directly helpful in determining the predictivity of the SM or CM for hazard classes, they may be helpful in assessing the general ability of the *in vitro* test to identify more or less irritating materials.

Subsequent to the construction of the scatter plots with the proposed cut-offs for the various hazard classifications, we prepared contingency tables to summarize the performance of the CM test and its prediction model in each of the major studies (and subsets of the studies where specific chemical classes were investigated). The parameters used to summarize the performance are reasonably standard for the analysis of toxicity tests and are defined below:

Concordance – the percentage of materials predicted by the CM to have the same hazard classification as determined by the rabbit test

Predictivity – the proportion of materials assigned to a specific hazard category by the animal test which were assigned to the same category by the CM.

In addition, analyses of how well the CM performed in separating “severe irritants” (those materials in the highest irritation category of any of the classification systems) from the rest (those in the remaining categories). The parameters used in this analysis are:

Concordance – the percentage of correctly identified severe irritants and those materials which were not severe irritants (a combination of mild irritants and non-irritants).

Sensitivity – the number of correctly identified severe irritants by the CM assay as a proportion of the number of actual severe irritants.

Specificity – the number of correctly identified non-severe irritants (“the rest”) as a proportion of the total number of actual non-severe irritants.

Positive Predictivity –the number of correctly predicted severe irritants as a proportion of the total number of predicted severe irritants.

Negative Predictivity –the number of correctly predicted non-severe irritants ("the rest") as a proportion of the total number of predicted non-severe irritants.

False Positive Rate – the number of non-severe irritants predicted by the CM to be severe irritants as a proportion of the total number of non-severe irritants.

False Negative Rate – the number of severe irritants predicted to be non-severe irritants as a proportion of the total number of severe irritants.

A third analysis of the ability of the CM to separate non-irritants (those materials in the lowest irritation category of any of the classification systems) from the rest (those in the remaining categories). The parameters used in this analysis are:

Concordance – the percentage of correctly identified non-irritants and irritants.

Sensitivity – the number of correctly identified irritants (e.g R36 plus R41) by the CM assay as a proportion of the number of actual irritants.

Specificity – the number of correctly identified non-irritants by the CM assay as a proportion of the total number of actual non-irritants.

Positive Predictivity – the number of correctly predicted irritants as a proportion of the total number of predicted irritants i.e. this analysis only determined that irritants were predicted to be irritants by the CM assay but did not distinguish between the different irritancy classifications predicted (e.g., R36 and R41).

Negative Predictivity – the number of correctly predicted non-irritants as a proportion of the total number of predicted non-irritants.

False Positive Rate – the number of non-irritants predicted to be irritants as a proportion of the total number of non-irritants.

False Negative Rate – the number of irritants predicted to be non-irritants as a proportion of the total number of irritants.

6.1.1 Relevant information for each study where raw data are available

Table 6.1.1 Table presenting the relevant information for each study where raw data are available

Studies	No. of chemicals	No. of products	Coded	No. of labs	No. of runs/ exp	Data format (raw, summary)	Chemical classes covered	Ranges of toxicity covered	Physico-chemical properties covered
EC/HO[1] (Balls, Botham et al. 1995)	31 of 60 tested	0	Coded	4	3-4	Raw data	Alcohols, surfactants, ketones, bases, acids, esters	MAS 0-108 EU: R36, R41, NC GHS: 1, 2A, 2B, NC EPA: I-IV	See Annex A and B
CTFA[2] (Gettings, Lordo et al. 1996)	0	24 of 25 tested	Coded	1	3	Raw data	Personal care products and their ingredients, mostly surfactant based	MAS 2.3-44 EU: R41, NC GHS: I, 2A, 2B, NC EPA: I-IV	Liquids and water-soluble materials
COLIPA[1] (Brantom, Bruner et al. 1997; Harbell, Osborne et al. 1999)	17 of 23 tested	9 of 32 tested	Coded	2	3	Raw data	Alcohols, surfactants, ketones, bases, acids, esters	MAS 0–106 EU: R36, R41, NC GHS: 1, 2A, 2B, NC EPA: I-IV	See Annex A and B
Company 1[1 & 2] (Unpublished)	76 unique of 101 tested		Coded	1	≥3	Raw data	Household and personal care products and ingredients	LVET MAS 0-65 EU: R36, R41, NC GHS: 1, 2A, 2B, NC EPA: I-IV	Liquids and water-soluble materials

[1] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[2] Performed using the silicon microphysiometer coverslip protocol with a 500-second exposure and L929 cells.

6.1.2 *In vivo* reference data used to assess the performance of the alternative method

Most of the reference data for the larger validation studies were collected from the exposure of rabbits using the standard Draize methodology. Within this general protocol; however, there could be significant variations including the use of anesthesia, the use of 1 to 12 animals, level of GLP compliance, level of blinding of the study, etc. In addition, most of the studies suffered from the confounder that historical data was used for the reference. In such studies it is often difficult, if not impossible, to determine after the fact (perhaps many years later) the composition and purity of the materials that were originally tested. Thus it becomes problematic whether the materials tested *in vitro* are at all similar to the materials that were tested *in vivo*.

Of the studies analyzed in this BRD, only the CTFA Phase III study utilized a concurrent testing scheme for all the test materials (Brantom, Bruner et al. 1997). Single large lots of the formulations to be tested were divided between the animal studies and the *in vitro* studies. Thus there was confidence that the materials used for *in vitro* testing were indeed the same materials that were tested *in vivo*. Although all of the *in vivo* data were collected before the *in vitro* testing began, the time differential was so small that it is very unlikely that significant changes occurred in the formulations before the *in vitro* testing commenced. COLIPA used concurrent testing for the 32 formulations which were formulated specifically for their program (Brantom, Bruner et al. 1997) following formulations used for the CTFA Phase I (Feder, Lordo et al. 1991), Phase II (Gettings, Dipasquale et al. 1994), and Phase III (Gettings, Lordo et al. 1996) evaluations.

Although the larger validation studies utilized a more or less standard Draize test, the extensive internal data supplied by the Company # 1 used the LVET as the reference standard. The LVET is similar to the Draize test (e.g. it has identical tissue scoring), but it uses a smaller dose of test material. The volume of material applied to the eye in the LVET is 10 µl – ten times less than the 100 µl applied in the traditional Draize. In addition, the LVET protocol calls for the test material to be applied directly to the surface of the cornea, while in the traditional Draize the test material is placed inside the conjunctival sac. The rational behind the LVET protocol is that it is said to be more predictive of the human reaction to an accidental ocular exposure than the Draize test which is thought to significantly over predict the human reaction. Support for this view comes from studies which show that when controlled human clinical studies are conducted concurrently with both the traditional Draize and the LVET, the LVET more closely matches the clinical response (Freeberg, Nixon et al. 1986). In general, both protocols seem to over predict the human response, but the LVET is claimed to over predict by a smaller amount than the traditional Draize. When LVET data are used as part of determining the predictive capacity of the CM or SM, that fact will be clearly noted. Table 6.1.2 provides a general description of the study design and methods of collection for the animal reference data examined in this BRD.

Table 6.1.2 Table presenting the *in vivo* reference data available

Studies	Species and protocols used as reference data	Sources of information	n. of labs	n. of chem	n. of animals	Quality of the reference data	Format of data available
EC/HO[1] (Balls, Botham et al. 1995)	Albino rabbits using traditional Draize and OECD TG405. No anesthesia. Whether every chemical was tested this way could not be determined	(Balls, Botham et al. 1995)	Labs not identified; number unknown	31 of 60 tested	>3	*In vivo* – principles of GLP *In vitro* – principles of GLP for at least 1 lab	*In vivo* – individual animal tissue scores *In vitro* – summaries (means); some raw data have been obtained (Annex L)
CTFA[2] (Gettings, Lordo et al. 1996)	Albino rabbits using traditional Draize and an experimental block design. Anesthesia used.	(Gettings, Lordo et al. 1996)	1	24 of 25 tested	6 total; 3 males, 3 females	*In vivo* GLP *In vitro* – no Concurrent testing	*In vivo* – individual animal tissue scores *In vitro* – each trial Raw data in Annex I
COLIPA [1] (Brantom, Bruner et al. 1997)	Assumed to be albino rabbit using traditional Draize test, but protocols varied for the formulations. No anesthesia used	(Brantom, Bruner et al. 1997)	Labs not identified; number unknown for 23 materials; 1 lab for 32 formulations	26 of 55 tested	Varied, 1-6	*In vivo* – GLP status not known *In vitro* - no	*In vivo* – individual animal and tissue scores *In vitro* – each trial Raw data in Annex J
Company 1[1&2] (Unpublished)	Albino rabbit using LVET protocol. No anesthesia used.	Communication from the Company 1	Labs not identified; number unknown	76 of 101 tested	Varied, 1-6	Generally GLP	Complete animal data (LVET) and summary MRD_{50} Raw data in Annex K

[1] Performed using the Cytosensor transwell protocol with an 810-second exposure and L929 cells.
[2] Performed using the silicon microphysiometer coverslip protocol with a 500-second exposure and L929 cells.

6.1.3 Brief description of the studies with available raw data

6.1.3.1 Analysis of the EC/HO study (Balls, Botham et al. 1995)

The reference data for the EC/HO study are all from historic tests; consequently it is difficult to know whether lack of concordance between the animal study and the *in vitro* study for individual chemicals is due to the result of underperformance of the *in vitro* test or to an intrinsic difference between the chemical that was tested in the animal and the chemical that was tested *in vitro*. Positive points about the *in vivo* data used for this study are that the tests were supposed to have been conducted since 1981 and according to OECD TG405 following the principles of GLP. The tests were not to have used anesthesia and to have been conducted long enough to enable reversibility/irreversibility to be assessed. Although the above were stated to be general rules for the selection of data, it is not clear from the publication whether all of these conditions were actually met for each of the chemicals selected for the study. Only some raw data could be obtained and they are shown in Annex L.

Four laboratories contributed CM data for this study (Annex H13), but there was not unanimity as to which chemicals could be appropriately tested in the CM and which could not. Consequently there are not four MRD_{50} values for each chemical. Further analysis was carried out by only using data where two or more labs produced results. Therefore, unless noted, all the graphs in this chapter contain mean MRD_{50} data (from 2, 3, or 4 labs) for 31 chemicals. See section 5.2.2.2 for more detail.

An overall summary of the EC/HO study including the chemical identities, animal scores and *in vitro* scores are given in Table 6.1.3.1.a. Graphical presentations of the results, for all materials tested by 2, 3, or 4 labs are given in Figure 6.1.3.1.a for the EU classifications, Figure 6.1.3.1.b for the GHS classifications, and Figure 6.1.3.1.c for the EPA classifications.

Table 6.1.3.1.a Summary of the EC/HO study (Balls, Botham et al. 1995). Surfactant materials are highlighted. N = 20 non-surfactant materials and N = 11 surfactant materials.

In Vivo Eye Irritation Classifications - EC/HO study

EC/HO chemical number	Substance	CASRN[1]	Concentration Tested	Purity (%)	n. of animals	*In Vivo* EU[2,3]	*In Vivo* GHS[4,5]	*In Vivo* EPA[6,7]	ECETOC MMAS Score[8]	MRD_{50} (mg/mL)[11]
1	Sodium hydroxide (10%)	1310-73-2	10%	reagent grade	1	R41	Category 1	Category I	108	2.26
2	Benzalkonium chloride (10%)	63449-41-2	10%	98	3	R41	Category 1	Category I	108	0.39
3	Trichloroacetic acid (30%)	76-03-9	30%	reagent grade	1	R41	Category 1	Category I	106	1.79
4	Cetylpyridinium bromide (10%)	140-72-7	10%	99	6	R41	Category 1	Category I	89.7	1.26
5	Cetylpyridinium bromide (6%)	140-72-7	6%	99	4	R41	Category 1	SCNM[9]	85.8	0.87
6	Benzalkonium chloride (5%)	63449-41-2	5%	98	4	R41	Category 1	Category I	83.8	1.13
9	Cyclohexanol	108-93-0	100%	97	4	R41	Category 1	Category I	79.8	8.03
11	Promethazine HCl	58-33-3	100%	98	3	R41	Category 1	SCNM	71.7	1.27
13	Triton X-100 (10%)	9002-93-1	10%	98	6	R41	Category 1	Category II	68.7	1.82
14	Acetone	67-64-1	100%	99	4	R36	Category 2A	Category II	65.8	148.82
18	Isobutanol	78-83-1	100%	99.9	4	R36	Category 2A	Category II	60.3	27.91
19	Imidazole	288-32-4	100%	99	3	R41	Category 1	SCNM	59.3	23.69
20	Sodium lauryl sulfate (15 %)	151-21-3	15%	98	6	R41	Category 1	Category I	59.2	0.62
23	Methyl ethyl ketone	78-93-3	100%	99	4	R36	Category 2A	Category III	50	58.08
24	Pyridine	110-86-1	100%	>99.9	3	R41	Category 1	SCNM	48	19.73
26	Benzalkonium chloride (1 %) (1x)	63449-41-2	1%	98	4	R41	Category 2A	Category I	34.4	4.53
28	Gammabutyrolactone	96-48-0	100%	>99	3	R36	Category 2A	Category II	43	93.79
31	Methyl acetate	79-20-9	100%	98	4	R36	Category 2A	Category II	39.5	94.68
34	Triton X-100 (5 %) (1x)	9002-93-1	5%	98	6	R36	Category 2A	Category III	32.3	3.23
36	Isopropanol	67-63-0	100%	99.9	4	Not classified	Category 2A	Category III	30.5	101.26
37	Sodium perborate, $4H_2O$	10486-00-7	100%	98.6	6	R41	Category 1	Category I	30.5	1.69
39	2,5-Dimethylhexanediol	110-03-2	100%	99.5	3	R41	Category 1	Category I	28.3	98.67
40	Methyl cyanoacetate	105-34-0	100%	99	3	R36	Category 2A	Category II	27.7	21.54
41	Sodium hydroxide (1%)	1310-73-2	1%	reagent grade	4	R36	Category 2B	Category III	25.8	27.04
42	Ethanol	64-17-5	100%	n.p.	3	Not classified	Category 2A	Category III	24	111.99
44	Ammonium nitrate	6484-52-2	100%	99.999	3	R36	Category 2A	Category III	18.3	71.27
46	Sodium lauryl sulfate (3 %)	151-21-3	3%	98	6	Not classified	No category	Category III	16	3.28
53	Trichloroacetic acid (3%)	76-03-9	3%	reagent grade	6	Not classified	No category	Category III	6.7	15.03
55	Tween 20	9005-64-5	100%	98	4	Not classified	No category	Category III	4	3.27
58	Cetylpyridinium bromide (0.1%)	140-72-7	0.10%	99	6	Not classified	No category	Category III	2.7	84.65
59	Glycerol	56-81-5	100%	>99.5	6	Not classified	No category	Category IV	1.7	129.88

[1]CASRN=Chemical Abstract Services Registry Number

[2]EU=European Union (EU [2001]).

[3]Risk phrase R41 = risk of serious damage to the eyes; R36 = irritating to the eyes; not classified.

[4]GHS=Globally Harmonized System (UN [2003])

[5]Eye Irritant Category 1 = irreversible effects on the eye/serious damage to the eye; Category 2A = reversible effects on the eye/irritating to the eyes; Category 2B = reversible effects on the eye/mildly irritating to the eyes; No category

[6]EPA=U.S. Environmental Protection Agency (EPA [1996]).

[7]Toxicity Category I for the Primary Eye Irritation Study = Corrosive, or corneal involvement or irritation not reversible within 21 days; Category II = Corneal involvement or irritation clearing in 8-21 days; Category III = Corneal involvement or irritation clearing in 1-7 daysl; Category IV: minimal effects clearing in less than 24 hr

[9]MMAS scores reported in Balls et al. (1995) and in the ECETOC Technical Report n. 48 (1998)

[9]SCNM=Study Crtieria Not Met

[10]n.p.=not provided

[11] Mean of the 2, 3, or 4 labs reporting results

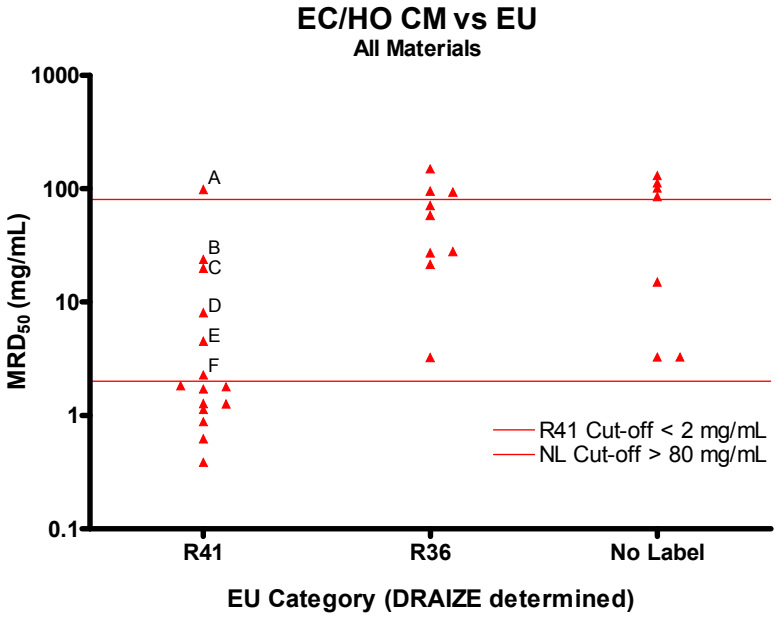

Figure 6.1.3.1.a Results of the EC/HO study related to EU classification. Data points indicate the mean MRD_{50} for the laboratories (2, 3, or 4) that provided data for that chemical. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. All materials including surfactants are included.

Table 6.1.3.1.b Identification of the six R41 materials which were underpredicted by EU classification. The lettering system remains the same for the GHS and EPA scatterplots.

Graph Letter	Material	Chemical Class	Mean MRD_{50}
A	2,5-dimethylhexanediol	Alcohol	98.67
B	Imidazole	Heterocyclics	23.69
C	Pyridine	Heterocyclics	19.73
D	Cyclohexanol	Alcohol	8.03
E	Benzalkonium chloride (1%)	Cationic Surfactant	4.53
F	Sodium hydroxide (10%)	Alkalis	2.26

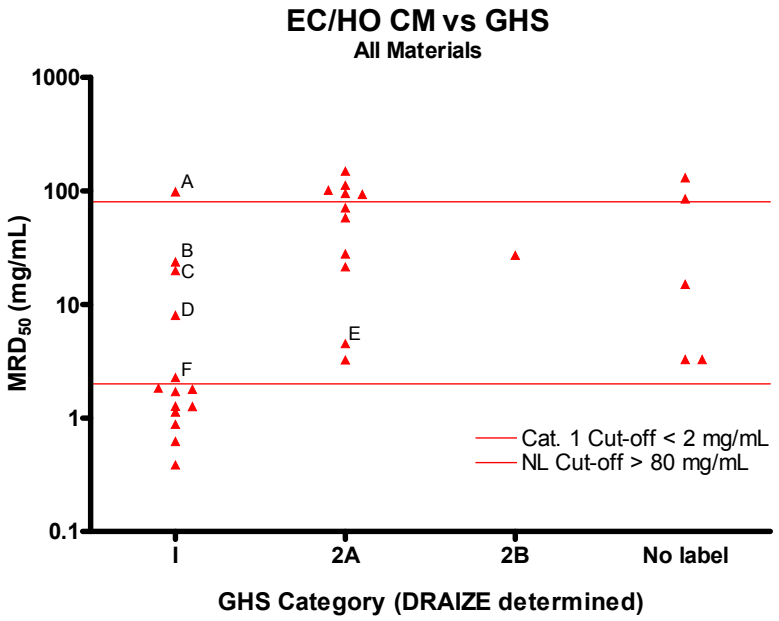

Figure 6.1.3.1.b Results of the EC/HO study related to GHS classification. Data points indicate the mean MRD_{50} for the laboratories (2, 3, or 4) that provided data for that chemical. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. All materials including surfactants are included.

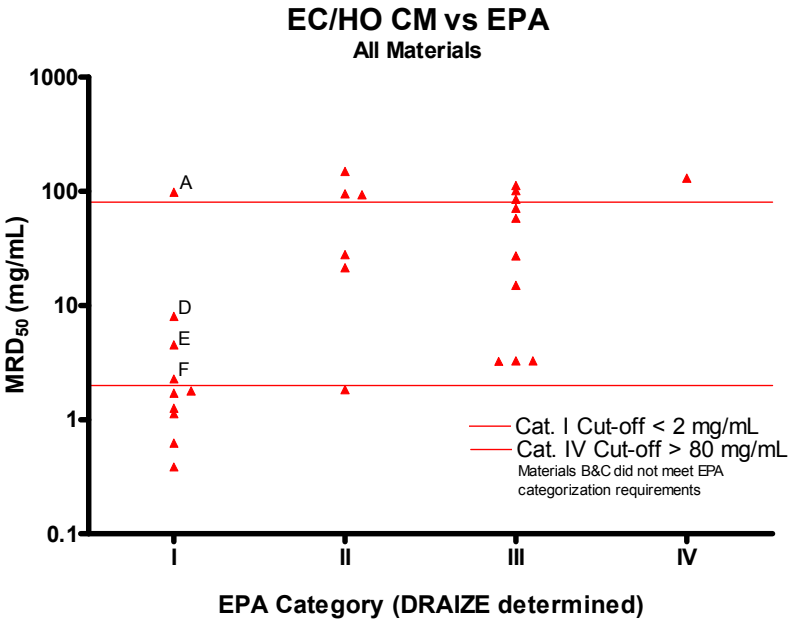

Figure 6.1.3.1.c Results of the EC/HO study related to EPA classification Data points indicate the mean MRD_{50} for the laboratories (2, 3, or 4) that provided data for that chemical. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. All materials including surfactants are included.

Figures 6.1.3.1.a, 6.1.3.1.b, and 6.1.3.1c show cut-off values for MRD_{50} scores that have been empirically chosen to identify, where possible, the various hazard categories. In the case of the GHS system and the EPA system which have 4 categories, the overlap of MRD_{50} response was so large that it was deemed impossible to differentiate between the two middle categories. Hence only upper (to possible identify non-irritants) and lower (to possibly identify severe irritants) cut-off values are shown.

It appears from the graphs that the SM with the transwell protocol does not have the ability to clearly separate the wide range of test materials used in the EC/HO study into the Draize test defined EU, GHS or EPA categories. One exception is that severe irritants seem to be reasonably predicted when MRD_{50} scores of less than 2 are used. Using this lower cut-off value, there is a high positive predictive value for EU category R41 (100%; 9 of 9 chemicals), GHS category 1 (100%; 9 of 9) and EPA category I (86%; 6 of 7 chemicals).

Even though the positive predictive value was high using a lower cut-off of MRD_{50} <2 mg/ml, the sensitivity was lower, with several severe chemicals being under predicted in each hazard classification system. Each of these chemicals is identified on the Figures with a letter code defined in Table 6.1.3.1.b. Within these outliers are 2 of the 3 heterocyclics in the study, 2 of the 9 alcohols, 1 of 6 dilutions of cationic surfactants, and 1 strong alkali.

Contingency tables were created to determine the performance of the SM assay and the proposed cut-offs for each of the hazard classification systems. Results with the EU system are found in Table 6.1.3.1.c, results with the GHS system in Table 6.1.3.1.d and results with the EPA system in Table 6.1.3.e. In all of the cases the proposed cut-off for the most irritating categories resulted in a high positive predictive value (100% for the EU and GHS systems; 85.7% for the EPA system). Predictivity for the less irritating materials was considerably lower.

Table 6.1.3.1.c EC/HO - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values shown in Figure 6.1.3.1.a are applied. N = 31 materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	9	5	1	15	60.0%	NA	40.0%
R36	0	6	3	9	66.7%	0.0%	33.3%
Not Classified	0	3	4	7	57.1%	42.9%	NA
Total	9	14	8	31	61.3%		
Predictivity	100.0%	42.9%	50.0%				
Category Underpredicted	NA	35.7%	50.0%				
Category Overpredicted	0.0%	21.4%	NA				

Table 6.1.3.1.d EC/HO - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values shown in Figure 6.1.3.1.b are applied.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	9	4		1	14	64.3%	NA	35.7%
2A	0	6		5	11	54.5%	0.0%	45.5%
2B	0	1		0	1	100.0%	0.0%	0.0%
No Label	0	3		2	5	40.0%	60.0%	NA
Total	9	14		8	31	58.1%		
Predictivity	100.0%	50.0%		25.0%				
Category Underpredicted	NA	28.6%		75.0%				
Category Overpredicted	0.0%	21.4%		NA				

Table 6.1.3.1.e EC/HO - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values shown in Figure 6.1.3.1.c are applied. N = 27 materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	6	3		1	10	60.0%	NA	40.0%
II	1	2		3	6	33.3%	16.7%	50.0%
III	0	7		3	10	70.0%	0.0%	30.0%
IV	0	0		1	1	100.0%	0.0%	NA
Total	7	12		8	27	59.3%		
Predictivity	85.7%	75.0%		12.5%				
Category Underpredicted	NA	25.0%		87.5%				
Category Overpredicted	14.3%	0.0%		NA				

An additional analysis was conducted to identify the ability of the *in vitro* assay to make two different binary classifications: severe irritants versus the rest, and non-irritating materials versus the rest. Severe irritants were defined to be those materials falling in the highest category (EU R41, GHS Category 1, or EPA category I) for each of the classification systems. Non-irritating materials were defined as those falling in the lowest category (EU Not Classified, GHS No Category, and EPA IV) for each of the classification systems. These results are given in Table 6.1.3.1.f.

Table 6.1.3.1.f EC/HO - Concordance table for severe irritants versus the rest and non-irritants versus the rest. Cut-off values from Figures 6.1.3.1.a, 6.1.3.1.b, and 6.1.3.1.c were applied. N = 31 materials (EU & GHS) and N = 27 materials (EPA).

	Vs Rest													
	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
EU Severe irritants	81%	25/31	60%	9/15	100%	16/16	100%	9/9	73%	16/22	0%	0/16	40%	6/15
EU non-irritants	77%	24/31	83%	20/24	57%	4/7	87%	20/23	50%	4/8	43%	3/7	17%	4/24
GHS Severe irritants	84%	26/31	64%	9/14	100%	17/17	100%	9/9	77%	17/22	0%	0/17	36%	5/14
GHS non-irritants	71%	22/31	77%	20/26	40%	2/5	87%	20/23	25%	2/8	60%	3/5	23%	6/26
EPA Severe irritants	81%	22/27	60%	6/10	94%	16/17	86%	6/7	80%	16/20	6%	1/17	40%	4/10
EPA non-irritants	74%	20/27	73%	19/26	100%	1/1	100%	19/19	13%	1/8	0%	0/1	27%	7/26

It can be seen from the above table that positive predictivity was relatively high for most of the situations, but only with the identification of severe irritants was the negative predictivity also relatively high (73%, 77%, and 80% for the EU, GHS and EPA systems, respectively).

Surfactant analysis

Since in recent years the applicability domain of the SM or CM assay has become more narrowly defined to be limited to test materials that are completely water soluble, and since much of the SM and CM testing over that same time period has been surfactants and surfactant-containing products, it was decided to investigate only the surfactants (no surfactant-based formulations were in the data set) from the EC/HO chemical set. There were 11 surfactants which had data from two or more labs. Figures 6.1.3.1.d, 6.1.3.1.e, and 6.1.3.1.f show the results of that analysis relative to EU, GHS, and EPA classifications, respectively.

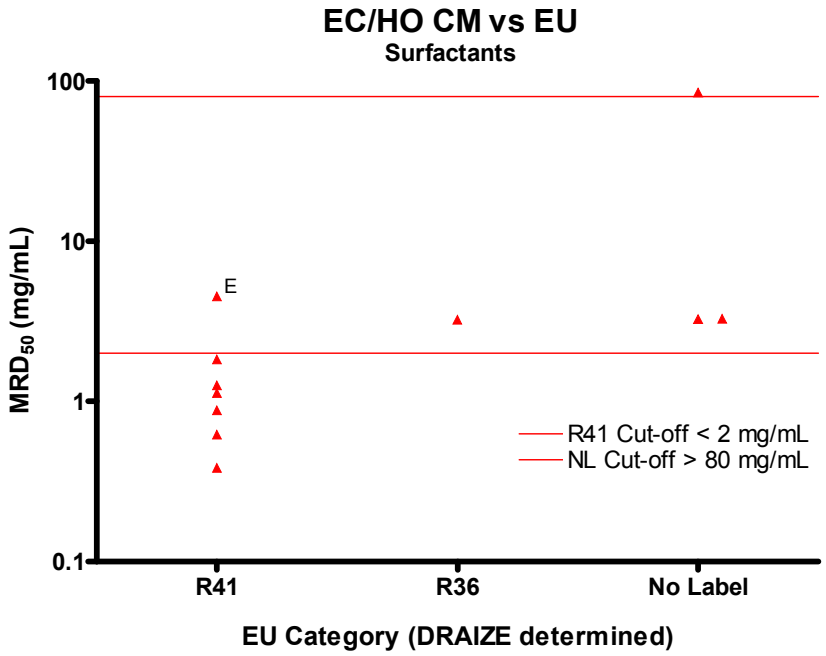

Figure 6.1.3.1.d Results of only the surfactants from the EC/HO study related to EU classification. Data points indicate the mean MRD_{50} for the laboratories (2, 3, or 4) that provided data for that chemical. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

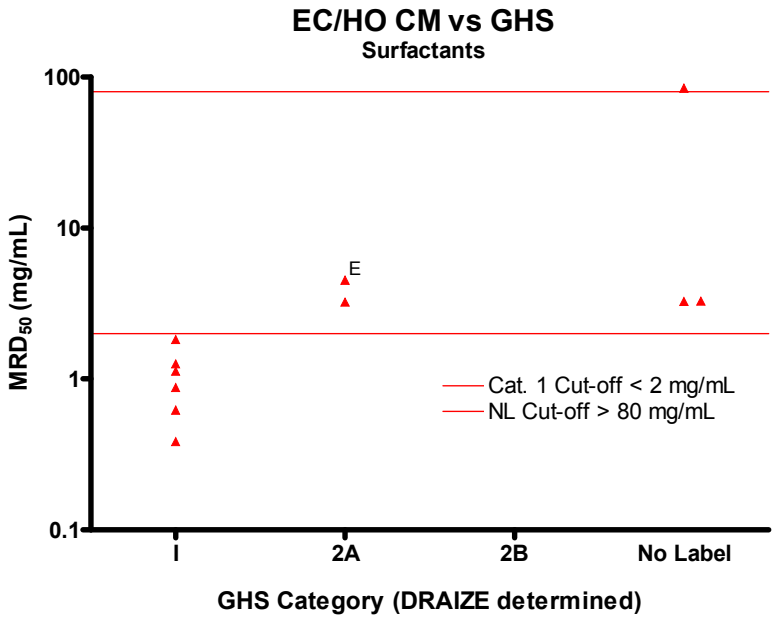

Figure 6.1.3.1.e Results of only the surfactants from the EC/HO study related to GHS classification. Data points indicate the mean MRD_{50} for the laboratories (2, 3, or 4) that provided data for that chemical. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

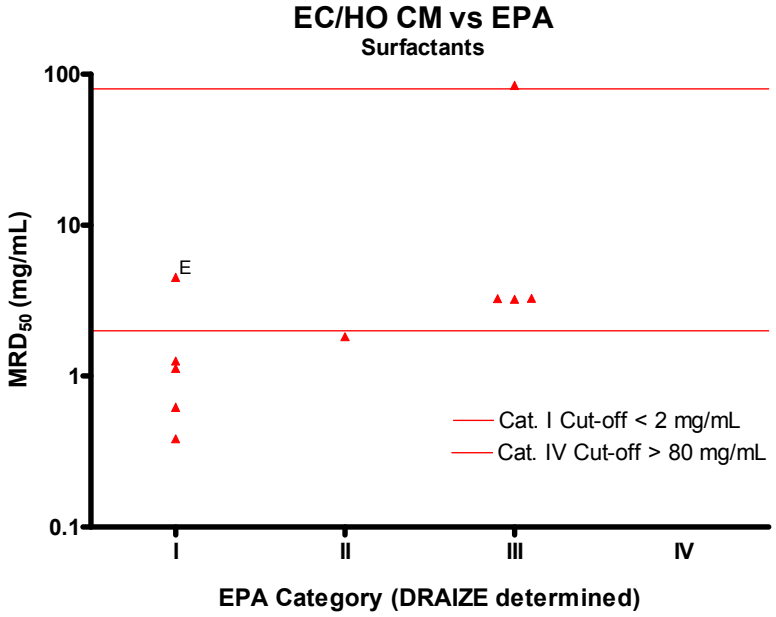

Figure 6.1.3.1.f Results of only the surfactants from the EC/HO study related to EPA classification. Data points indicate the mean MRD_{50} for the laboratories (2, 3, or 4) that provided data for that chemical. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

The smaller number of data points (11 for the EU and GHS analysis; 10 for the EPA analysis) make it difficult to set cut-off values based on these data sets alone. Therefore, we have kept the cut-offs used when analyzing the full range of chemicals, but it should be noted that these might not be optimal when a greater number of surfactants or surfactant-containing materials are assessed. Because of the limited amount of data, it is difficult to determine if the CM adequately separates the classifications for the EU, GHS, or EPA systems although (as with the complete set of chemicals) the lower cut-off of <2 mg/ml generally results in a high positive predictive value for R41 (6 of 6 materials; 100%), GHS 1's (6 of 6 materials; 100%), or EPA I's (4 of 5 materials; 80%). The one R41 and EPA I material which was under classified was 1% benzalkonium chloride.

Although a lower cut-off value may exist which separates EU Not Classified, GHS No Category or EPA Category IV from the other GHS, EU, or EPA Categories (we have hypothesized >80 mg/ml on Figures 6.1.3.1.d, 6.1.3.1.e, and 6.1.3.1.f), we believe there are insufficient data available from this study to make a definitive decision.

One significant difference between the surfactant analysis and the total chemical analysis is that the number of false negative materials is reduced significantly regardless of the hazard classification scheme. Comparing Figures 6.1.3.1.a, 6.1.3.1.b, and 6.1.3.1.c with Figures 6.1.3.1.d, 6.1.6.1.e, and 6.1.3.1.f shows that only material E (1% Benzalkonium Chloride) remains as an under predicted severe irritant in the EU or EPA classification systems.

Table 6.1.3.1.g EC/HO Surfactants - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values shown in Figure 6.1.3.1.d are applied. N = 11 materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	6	1	0	7	85.7%	NA	14.3%
R36	0	1	0	1	100.0%	0%	0%
Not Classified	0	2	1	3	33.3%	66.7%	NA
Total	6	4	1	11	72.7%		
Predictivity	100.0%	25.0%	100.0%				
Category Underpredicted	NA	25.0%	0%				
Category Overpredicted	0%	50.0%	NA				

Table 6.1.3.1.h EC/HO Surfactants - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values shown in Figure 6.1.3.1.e are applied. N = 11 materials.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	6	0		0	6	100.0%	NA	0%
2A	0	2		0	2	100.0%	0%	0%
2B	0	0		0	0	NA	NA	NA
No Label	0	2		1	3	33.3%	66.7%	NA
Total	6	4		1	11	81.8%		
Predictivity	100.0%	50.0%		100.0%				
Category Underpredicted	NA	0%		0%				
Category Overpredicted	0%	50.0%		NA				

Table 6.1.3.1.i EC/HO - SurfactantsContingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values shown in Figure 6.1.3.1.f are applied. N = 10 materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	4	1		0	5	80.0%	NA	20.0%
II	1	0		0	1	0.0%	100.0%	0.0%
III	0	3		1	4	75.0%	0%	25.0%
IV	0	0		0	0	NA	NA	NA
Total	5	4		1	10	70.0%		
Predictivity	80.0%	75.0%		0.0%				
Category Underpredicted	NA	25.0%		100.0%				
Category Overpredicted	20.0%	0%		NA				

Table 6.1.3.1.j EC/HO Surfactants - Concordance table for severe irritants versus the rest and non-irritants versus the rest. Cut-off values from Figures 6.1.3.1.d, 6.1.3.1.e, and 6.1.3.1.f were applied. N = 11 materials (EU & GHS) and N = 10 materials (EPA)

	Vs Rest													
	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
EU Severe irritants	91%	10/11	86%	6/7	100%	4/4	100%	6/6	80%	4/5	0%	0/4	14%	1/7
EU non-irritants	82%	9/11	100%	8/8	33%	1/3	80%	8/10	100%	1/1	67%	2/3	0%	0/8
GHS Severe irritants	100%	11/11	100%	6/6	100%	5/5	100%	6/6	100%	5/5	0%	0/5	0%	0/6
GHS non-irritants	82%	9/11	100%	8/8	33%	1/3	80%	8/10	100%	1/1	67%	2/3	0%	0/8
EPA Severe irritants	80%	8/10	80%	4/5	80%	4/5	80%	4/5	80%	4/5	20%	1/5	20%	1/5
EPA non-irritants	90%	9/10	90%	9/10	NA	0/0	100%	9/9	0%	0/1	NA	0/0	10%	1/10

Table 6.1.3.1.j presents the two binary classification systems of Severe Irritants versus the rest, and Non-Irritants versus the rest. The performance for all situations appears much better for the surfactants than for the analysis with all chemicals; however, there were only 11 data points in the surfactant analysis, while there were 28 data points in the total chemical analysis. In the case of the GHS severe irritants versus the rest there was both 100% positive predictive value and 100% negative predictive value. The analysis for EU non-irritants versus the rest, and GHS non-irritants versus the rest resulted in 100% negative predictive value. Using such a test to identify negative materials so they did not have to be animal tested would be helpful.

6.1.3.2 Analysis of the CTFA Phase III study (Gettings, Lordo et al. 1996)

The reference data for the CTFA Phase III study (Annex I) are arguably the most useful of the animal data used for any of the studies in this BRD. They were all obtained under GLP-compliant conditions and with a randomized block design utilizing three male and three female rabbits for each chemical. There are several advantages to the block design: 1) it simulates to some extent within lab day-to-day variability since for each chemical not all rabbits are dosed on the same day, and 2) it eliminates some of the scoring bias since the scorers read each animal independently and are unaware of which six rabbits were treated with the same test article. However, the main positive point about the study is that the *in vitro* and *in vivo* assays were run nearly currently (separated only by a few weeks) using samples from the same batch of chemical or formulation. The one negative point to this study is that ocular anesthesia was used during the rabbit test. Anesthesia is generally not used when conducting the Draize test, so this set of animal data is not completely compatible with the reference data for most of the other studies addressed in this BRD (see Section 6.1.3.4 for additional discussion).

A single laboratory (Company # 4) contributed SM data (Annex F3) for this study. All 25 chemicals in the study were deemed compatible for testing with the SM. An overall summary of the CTFA Phase III study including the chemical identities, animal scores, and *in vitro* scores are given in Table 6.1.3.2.a. Since these studies were conducted with the SM, for ease of comparison with the other studies in this section of the BRD, the *in vitro* MRD_{50} values have been converted to CM values using the relationship presented in section 2.2.1.

Table 6.1.3.2.a Summary of CTFA Phase III study (Gettings, Lordo et al. 1996). N = 25 materials.

In Vivo Eye Irritation Classifications - CTFA Phase III Study

CTFA chemical number	Substance	Test Code	CASRN[1]	Concentration Tested	Purity (%)	n. of animals	*In Vivo* EU[2,3]	*In Vivo* GHS[4,5]	*In Vivo* EPA[6,7]	MMAS	CM converted value MRD_{50} (mg/mL)
1	Shampoo 7	HZA	NA	100%	NA	6	R41	Category 1	Category I	37.8	1.18
2	Liquid Soap 1	HZB	NA	25%	NA	6	Not classified	No category	Category III	20.7	2.80
3	Shampoo 1	HZC	NA	25%	NA	6	Not classified	No category	Category III	36.0	1.72
4	Shampoo 5	HZD	NA	25%	NA	6	Not classified	No category	Category III	19.5	2.78
5	Gel Cleanser	HZE	NA	100%	NA	6	R41	No category	Category I	22	3.19
6	Baby Shampoo 2	HZF	NA	100%	NA	6	R41	Category 1	Category I	37.5	1.50
7	Shampoo 8	HZG	NA	25%	NA	6	Not classified	No category	Category III	17.8	2.80
8	Eye Makeup re.	HZH	NA	100%	NA	6	Not classified	No category	Category IV	2.3	20.0
9	Skin Cleaner	HZI	NA	100%	NA	6	R41	Category 1	Category I	41.0	1.09
10	Mild Shampoo	HZJ	NA	100%	NA	6	Not classified	No category	Category IV	8.2	6.38
11	Bubble bath	HZK	NA	100%	NA	6	R41	Category 1	Category I	39.7	0.97
12	Foam Bath	HZL	NA	100%	NA	6	R41	Category 1	Category I	37.8	1.09
13	Shampoo 3	HZM	NA	25%	NA	6	Not classified	No category	Category III	12.7	3.11
14	Shampoo 6	HZN	NA	25%	NA	6	Not classified	No category	Category III	18.0	2.56
15	Baby Shampoo 1	HZP	NA	100%	NA	6	Not classified	No category	Category III	11.7	2.45
16	Cleansing Gel	HZQ	NA	100%	NA	6	Not classified	No category	Category III	17.2	5.85
17	Facial Cleansing Foam	HZR	NA	25%	NA	6	R41	No category	Category I	39.0	5.60
18	Shower Gel	HZS	NA	100%	NA	6	R41	Category 1	Category I	41.4	1.13
19	Polishing Scrub	HZT	NA	100%	NA	6	Not classified	No category	Category IV	7.0	30.9
20	Hand Soap	HZU	NA	25%	NA	6	Not classified	No category	Category III	33.7	4.85
21	Shampoo 4	HZV	NA	25%	NA	6	Not classified	No category	Category III	25.2	2.34
22	Liquid Soap 2	HZW	NA	25%	NA	6	Not classified	2B	Category III	31.0	2.64
23	Shampoo 2	HZX	NA	100%	NA	6	R41	Category 1	Category I	40.0	1.20
24	Shampoo AntiD	HZY	NA	100%	NA	6	R41	Category 1	Category I	43.0	1.14
25	Facial Cleanser	HZZ	NA	100%	NA	6	Not classified	No category	Category IV	3.7	>168.9

[1]CASRN=Chemical Abstract Services Registry Number

[2]EU=European Union (EU [2001]).

[3]Risk phrase R41 = risk of serious damage to the eyes; R36 = irritating to the eyes; not classified.

[4]GHS=Globally Harmonized System (UN [2003])

[5]Eye Irritant Category 1 = irreversible effects on the eye/serious damage to the eye; Category 2A = reversible effects on the eye/irritating to the eyes; Category 2B = reversible effects on the eye/mildly irritating to the eyes; No category

[6]EPA=U.S. Environmental Protection Agency (EPA [1996]).

[7]Toxicity Category I for the Primary Eye Irritation Study = Corrosive, or corneal involvement or irritation not reversible within 21 days; Category II = Corneal involvement or irritation clearing in 8-21 days; Category III = Corneal involvement or irritation clearing in 1-7 daysl; Category IV: minimal effects clearing in less than 24 hr

[9]MMAS scores reported in Gettings et al. (1996)

[9]SCNM=Study Crtieria Not Met

[10]n.p.=not provided

[11] NA = not applicable

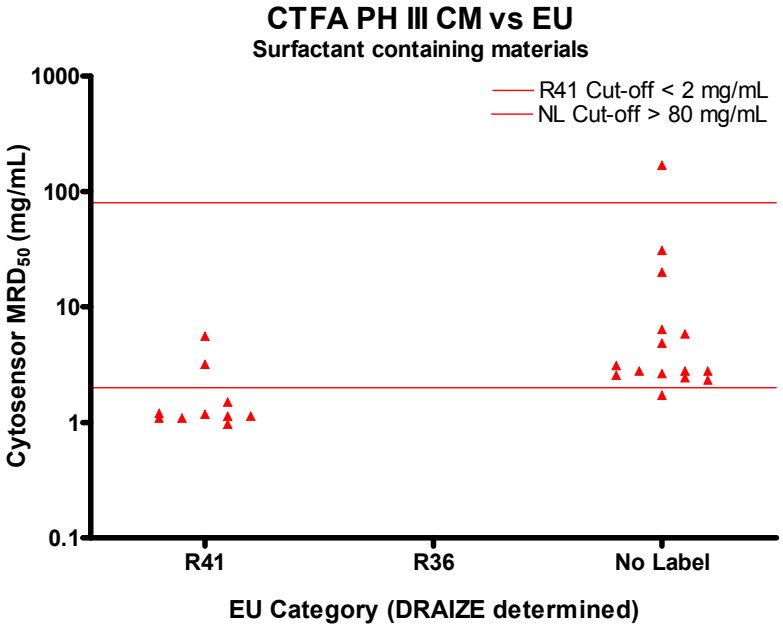

Figure 6.1.3.2.a Results of the CTFA Phase III study related to EU classification. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. MRD_{50} is reported in converted CM values.

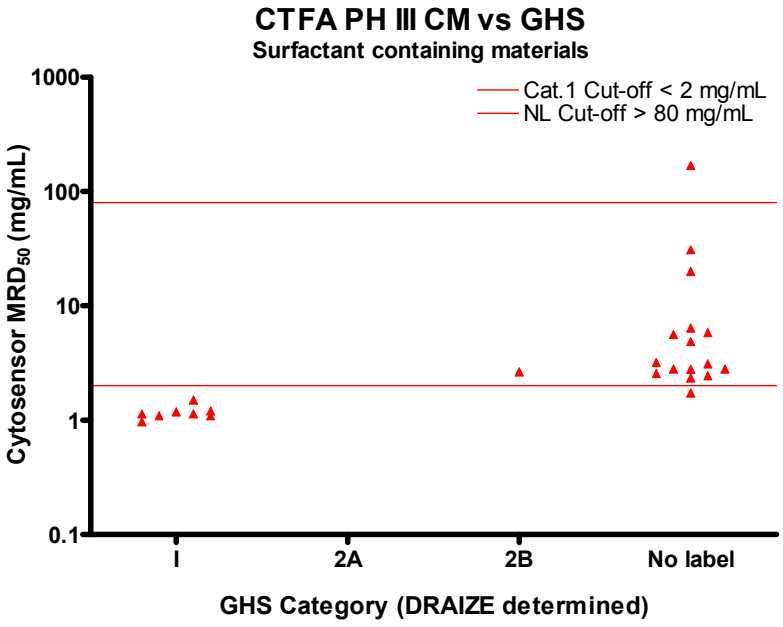

Figure 6.1.3.2.b Results of the CTFA Phase III study related to GHS classification. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. MRD_{50} is reported in converted CM values.

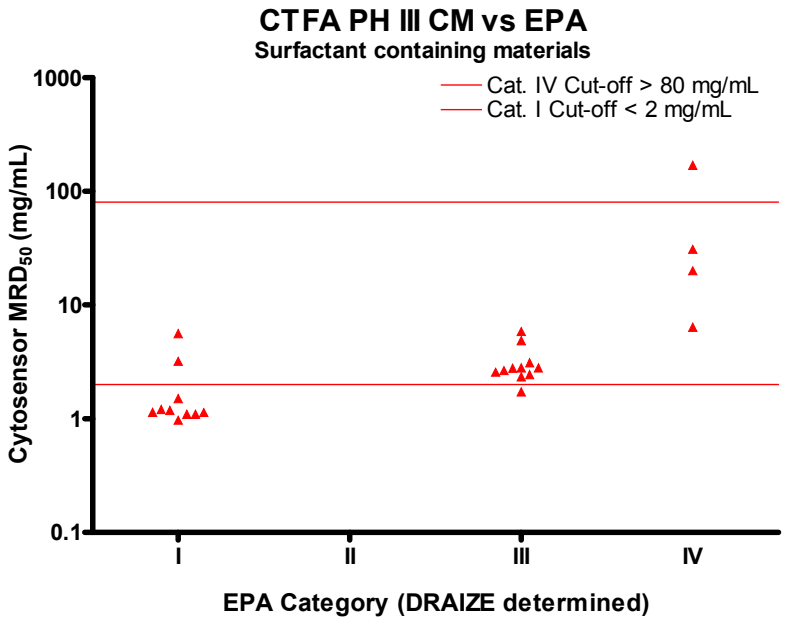

Figure 6.1.3.2.c Results of the CTFA Phase III study related to EPA classification. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. MRD_{50} is reported in converted CM values.

Figures 6.1.3.2.a, 6.1.3.2.b, and 6.1.3.2.c show cut-off values for MRD_{50} scores that have been empirically chosen to identify, where possible, the various hazard categories. In attempting to select cut-off values we first tried those that were chosen from the EC/HO study (see preceding section). Since these appeared adequate, we continued the analysis with these values for the sake of consistency. As with the EC/HO study, in the case of the GHS system and the EPA system which have 4 categories, the overlap of MRD_{50} response was so large that it was deemed impossible to differentiate between the two middle categories. This analysis was made even more difficult because of the distribution of the hazard classifications. There were no R36 materials and only 1 GHS 2A or 2B material. Hence only upper (to possibly identify non-irritants) and lower (to possibly identify severe irritants) cut-off values are shown.

It appears from the graphs that the SM does not have the ability to clearly separate the surfactant-containing materials used in the CTFA Phase III study into the Draize test defined EU, GHS or EPA categories. One exception is that severe irritants seem to be reasonably predicted when MRD_{50} scores of less than 2 are used. Using this lower cut-off value, there is a high positive predictive value for EU category R41 (89%; 8 of 9 materials), GHS category 1 (89%; 8 of 9 materials) and EPA category I (89%; 8 of 9 materials).

Even though the positive predictive value was high using a lower cut-off of MRD_{50} <2 mg/ml, the sensitivity was lower, with several severe chemicals being under predicted by the EU and EPA classification system. Over predictions of mild materials (EU Not Classified, GHS No Label, and EPA IV) were very frequent in this study.

A more detailed analysis of the performance of the selected cut-off values is given in Tables 6.1.3.2.b, 6.1.3.2.c, and 6.1.3.2.d for the EU, GHS and EPA classification systems, respectively.

All materials in the CTFA Phase III study were surfactant-containing materials, so no additional analysis of subsets of the test materials was necessary.

Table 6.1.3.2.b CTFA - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values shown in Figure 6.1.3.2.a are applied. N = 25 materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	8	2	0	10	80.0%	NA	20.0%
R36	0	0	0	0	NA	NA	NA
Not Classified	1	13	1	15	6.7%	93.3%	NA
Total	9	15	1	25	36.0%		
Predictivity	88.9%	0.0%	100.0%				
Category Underpredicted	NA	13.3%	0%				
Category Overpredicted	11.1%	86.7%	NA				

Table 6.1.3.2.c CTFA - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values shown in Figure 6.1.3.2.b are applied. N = 25 materials.

Draize Determined GHS Category	GHS Category Predicted By CM						
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	8	0		0	8	100.0%	NA	0%
2A	0	0		0	0	NA	NA	NA
2B	0	1		0	1	100.0%	0%	0%
No Label	1	14		1	16	6.3%	93.8%	NA
Total	9	15		1	25	44.0%		
Predictivity	88.9%	6.7%		100.0%				
Category Underpredicted	NA	0%		0%				
Category Overpredicted	11.1%	93.3%		NA				

Table 6.1.3.2.d CTFA - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values shown in Figure 6.1.3.2.c are applied. N = 25 materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	8	2		0	10	80.0%	NA	20.0%
II	0	0		0	0	NA	NA	NA
III	1	10		0	11	90.9%	9.1%	0%
IV	0	3		1	4	25.0%	75.0%	NA
Total	9	15		1	25	76.0%		
Predictivity	88.9%	66.7%		100.0%				
Category Underpredicted	NA	13.3%		0%				
Category Overpredicted	11.1%	20.0%		NA				

An analysis of the ability of the SM assay to separate severe irritants from “the rest”, and non-irritants from “the rest” is shown in Table 6.1.3.2.e. It can be seen that the SM assay performed quite well in some situations. For example when separating GHS severe irritants from “the rest”, there was a 96% concordance, 89% positive predictivity and 100% negative predictivity. Also when separating EPA non-irritants from “the rest” there was an 88% concordance, an 88% positive predictivity and a 100% negative predictivity; however, there were only 4 non-irritants in the sample so the number of non-irritants would have to be increased before any significant conclusions could be drawn.

Table 6.1.3.2.e CTFA - Concordance table for severe irritants versus the rest and non-irritants versus the rest. Cut-off values from Figures 6.1.3.2.a, 6.1.3.2.b, and 6.1.3.2.c were applied. N = 25 materials.

	Vs Rest													
	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
EU Severe irritants	88%	22/25	80%	8/10	93%	14/15	89%	8/9	88%	14/16	7%	1/15	20%	2/10
EU non-irritants	44%	11/25	100%	10/10	7%	1/15	42%	10/24	100%	1/1	93%	14/15	0%	0/10
GHS Severe irritants	96%	24/25	100%	8/8	94%	16/17	89%	8/9	100%	16/16	6%	1/17	0%	0/8
GHS non-irritants	40%	10/25	100%	9/9	6%	1/16	38%	9/24	100%	1/1	94%	15/16	0%	0/9
EPA Severe irritants	88%	22/25	80%	8/10	93%	14/15	89%	8/9	88%	14/16	7%	1/15	20%	2/10
EPA non-irritants	88%	22/25	100%	21/21	25%	1/4	88%	21/24	100%	1/1	75%	3/4	0%	0/21

6.1.3.3 Analysis of the COLIPA study (Brantom, Bruner et al. 1997)

The reference data for the COLIPA study (Annex J) came from three main sources; two for the neat chemicals and one for the formulations. The data for the chemicals came from the ECETOC data bank (ECETOC 1992) and the EU isolated cornea study (Gautheron, Giroux et al. 1994). All of these data are now available in a new edition of the ECETOC data bank (ECETOC 1998). Twenty-three chemicals were used in this study, and 20 of the 23 are identical to – and use the same Draize values – as a portion of the chemicals used in the EC/HO study. All of these data were from historical studies.

Thirty-two formulations were used in the COLIPA study (Annex D11), and the Draize scores for these formulations come from Draize tests conducted contemporaneous with this study. The formulations were newly prepared for this study, but most were based on formulations that had been tested in Phases I, II, and III of the CTFA evaluation program (Feder, Lordo et al. 1991; Gettings, Dipasquale et al. 1994; Gettings, Lordo et al. 1996). Thus, it is likely that for the formulations, the *in vitro* tests were challenged with exactly the same material as the *in vivo* test. The same cannot be said for the chemicals since historical data were used for them.

Two laboratories (Company # 4 and Company # 5) contributed CM data for this study (Annexes F13 & H29). Company # 4 found that 29 of the 55 materials were compatible with the CM, while Company # 5 found that only 26 of 55 materials were compatible. Because of this, only the mean MRD_{50} values from the 26 materials where both laboratories provided data were used in this analysis. See section 5.2.3.2 for more details. An overall summary of the COLIPA study including the chemical identities, animal scores and *in vitro* scores (averages from Company # 4 and Company # 5) are given in Table 6.1.3.3.a.

Table 6.1.3.3.a Summary of the COLIPA study which includes average values from Company # 4 and Company # 5 laboratories (Brantom, Bruner et al. 1997). Surfactant materials are highlighted. N = 10 non-surfactant materials and N = 19 surfactant materials.

In Vivo Eye Irritation Classifications - COLIPA study

COLIPA chemical number	Substance	CASRN[1]	Concentration Tested	Purity (%)	n. of animals	*In Vivo* EU[2,3]	*In Vivo* GHS[4,5]	*In Vivo* EPA[6,7]	ECETOC MMAS Score[8]	Average MRD_{50} (mg/mL)
5	Shampoo no. 1 - normal		100%		3	R41	Category 1	Category I	33.3	0.735
6	Eye make-up remover		100%		3	Not classified	No Category	Category IV	0.7	93.5
10	Imidazole		100%		3	R41	Category 1	SCNM	59.3	22.4
11	Polyethylene glycol 400		100%		6	Not classified	No Category	Category IV	0.0	306.4
12	Propylene glycol		100%		3	Not classified	No Category	Category IV	1.3	242.0
13	Triton X-100		1%		3	Not classified	No Category	Category III	1.7	19.0
14	Glycerol		100%		6	Not classified	No Category	Category IV	1.7	211.8
15	Tween 20		100%		4	Not classified	No Category	Category III	4.0	6.50
17	Sodium lauryl sulphate		3%		6	Not classified	No Category	Category III	16.0	3.00
18	Sodium hydroxide		1%		4	R36	Category 2B	Category III	25.8	11.34
19	Isopropanol		100%		4	Not classified	Category 2A	Category III	30.5	88.5
20	Triton X-100 [2]		5%		6	Not classified	Category 2A	Category III	32.3	3.54
21	Benzalkonium chloride [2]		1%		4	R41	Category 1	Category I	34.3	4.22
22	Methyl ethyl ketone		100%		6	R36	Category 2A	Category III	50.0	54.2*
23	Sodium lauryl sulphate		15%		6	R41	Category 1	Category I	59.2	0.513
24	Sodium lauryl sulphate		30%		6	R36	Category 2A	Category II	60.5	0.312*
25	Triton X-100		10%		6	R41	Category 1	Category II	59.0	1.85
26	Benzalkonium chloride		5%		4	R41	Category 1	Category I	83.8	1.095
27	Benzalkonium chloride		10%		3	R41	Category 1	Category I	108.0	0.314
28	Pump deodorant / antiperspirant		100%		3	Not classified	No Category	Category III	14.7	33.54
34	Gel cleanser		100%		3	Not classified	No Category	Category III	15.7	5.58
36	Shampoo - baby		100%		3	R41	Category 1	Category I	36.0	2.33
38	Hair styling lotion		100%		3	R36	Category 2A	Category III	19.3	228.4
39	Liquid soap no.1		100%		3	R41	Category 1	Category I	37.0	0.78
43	Mouthwash		100%		3	Not classified	No Category	Category IV	0.7	42.35
49	Skin cleanser		100%		3	R41	Category 1	Category I	34.3	0.70
52	Cetylpyridinium bromide		6%		4	R41	Category I	Category I	85.8	1.36*
54	Sodium hydroxide		10%		3	R41	Category 1	Category I	108.0	2.47
55	Trichloroacetic acid		30%		3	R41	Category 1	Category I	106.0	1.18

* - MA value only, CellTox AB designated unsuitable for testing; therefore, these values were not included in the analysis

[1]CASRN=Chemical Abstract Services Registry Number

[2]EU=European Union (EU [2001]).

[3]Risk phrase R41 = risk of serious damage to the eyes; R36 = irritating to the eyes; not classified.

[4]GHS=Globally Harmonized System (UN [2003])

[5]Eye Irritant Category 1 = irreversible effects on the eye/serious damage to the eye; Category 2A = reversible effects on the eye/irritating to the eyes; Category 2B = reversible effects on the eye/mildly irritating to the eyes; No category

[6]EPA=U.S. Environmental Protection Agency (EPA [1996]).

[7]Toxicity Category I for the Primary Eye Irritation Study = Corrosive, or corneal involvement or irritation not reversible within 21 days; Category II = Corneal involvement or irritation clearing in 8-21 days; Category III = Corneal involvement or irritation clearing in 1-7 daysl; Category IV: minimal effects clearing in less than 24 hr

[8]MMAS scores reported in Brantom et al. (1997)

[9]SCNM=Study Crtieria Not Met

[10]n.p.=not provided

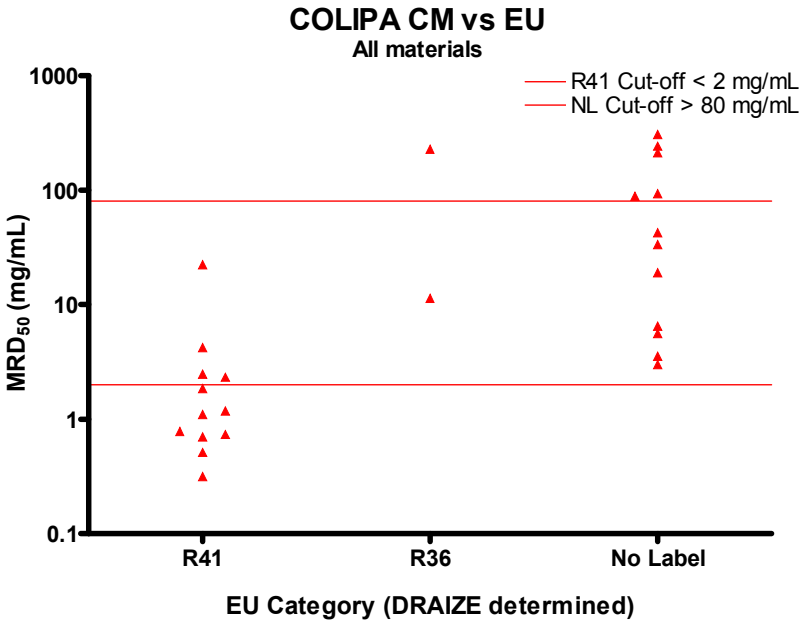

Figure 6.1.3.3.a Results of the COLIPA study related to EU classification. Data points indicate the mean MRD_{50} for both laboratories. Only data from test materials which were tested in both laboratories are shown here. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

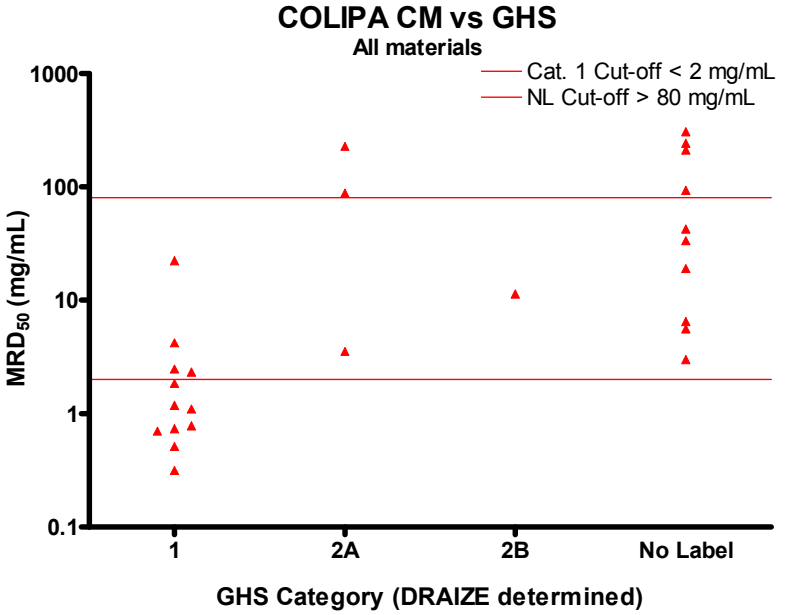

Figure 6.1.3.3.b Results of the COLIPA study related to GHS classification. Data points indicate the mean MRD_{50} for both laboratories. Only data from test materials which were tested in both laboratories are shown here. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

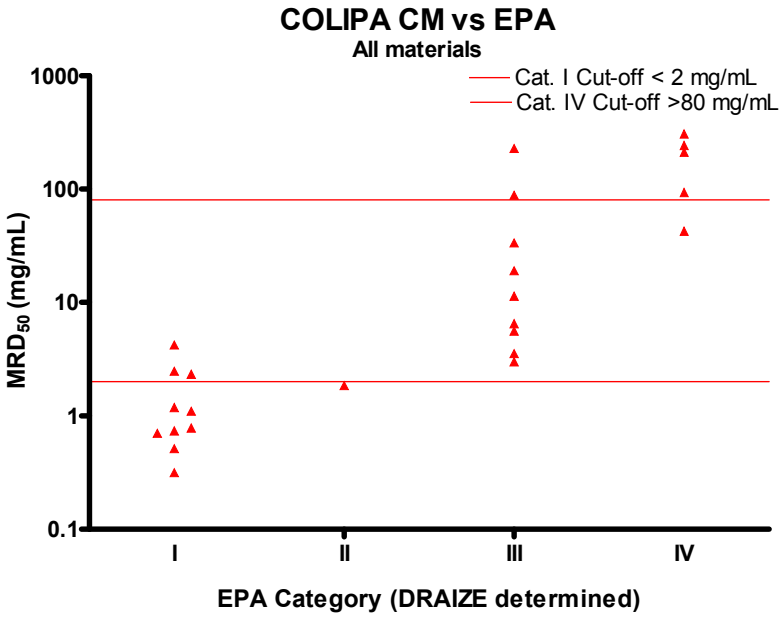

Figure 6.1.3.3.c Results of the COLIPA study related to EPA classification. Data points indicate the mean MRD_{50} for both laboratories. Only data from test materials which were tested in both laboratories are shown here. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

Figures 6.1.3.3.a, 6.1.3.3.b, and 6.1.3.3.c show cut-off values for MRD_{50} scores that have been empirically chosen to identify, where possible, the various hazard categories. In attempting to select cut-off values we first tried those that were chosen from the EC/HO study and CTFA Phase III studies (see preceding sections). Since these appeared adequate, we continued the analysis with these values for the sake of consistency. As with the EC/HO and CTFA Phase III studies, in the case of the GHS system and the EPA system which have 4 categories, the overlap of MRD_{50} response was so large that it was deemed impossible to differentiate between the two middle categories. This analysis was made even more difficult because of the distribution of the hazard classifications. There were only 2 R36 materials and only 4 GHS 2A or 2B materials. Hence only upper (to possibly identify non-irritants) and lower (to possibly identify severe irritants) cut-off values are shown.

It appears from the graphs that the CM does not have the ability to clearly separate the chemicals or surfactant-containing materials used in the COLIPA study into the Draize test defined EU, GHS or EPA categories. One exception is that severe irritants seem to be reasonably predicted when MRD_{50} scores of less than 2 are used. Using this lower cut-off value, there is a high positive predictive value for EU category R41 (100%; 8 of 8 materials), GHS category 1 (100%; 8 of 8 materials) and EPA category I (88%; 7 of 8 materials).

Even though the positive predictive value was high using a lower cut-off of MRD_{50} <2 mg/ml, the sensitivity was lower, with several severe chemicals being under predicted

by at least one hazard category by the EU, GHS, and EPA classification system. Over predictions of mild materials (EU Not Classified, GHS No Label, and EPA IV), especially by the EU and GHS system, occurred very frequent in this study.

Table 6.1.3.3.b COLIPA - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values shown in Figure 6.1.3.3.a are applied. N = 26 materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity over predicted	Toxicity under predicted
R41	8	4	0	12	66.7%	NA	33.3%
R36	0	1	1	2	50.0%	0%	50.0%
Not Classified	0	7	5	12	41.7%	58.3%	NA
Total	8	12	6	26	53.8%		
Predictivity	100.0%	8.3%	83.3%				
Category Underpredicted	NA	33.3%	16.7%				
Category Overpredicted	0.0%	58.3%	NA				

Table 6.1.3.3.c COLIPA - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values shown in Figure 6.1.3.3.b are applied. N = 26 materials.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity over predicted	Toxicity under predicted
1	8	4		0	12	66.7%	NA	33.3%
2A	0	1		2	3	33.3%	0%	66.7%
2B	0	1		0	1	100.0%	0%	0%
No Label	0	6		4	10	40.0%	60.0%	NA
Total	8	12		6	26	53.8%		
Predictivity	100.0%	16.7%		66.7%				
Category Under predicted	NA	33.3%		33.3%				
Category Overpredicted	0%	50.0%		NA				

Table 6.1.3.3.d COLIPA - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values shown in Figure 6.1.3.3.c are applied. N = 25 materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	7	3		0	10	70.0%	NA	30.0%
II	1	0		0	1	0.0%	100.0%	0%
III	0	7		2	9	77.8%	0.0%	22.2%
IV	0	1		4	5	80.0%	20.0%	NA
Total	8	11		6	25	72.0%		
Predictivity	87.5%	63.6%		66.7%				
Category Underpredicted	NA	27.3%		33.3%				
Category Overpredicted	12.5%	9.1%		NA				

An analysis of the ability of the CM assay to separate severe irritants from “the rest”, and non-irritants from “the rest” is shown in Table 6.1.3.3.e. In general, the CM assay did not perform as well with this set of materials (a combination of surfactants and non-surfactant materials) as it did in the COLIPA study. In no case was there both a high positive predictivity and a high negative predictivity.

Table 6.1.3.3.e COLIPA - Concordance table for severe irritants versus the rest and non-irritants versus the rest. Cut-off values from Figures 6.1.3.3.a, 6.1.3.3.b, and 6.1.3.3.c were applied. N = 26 materials (EU & GHS) and N = 25 materials (EPA)

	Vs Rest													
	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
EU Severe irritants	85%	22/26	67%	8/12	100%	14/14	100%	8/8	78%	14/18	0%	0/14	33%	4/12
EU non-irritants	69%	18/26	93%	13/14	42%	5/12	65%	13/20	83%	5/6	58%	7/12	7%	1/14
GHS Severe irritants	85%	22/26	67%	8/12	100%	14/14	100%	8/8	78%	14/18	0%	0/14	33%	4/12
GHS non-irritants	69%	18/26	88%	14/16	40%	4/10	70%	14/20	67%	4/6	60%	6/10	13%	2/16
EPA Severe irritants	84%	21/25	70%	7/10	93%	14/15	88%	7/8	82%	14/17	7%	1/15	30%	3/10
EPA non-irritants	88%	22/25	90%	18/20	80%	4/5	95%	18/19	67%	4/6	20%	1/5	10%	2/20

Analysis of surfactants and surfactant-containing formulations

Since in recent years the applicability domain of the SM or CM assay has become more narrowly defined to be limited to test materials that are completely water soluble, and since much of the SM and CM testing over that same time period has been surfactants and surfactant-containing products, it was decided to investigate only the surfactants and surfactant-containing materials from the COLIPA test material set. There were 17 surfactants and surfactant-containing materials which had data from two labs. Figures 6.1.3.3.d, 6.1.3.3.e, and 6.1.3.3.f show the results of that analysis relative to EU, GHS, and EPA classifications, respectively.

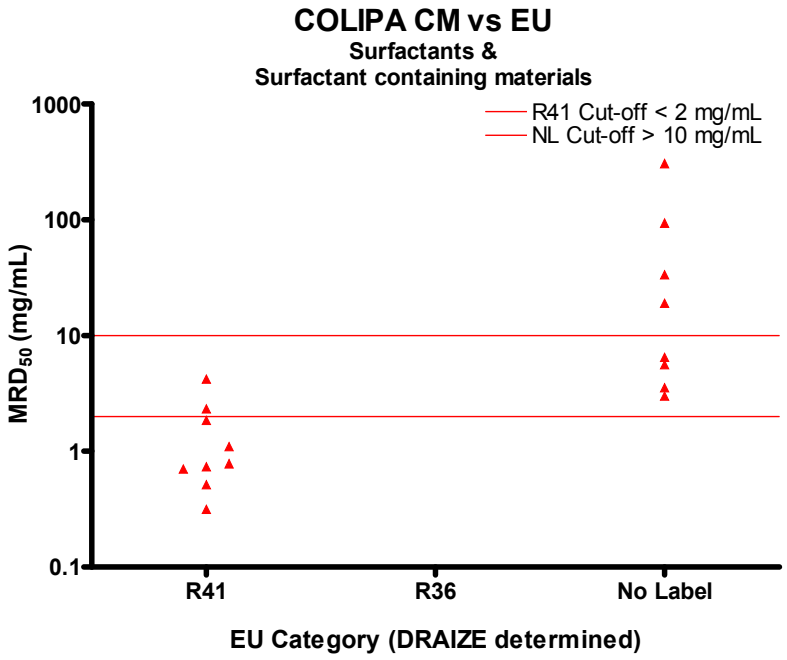

Figure 6.1.3.3.d Surfactant results of the COLIPA study related to EU classification. Data points indicate the mean MRD_{50} for both laboratories. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

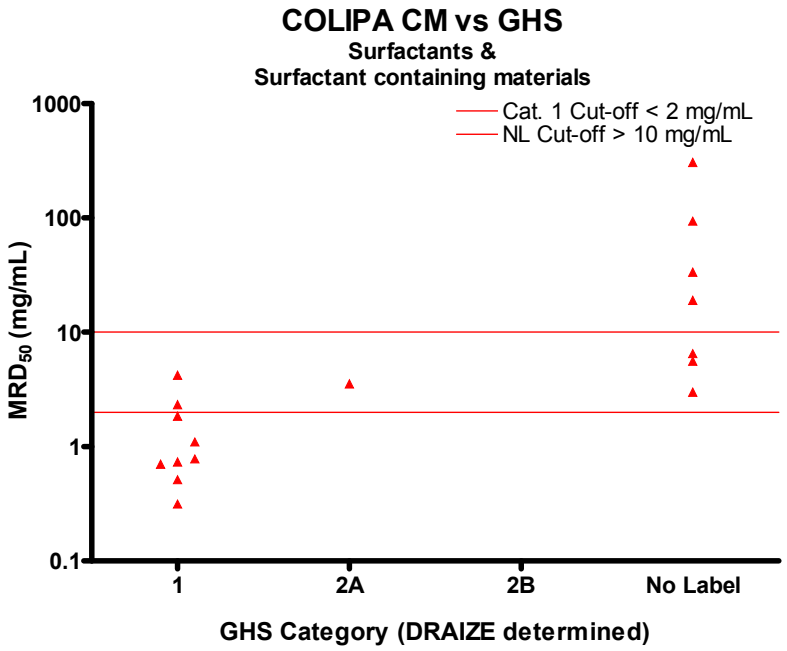

Figure 6.1.3.3.e Surfactant results of the COLIPA study related to GHS classification. Data points indicate the mean MRD_{50} for both laboratories. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

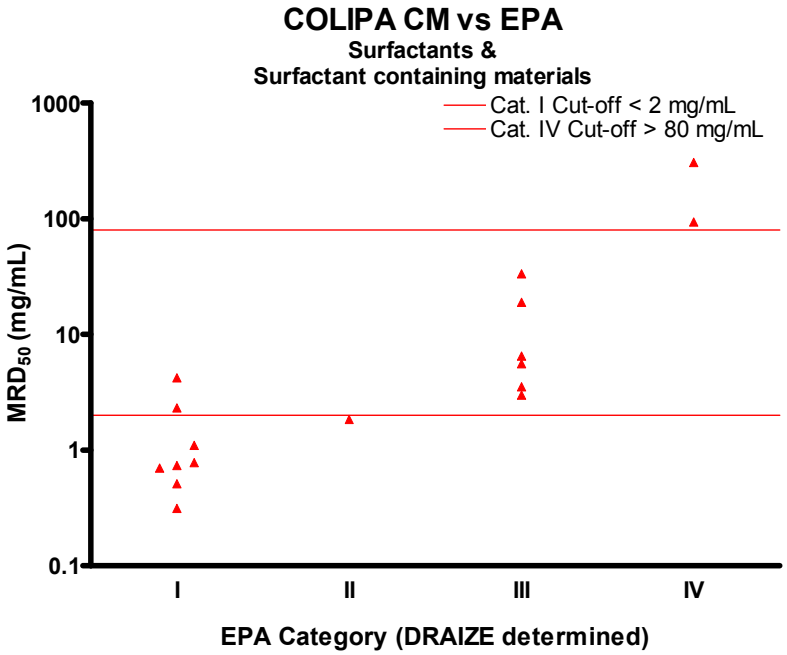

Figure 6.1.3.3.f Surfactant results of the COLIPA study related to EPA classification. Data points indicate the mean MRD_{50} for both laboratories. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude.

For these three data sets it appeared that a lower cut-off value (>10 mg/mL) than the previously used MRD_{50} > 80 mg/ml might be appropriate to identify the EU not classified and GHS no label from the more irritating. The cut-off of <2 mg/ml was retained for identifying R41, GHS 1 or EPA I materials. However, as seen in most of the previous analyses, there were very few materials in the EU R36, GHS 2A or 2B, or EPA II categories. This makes it difficult to determine exactly where the cut-off between these intermediate irritating categories and the mild categories lies. Additionally the EPA classification had only two Category IV materials, again making a decision for a cut-off almost impossible. To make the analysis even more difficult there were no R36 materials and only 1 GHS 2A or 2B materials. Hence only upper (to possibly identify non-irritants) and lower (to possibly identify severe irritants) cut-off values are shown on the three scatter plots.

It appears from the graphs that the CM does not have the ability to clearly separate the surfactants or surfactant-containing materials used in the COLIPA study into the Draize test defined EU, GHS or EPA categories. One exception is that severe irritants seem to be reasonably predicted when MRD_{50} scores of less than 2 are used. Using this lower cut-off value, there is a high positive predictive value for EU category R41 (100%; 7 of 7 materials), GHS category 1 (100%; 7 of 7 materials) and EPA category I (86%; 6 of 7 materials).

Even though the positive predictive value was high using a lower cut-off of MRD_{50} <2 mg/ml, the sensitivity was lower, with several severe chemicals being under predicted by at least one hazard category by the EU, GHS, and EPA classification system. Over predictions of mild materials (EU Not Classified, GHS No Label, and EPA IV), were not as great as found in the previous studies.

Table 6.1.3.3.f COLIPA Surfactant and surfactant containing materials - Contingency table for depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values shown in Figure 6.1.3.3.d are applied. N = 17 surfactant materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	7	2	0	9	77.8%	NA	22.2%
R36	0	0	0	0	NA	NA	NA
Not Classified	0	4	4	8	50.0%	50.0%	NA
Total	7	6	4	17	64.7%		
Predictivity	100.0%	0.0%	100.0%				
Category Underpredicted	NA	33.3%	0%				
Category Overpredicted	0%	66.7%	NA				

Table 6.1.3.3.g COLIPA Surfactant and surfactant containing materials - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values shown in Figure 6.1.3.3.e are applied. N = 17 surfactant materials.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	7	2		0	9	77.8%	NA	22.2%
2A	0	1		0	1	100.0%	0%	0%
2B	0	0		0	0	NA	NA	NA
No Label	0	3		4	7	57.1%	42.9%	NA
Total	7	6		4	17	70.6%		
Predictivity	100.0%	16.7%		100.0%				
Category Underpredicted	NA	33.3%		0%				
Category Overpredicted	0%	50.0%		NA				

Table 6.1.3.3.h COLIPA Surfactant and surfactant containing materials - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values shown in Figure 6.1.3.3.f are applied. N = 17 surfactant materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	6	2		0	8	75.0%	NA	25.0%
II	1	0		0	1	0.0%	100.0%	0%
III	0	6		0	6	100.0%	0%	0%
IV	0	0		2	2	100.0%	0%	NA
Total	7	8		2	17	82.4%		
Predictivity	85.7%	75.0%		100.0%				
Category Underpredicted	NA	25.0%		0%				
Category Overpredicted	14.3%	NA		NA				

An analysis of the ability of the CM assay to test surfactants and surfactant-based materials and separate severe irritants from “the rest”, and non-irritants from “the rest” is shown in Table 6.1.3.3.h. Although the assay seemed to perform well in some situations, e.g. when separating EPA non-irritants from “the rest” both the positive and negative predictivities were 100%, the fact that only 17 materials were included in the analysis prevents any strong conclusions from being drawn.

Table 6.1.3.3.i COLIPA Surfactant and surfactant containing materials - Concordance table for surfactant and surfactant containing matierlas for severe irritants versus the rest and non-irritants versus the rest. Cut-off values from Figures 6.1.3.3.d, 6.1.3.3.e, and 6.1.3.3.f were applied. N = 17 surfactant materials.

	Vs Rest													
	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Posivite Rate		False Negative Rate	
	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
EU Severe irritants	88%	15/17	78%	7/9	100%	8/8	100%	7/7	80%	8/10	0%	0/8	22%	2/9
EU non-irritants	76%	13/17	100%	9/9	50%	4/8	69%	9/13	100%	4/4	50%	4/8	0%	0/9
GHS Severe irritants	88%	15/17	78%	7/9	100%	8/8	100%	7/7	80%	8/10	0%	0/8	22%	2/9
GHS non-irritants	82%	14/17	100%	10/10	57%	4/7	77%	10/13	100%	4/4	43%	3/7	0%	0/10
EPA Severe irritants	82%	14/17	75%	6/8	89%	8/9	86%	6/7	80%	8/10	11%	1/9	25%	2/8
EPA non-irritants	100%	17/17	100%	15/15	100%	2/2	100%	15/15	100%	2/2	0%	0/2	0%	0/15

6.1.3.4 Analysis of the combined CTFA Phase III, EC/HO, and COLIPA data

Since each of the previously described studies had a relatively small number of data points – and the prediction models were being developed *post hoc* - we felt it would be more accurate to combine information from all three studies so that a more comprehensive prediction model(s) could be developed. At the same time we analyzed the data according to their distribution into the more defined applicability domains of surfactant materials, non-surfactant materials, surfactant-containing products (or mixtures) and non surfactant-containing products. Even so, the number of data points was still rather low, e.g. only 53 materials (not necessarily unique) were available when just the surfactants and surfactant-containing materials from the CTFA, EC/HO and COLIPA studies were considered. We justified combining data from these three studies with the knowledge that identical CM protocols were used for the EC/HO and COLIPA studies, and that the CTFA data could be converted from SM data to CM data.

The positive justifications for combining data notwithstanding, there are at least two caveats that should be considered in reviewing the analysis. First, the animal tests used to categorize the test materials differed slightly among the studies; topical ocular anesthesia was used for the CTFA studies but was not used for the EC/HO study or the COLIPA study. Secondly – and perhaps more importantly – the materials used in the three studies have some amount of overlap. However, the degree of overlap involved is not clear. Specifically, the COLIPA study attempted to use six of the surfactant formulations that were used in the CTFA study, but the test materials had to be reformulated for the COLIPA study which took place several years later. Thus it is likely that the formulations were very similar, but probably slightly different. Also seven surfactant/concentration combinations were duplicated between the EC/HO study and the COLIPA study, but there is no evidence that the materials were truly identical since they may have been procured from different sources and at different purities.

Therefore, we have chosen to graph all of the data points from the combined studies and use them to determine the prediction models and the performance statistics; but in each of these cases the number of truly unique materials is listed. In addition we also list the number of unique chemicals since some of the chemicals were tested at several different concentrations.

Table 6.1.3.4.a Surfactant Chemicals from the EC/HO, COLIPA, and CTFA studies. There are 21 total materials; 13 unique chemical:concentration combinations, and 6 totally unique surfactants. N = 21 surfactant chemicals.

Substance	Concentration Tested	Purity (%)	no. of animals	In Vivo EU[2,3]	In Vivo GHS[4,5]	In Vivo EPA[6,7]	ECETOC MMAS Score[8]	MRD_{50} (mg/mL) [11]	Protocol
Benzalkonium chloride	5		4	R41	Category 1	Category I	83.75	1.1	COLIPA
Benzalkonium chloride	10		3	R41	Category 1	Category I	108	0.31	COLIPA
Benzalkonium chloride (1%) (1x)	1	98	4	R41	Category 2A	Category I	34.4	4.53	EC/HO
Benzalkonium chloride (10%)	10	98	3	R41	Category 1	Category I	108	0.39	EC/HO
Benzalkonium chloride (5%)	5	98	4	R41	Category 1	Category I	83.8	1.13	EC/HO
Benzalkonium chloride [2]	1		4	R41	Category 1	Category I	34.25	4.22	COLIPA
Cetylpyridinium bromide (0.1%)	0.1	99	6	Not classified	No Category	Category III	2.7	84.65	EC/HO
Cetylpyridinium bromide (10%)	10	99	6	R41	Category 1	Category I	89.7	1.26	EC/HO
Cetylpyridinium bromide (6%)	6	99	4	R41	Category 1	SCNM[9]	85.8	0.87	EC/HO
Polyethylene glycol 400	100		6	Not classified	No Category	Category IV	0	306.36	COLIPA
Sodium lauryl sulfate (15%)	15	98	6	R41	Category 1	Category I	59.2	0.62	EC/HO
Sodium lauryl sulfate (3%)	3	98	6	Not classified	No Category	Category III	16	3.28	EC/HO
Sodium lauryl sulphate	3		6	Not classified	No Category	Category III	16	3	COLIPA
Sodium lauryl sulphate	15		6	R41	Category 1	Category I	59.17	0.51	COLIPA
Triton X-100	1		3	Not classified	No Category	Category III	1.67	18.98	COLIPA
Triton X-100	10		6	R41	Category 1	Category II	59	1.85	COLIPA
Triton X-100 (10%)	10	98	6	R41	Category 1	Category II	68.7	1.82	EC/HO
Triton X-100 (5%) (1x)	5	98	6	R36	Category 2A	Category III	32.3	3.23	EC/HO
Triton X-100 [2]	5		6	Not classified	Category 2A	Category III	32.33	3.54	COLIPA
Tween 20	100	98	4	Not classified	No Category	Category III	4	3.27	EC/HO
Tween 20	100		4	Not classified	No Category	Category III	4	6.5	COLIPA

Table 6.1.3.4.b Surfactant Formulations from the EC/HO, COLIPA, and CTFA studies. There are 32 test materials; most likely all are unique. N = 32 surfactant formulations.

Substance	Concentration Tested	Purity (%)	no. of animals	In Vivo EU[2,3]	In Vivo GHS[4,5]	In Vivo EPA[6,7]	ECETOC MMAS Score[8]	MRD_{50} (mg/mL) [11]	Protocol
Baby Shampoo 1	100	NA	6	Not classified	No category	Category III	11.7	2.45	CTFA
Baby Shampoo 2	100	NA	6	R41	Category 1	Category I	37.5	1.5	CTFA
Bubble bath	100	NA	6	R41	Category 1	Category I	39.7	0.97	CTFA
Cleansing Gel	100	NA	6	Not classified	No category	Category III	17.2	5.85	CTFA
Eye Makeup re.	100	NA	6	Not classified	No category	Category IV	2.3	20.03	CTFA
Eye make-up remover	100		3	Not classified	No Category	Category IV	0.67	93.54	COLIPA
Facial Cleanser	100	NA	6	Not classified	No category	Category IV	3.7	>168.9	CTFA
Facial Cleansing Foam	25	NA	6	R41	Category 2A	Category I	39	5.6	CTFA
Foam Bath	100	NA	6	R41	Category 1	Category I	37.8	1.09	CTFA
Gel cleanser	100		3	Not classified	No Category	Category III	15.67	5.58	COLIPA
Gel Cleanser	100	NA	6	R41	No category	Category I	22	3.19	CTFA
Hand Soap	25	NA	6	Not classified	No category	Category III	33.7	4.85	CTFA
Liquid Soap 1	25	NA	6	Not classified	No category	Category III	20.7	2.8	CTFA
Liquid Soap 2	25	NA	6	Not classified	Category 2B	Category III	31	2.64	CTFA
Liquid soap no.1	100		3	R41	Category 1	Category I	37	0.78	COLIPA
Mild Shampoo	100	NA	6	Not classified	No category	Category IV	8.2	6.38	CTFA
Polishing Scrub	100	NA	6	Not classified	No category	Category IV	7	30.85	CTFA
Pump deodorant / antiperspirant	100		3	Not classified	No Category	Category III	14.67	33.54	COLIPA
Shampoo – baby	100		3	R41	Category 1	Category I	36	2.33	COLIPA
Shampoo 1	25	NA	6	Not classified	No category	Category III	36	1.72	CTFA
Shampoo 2	100	NA	6	R41	Category 1	Category I	40	1.2	CTFA
Shampoo 3	25	NA	6	Not classified	No category	Category III	12.7	3.11	CTFA
Shampoo 4	25	NA	6	Not classified	No category	Category III	25.2	2.34	CTFA
Shampoo 5	25	NA	6	Not classified	No category	Category III	19.5	2.78	CTFA
Shampoo 6	25	NA	6	Not classified	No category	Category III	18	2.56	CTFA
Shampoo 7	100	NA	6	R41	Category 1	Category I	37.8	1.18	CTFA
Shampoo 8	25	NA	6	Not classified	No category	Category III	17.8	2.8	CTFA
Shampoo AntiD	100	NA	6	R41	Category 1	Category I	43	1.14	CTFA
Shampoo no. 1 – normal	100		3	R41	Category 1	Category I	33.33	0.74	COLIPA
Shower Gel	100	NA	6	R41	Category 1	Category I	41.4	1.13	CTFA
Skin Cleaner	100	NA	6	R41	Category 1	Category I	41	1.09	CTFA
Skin cleanser	100		3	R41	Category 1	Category I	34.33	0.7	COLIPA

Table 6.1.3.4.c Non-surfactant Chemicals from the EC/HO, COLIPA, and CTFA studies. There are 27 total materials; 21 unique chemical:concentration combinations, and 19 totally unique chemicals. N = 27 non-surfactant chemicals.

Substance	Concentration Tested (%)	Purity (%)	no. of animals	In Vivo EU[2,3]	In Vivo GHS[4,5]	In Vivo EPA[6,7]	ECETOC MMAS Score[8]	MRD_{50} (mg/mL) [11]	Protocol
2,5-Dimethylhexanediol	100	99.5	3	R41	Category 1	Category I	28.3	98.67	EC/HO
Acetone	100	99	4	R36	Category 2A	Category II	65.8	148.82	EC/HO
Ammonium nitrate	100	99.999	3	R36	Category 2A	Category III	18.3	71.27	EC/HO
Cyclohexanol	100	97	4	R41	Category 1	Category I	79.8	8.03	EC/HO
Ethanol	100	n.p.	3	Not classified	Category 2A	Category III	24	111.99	EC/HO
Gammabutyrolactone	100	>99	3	R36	Category 2A	Category II	43	93.79	EC/HO
Glycerol	100	>99.5	6	Not classified	No category	Category IV	1.7	129.88	EC/HO
Glycerol	100		6	Not classified	No Category	Category IV	1.7	211.77	COLIPA
Imidazole	100	99	3	R41	Category 1	SCNM	59.3	23.69	EC/HO
Imidazole	100		3	R41	Category 1	SCNM	59.3	22.43	COLIPA
Isobutanol	100	99.9	4	R36	Category 2A	Category II	60.3	27.91	EC/HO
Isopropanol	100	99.9	4	Not classified	Category 2A	Category III	30.5	101.26	EC/HO
Isopropanol	100		4	Not classified	Category 2A	Category III	30.5	88.55	COLIPA
Methyl acetate	100	98	4	R36	Category 2A	Category II	39.5	94.68	EC/HO
Methyl cyanoacetate	100	99	3	R36	Category 2A	Category II	27.7	21.54	EC/HO
Methyl ethyl ketone	100	99	4	R36	Category 2A	Category III	50	58.08	EC/HO
Promethazine HCl	100	98	3	R41	Category 1	SCNM	71.7	1.27	EC/HO
Propylene glycol	100		3	Not classified	No Category	Category IV	1.3	241.97	COLIPA
Pyridine	100	>99.9	3	R41	Category 1	SCNM	48	19.73	EC/HO
Sodium hydroxide	1		4	R36	Category 2B	Category III	25.8	11.34	COLIPA
Sodium hydroxide (1%)	1	reagent grade	4	R36	Category 2B	Category III	25.8	27.04	EC/HO
Sodium hydroxide	10		3	R41	Category 1	Category I	108	2.47	COLIPA
Sodium hydroxide (10%)	10	reagent grade	1	R41	Category 1	Category I	108	2.26	EC/HO
Sodium perborate, 4H2O	100	98.6	6	R41	Category 1	Category I	30.5	1.69	EC/HO
Trichloroacetic acid	30		3	R41	Category 1	Category I	106	1.18	COLIPA
Trichloroacetic acid (30%)	30	reagent grade	1	R41	Category 1	Category I	106	1.79	EC/HO
Trichloroacetic acid (3%)	3	reagent grade	6	Not classified	No category	Category III	6.7	15.03	EC/HO

Table 6.1.3.4.d Non-surfactant Formulations from the EC/HO, COLIPA, and CTFA studies. N = 2 non-surfactant formulations.

Substance	Concentration Tested	Purity (%)	no. of animals	In Vivo EU[2,3]	In Vivo GHS[4,5]	In Vivo EPA[6,7]	ECETOC MMAS Score[8]	MRD_{50} (mg/mL) [11]	Protocol
Hair styling lotion	100		3	R36	Category 2A	Category III	19.33	228.41	COLIPA
Mouthwash	100		3	Not classified	No Category	Category IV	0.67	42.35	COLIPA

We attempted to investigate the effect of topical ocular anesthesia on the Draize test results by comparing the 6 materials that were formulated to be identical between the CTFA and the COLIPA studies. This comparison is shown in Table 6.1.3.4.e. No clear cut pattern could be seen; for two materials the hazard categories were lower in the CTFA study, for one material the categories were higher, and for three materials the categories were the same. This discrepancy, combined with the fact that it was impossible to determine whether or not the formulations were identical since they were formulated several years apart, made it clear that it would be very difficult to determine the effect of topical anesthesia on the Draize score. This made it even more difficult to determine whether the six materials should be treated as identical; however we felt that it was not unreasonable to continue the evaluation of predictive capacity by treating the twelve data sets (six from the CTFA study, six from the COLIPA study) as if they had come from unique formulations.

Table 6.1.3.4.e Comparison of animal data for the CTFA and COLIPA studies. Liquid Soap 1 and Shampoo #1 used in the COLIPA study were formulated at 25% the strength used in the CTFA study, but were tested without dilution as was done in the CTFA study. Thus the actual tested concentrations of the formulations were the same.

CTFA								COLIPA							
Substance Designation	Conc. Tested	n. of animals	EU	GHS	EPA	MMAS	CM converted value MRD_{50} (mg/mL)	Substance Designation	Conc. Tested	n. of animals	EU	GHS	EPA	MMAS	Average MRD_{50} (mg/mL) (avg MA & CT)
Liquid Soap 1	25%	6	Not classified	No category	Category III	20.7	2.80	Liquid Soap 1	100%	3	R41	Category 1	Category I	37.0	0.78
Gel Cleanser	100%	6	R41	No category	Category I	22	3.19	Gel Cleanser	100%	3	Not classified	No Category	Category III	15.7	5.58
Baby Shampoo 2	100%	6	R41	Category 1	Category I	37.5	1.50	Shampoo - Baby	100%**	3	R41	Category 1	Category I	36.0	2.33
Shampoo 1	25%	6	Not classified	No category	Category III	36	1.72	Shampoo #1	100%**	3	R41	Category 1	Category I	33.3	0.735
Eye Makeup Remover	100%	6	Not classified	No category	Category IV	2.3	20.0	Eye Makeup Remover	100%	3	Not classified	No Category	Category IV	0.67	93.5
Skin Cleaner	100%	6	R41	Category 1	Category I	41	1.09	Skin Cleaner	100%	3	R41	Category 1	Category I	34.3	0.696

** - COLIPA formulation slightly different than the CTFA formulation

We next constructed scatter plots of the combined data for surfactants, surfactant containing formulations and non-surfactant materials plotted against the EU, GHS and EPA hazard categories in order to determine the appropriate cut-offs for the prediction model(s). Because only two materials were available in the non-surfactant formulations category, data for these materials were not plotted as a separate group, but were combined with the non-surfactant materials. Non-surfactant materials were also plotted by themselves. Finally the surfactant chemicals and the surfactant formulations were plotted together to determine if a single set of cut-off MRD_{50} values could be used for both.

6.1.3.4.1 Analysis of the combined CTFA, EC/HO, and COLIPA data with EU category

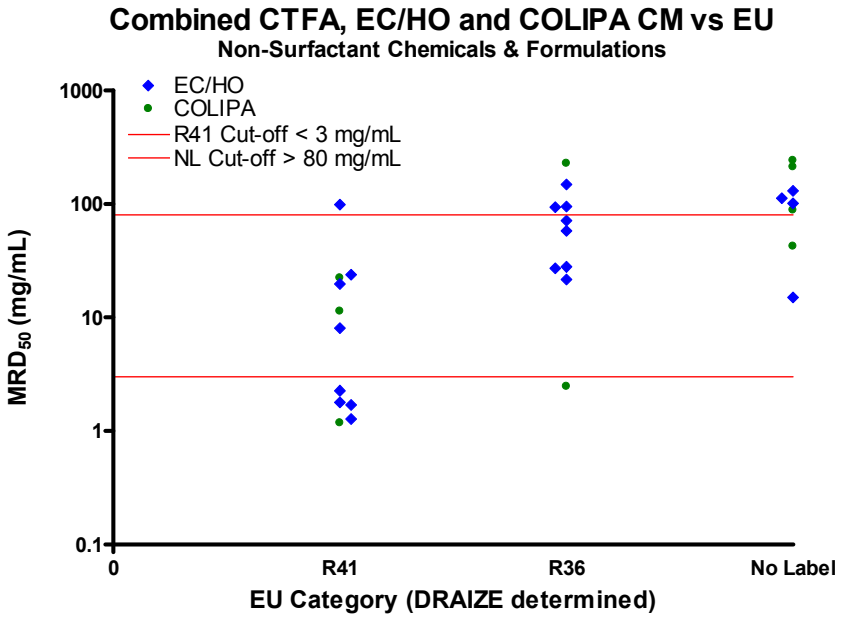

Figure 6.1.3.4.1.a Combined data for all non-surfactant containing chemicals and formulations from the CTFA, EC/HO and COLIPA studies graphed against EU hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define No Label materials and <3 mg/ml to define Category R41 materials are shown. There are 29 total materials; 23 unique chemical:concentration combinations, and 21 totally unique chemicals.

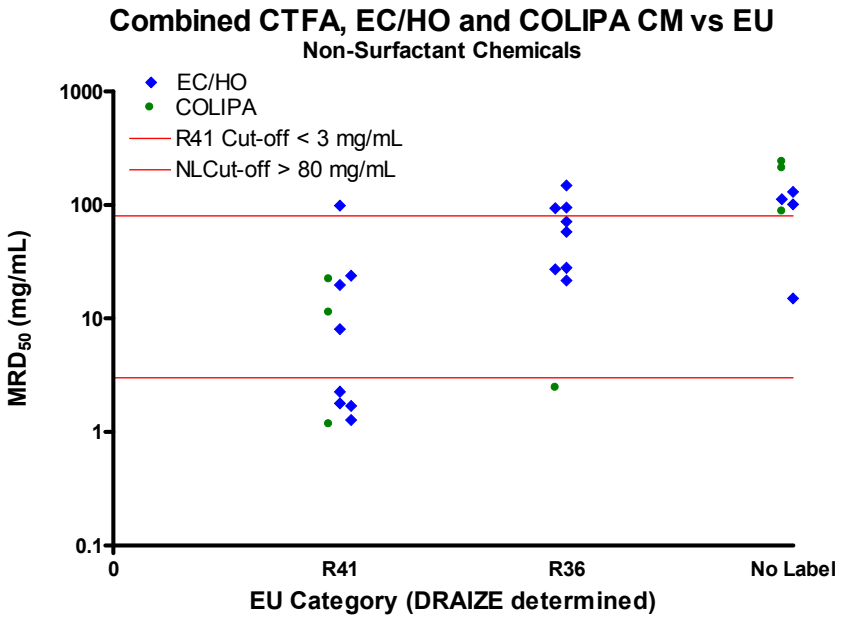

Figure 6.1.3.4.1.b Combined data for non-surfactant chemicals from the CTFA, EC/HO and COLIPA studies graphed against EU hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define No Label materials and <3 mg/ml to define Category R41 materials are shown. There are 27 total materials; 21 unique chemical:concentration combinations, and 19 totally unique chemicals.

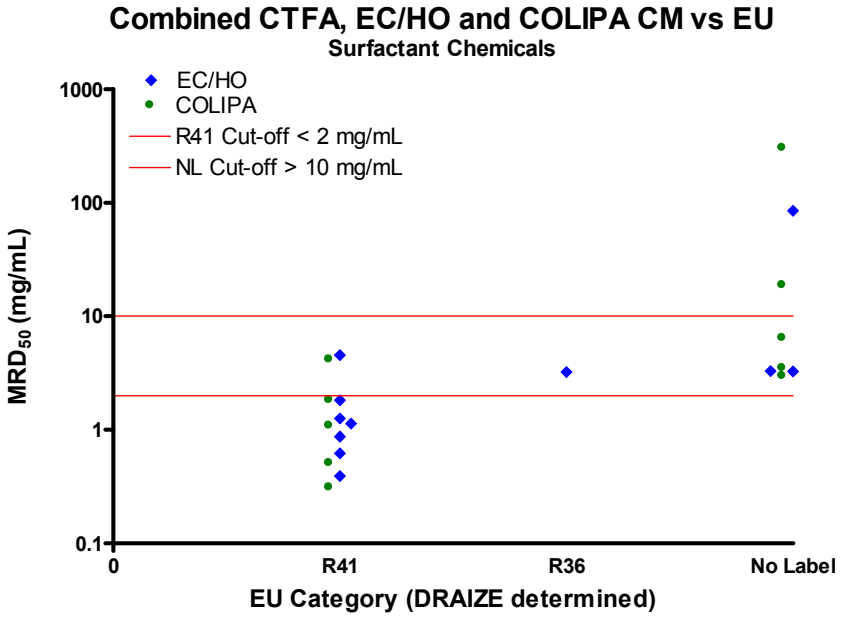

Figure 6.1.3.4.1.c Combined data for surfactant chemicals from the CTFA, EC/HO and COLIPA studies graphed against EU hazard classifications. Proposed MRD_{50} cut-off values of >10 mg/ml to define No Label materials and <2 mg/ml to define Category R41 materials are shown. There are 21 total materials; 13 unique chemical:concentration combinations, and 6 totally unique chemicals.

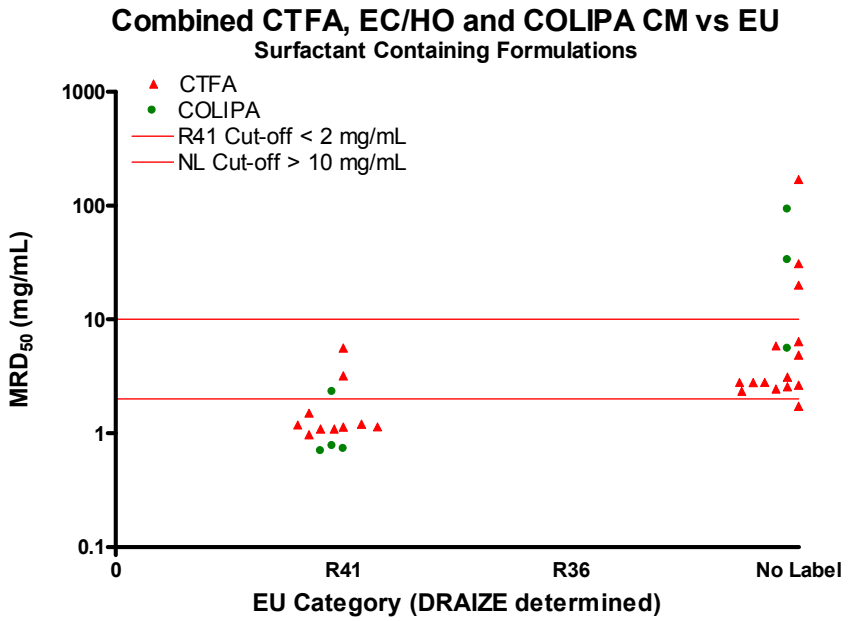

Figure 6.1.3.4.1.d Combined data for surfactant formulations from the CTFA, EC/HO and COLIPA studies graphed against EU hazard classifications. Proposed MRD_{50} cut-off values of >10 mg/ml to define No Label materials and <2 mg/ml to define Category R41 materials are shown. There are 32 totally unique chemicals. The assumption is that the six reformulated test materials for the COLIPA study are different from similar test materials used in the CTFA study.

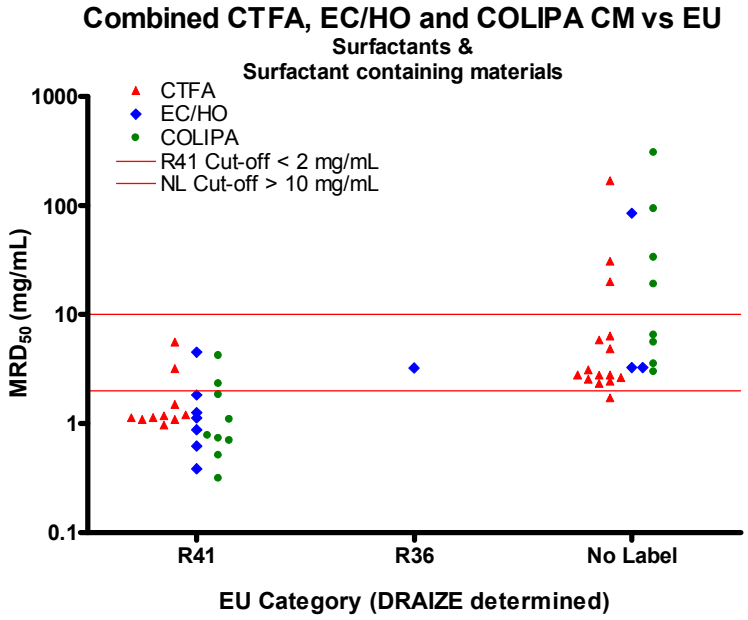

Figure 6.1.3.4.1.e Combined data for surfactants and surfactant containing materials from the CTFA, EC/HO and COLIPA studies graphed against EU hazard classifications. Proposed MRD_{50} cut-off values of >10 mg/ml to define No Label materials and <2 mg/ml to define R41 materials are shown. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. There are 53 total materials; 45 unique chemical:concentration combinations, and 38 totally unique chemicals/formulations.

Starting with the EU hazard classification, a possible prediction model was determined empirically by trying to balance over predictions and under predictions in a conservative manner that kept under predictions to a minimum. For the combined non-surfactant chemicals and formulations, and non-surfactant chemicals alone (Figures 6.1.3.4.1.a and 6.1.3.4.1.b, this seemed to be done by cut-off values at 10 mg/ml and 3 mg/ml such that EU R41 predictions would be made for materials with MRD_{50} values ≤ 3.0 mg/ml, EU R36 predictions would be made for materials where their MRD_{50} was between 3 and 80 mg/ml and Not Classified for materials with MRD_{50} ≥ 80 mg/ml. Cut-off values are represented in the figures by horizontal lines. Some data points which overlapped have been displaced horizontally along the X-axis so that they are more easily visible. The predictive capacity for this prediction model for non-surfactant chemicals (there were only two non-surfactant formulations) is shown in Table 6.1.3.4.1.b. It can be seen from both the Table and the Figure 6.1.3.4.1.b that there is a high level of under prediction of both R41 (55%) and R36 (33%) categories.

A similar approach was taken for the surfactants and surfactant containing formulations. Figures 6.1.3.4.1.c and 6.1.2.4.1.d show that for these materials a slightly lower cut-off (2 mg/ml) was selected to identify the most irritating R41 materials. Materials with a MRD_{50} score of between 2 mg/ml and 10 mg/ml were considered R36 materials, and materials with an MRD_{50} >10 mg/ml were considered to have Not Classified. A significant difficulty with choosing these cut-off values was that there was only one surfactant chemical and no surfactant formulations that carried an R36 classification. Thus it is difficult to have high confidence in the 10 mg/ml cut-off value for the Not Classified classification.

An analysis of the performance statistics that result from using this prediction model are shown in Table 6.1.3.4.1.c (surfactant chemicals) and Table 6.1.3.4.1.d (surfactant containing formulations). These statistics are considerably better than for the non-surfactant materials. There was some under prediction of R41 surfactants or surfactant containing materials (17% and 21%, respectively), but there was considerable over prediction of Not Classified materials (63% for surfactants and 72% for surfactant containing materials).

A further analysis of the ability of the CM to predict either R41 versus the other categories, or EU Not Classified versus the other categories for surfactants and surfactant containing materials is shown in Table 6.1.3.4.f. It can be seen that the positive predictive value (PPV) is much higher for the surfactants and surfactant containing materials (100% for severes vs. the rest [surfactants], 92% severes vs. the rest [surfactant formulations] than it is for non-surfactant chemicals (83% for severes vs. the rest [non-surfactant chemicals].

Because the number of materials that were analyzed either as surfactants or as surfactant-containing materials was still small for each group, we combined these two classes to obtain a data set of 53 test materials of which 50 have unique chemical:concentration combinations and 38 are completely unique chemicals or formulations. For these materials the PPV for severe irritants versus the rest was 95% and

the negative predictive value (NPV) was 84%. For EU non irritants versus the rest, the PPV was 60% and the NPV was 100%.

Table 6.1.3.4.1.a Combined studies – Non-surfactant chemicals and formulations - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values from Figure 6.1.3.4.1.a are applied. N = 29 non-surfactant materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	5	5	1	11	45%	NA	55%
R36	1	5	4	10	50%	10%	40%
Not Classified	0	2	6	8	75%	25%	NA
Total	6	12	11	29	55%		
Predictivity	83%	42%	55%				
Category Under predicted	NA	42%	45%				
Category Overpredicted	17%	17%	NA				

Table 6.1.3.4.1.b Combined studies – Non-surfactant chemicals - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values from Figure 6.1.3.4.1.b are applied. N = 27 non-surfactant chemicals.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	5	5	1	11	45%	NA	55%
R36	1	5	3	9	56%	11%	33%
Not Classified	0	1	6	7	86%	14%	NA
Total	6	11	10	27	59%		
Predictivity	83%	45%	60%				
Category Under predicted	NA	45%	40%				
Category Overpredicted	17%	9%	NA				

Table 6.1.3.4.1.c Combined studies – Surfactant chemicals - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values from Figure 6.1.3.4.1.c are applied. N = 21 surfactant chemicals.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	10	2	0	12	83%	NA	17%
R36	0	1	0	1	100%	0%	0%
Not Classified	0	5	3	8	38%	63%	NA
Total	10	8	3	21	67%		
Predictivity	100%	13%	100%				
Category Under predicted	NA	25%	0%				
Category Overpredicted	0%	63%	NA				

Table 6.1.3.4.1.d Combined studies – Surfactant containing formulations - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values from Figure 6.1.3.4.1.d are applied. N = 32 surfactant formulations.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	11	3	0	14	79%	NA	21%
R36	0	0	0	0	NA	NA	NA
Not Classified	1	12	5	18	28%	72%	NA
Total	12	15	5	32	50%		
Predictivity	92%	0%	100%				
Category Under predicted	NA	20%	0%				
Category Overpredicted	8%	80%	NA				

Table 6.1.3.4.1.e Combined studies – Surfactant Chemicals and Formulations - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values from Figure 6.1.3.4.1.e are applied. N = 53 surfactant materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	21	5	0	26	80.8%	NA	19.2%
R36	0	1	0	1	100.0%	0%	0%
Not Classified	1	17	8	26	30.8%	69.2%	NA
Total	22	23	8	53	56.6%		
Predictivity	95.5%	4.3%	100.0%				
Category Under predicted	NA	21.7%	0%				
Category Overpredicted	4.5%	73.9%	NA				

6.1.3.4.2 Analysis of the combined CTFA, EC/HO, and COLIPA data with GHS category

Possible prediction models for the GHS hazard classification were determined similarly to that described for the EU classification system described in Section 6.1.3.4.1 by setting cut-off values empirically in a conservative manner that kept under predictions to a minimum. For the combined non-surfactant chemicals and formulations, and non-surfactant chemicals (Figures 6.1.3.4.2.a and 6.1.3.4.2.b), this seemed to be done by cut-off values at 80 mg/ml and 3 mg/ml such that GHS 1 predictions would be made for materials with MRD_{50} values ≤ 3.0 mg/ml, GHS 2A predictions would be made for materials where their MRD_{50} was between 3 and 10 mg/ml (it was not possible to discriminate between 2A and 2B categories) and No Label for materials with MRD_{50} ≥ 80 mg/ml. Cut-off values are represented in the figures by horizontal lines. Some data points which overlapped have been displaced horizontally along the x-axis so that they are more

easily visible. The predictive capacity for this prediction model for non-surfactant chemicals (there were only two non-surfactant formulations) is shown in Table 6.1.3.4.2.b. It can be seen from both the Table and the Figure 6.1.3.4.2.b that there is a high level of under prediction of both GHS 1 (45%) and GHS 2A (60%) categories.

A similar approach was taken for the surfactants and surfactant containing formulations. Figures 6.1.3.4.2.c and 6.1.2.4.2.d show that for these materials a slightly lower cut-off (2 mg/ml) was selected to identify the most irritating GHS 1 materials. Materials with a MRD_{50} score of between 2 mg/ml and 10 mg/ml were considered category 2A materials, and materials with an MRD_{50} >10 mg/ml were considered to have No Label. One main difficulty with choosing these cut-off values was that there were only three surfactant chemicals and two surfactant formulations that carried either a 2A or 2B classification. Thus it is difficult to have high confidence in the 10 mg/ml cut-off value for the No Label classification.

An analysis of the performance statistics that result from using this prediction model are shown in Table 6.1.3.4.2.c (surfactant chemicals) and Table 6.1.3.4.2.d (surfactant containing formulations). These statistics are considerably better than for the non-surfactant materials. There was very little under prediction of GHS 1 surfactants or surfactant containing materials (9% and 8%, respectively), but there was considerable over prediction of No Label materials (57% for surfactants and 73% for surfactant containing materials). The PPV for the extremes of the irritation categories were very good for both surfactants and surfactant containing formulations. For surfactants, the predictivity was 100% for GHS 1's and No Label. For surfactant containing formulations, the predictivity was 92% for GHS 1's and 100% for No Label.

For the combination of surfactants and surfactant containing materials the predictivity was 95.5% for GHS 1's (22/22) and 100% for No Label (8/8).

A further analysis of the ability of the CM to predict either GHS 1 versus the other categories, or GHS No Label versus the other categories is shown in Table 6.1.3.4.g. It can be seen that the PPV is high for the non-surfactant chemicals, surfactants and surfactant containing materials (100% for severes vs. the rest [non-surfactant chemicals], 100% for severes vs. the rest [surfactants] and 92% severes vs. the rest [surfactant formulations]. However, the NPV for the non-surfactants is only 76% while it is 91% for the surfactant chemicals and 95% for the surfactant formulations.

Because the number of materials that were analyzed either as surfactants or as surfactant-containing materials was still small for each group, we combined these two classes to obtain a data set of 53 test materials of which 50 have unique chemical:concentration combinations and 38 are completely unique chemicals or formulations. For these materials the PPV for severe irritants versus the rest was 95% and the NPV was 94%. For GHS non irritants versus the rest, the PPV was 62% and the NPV was 100%.

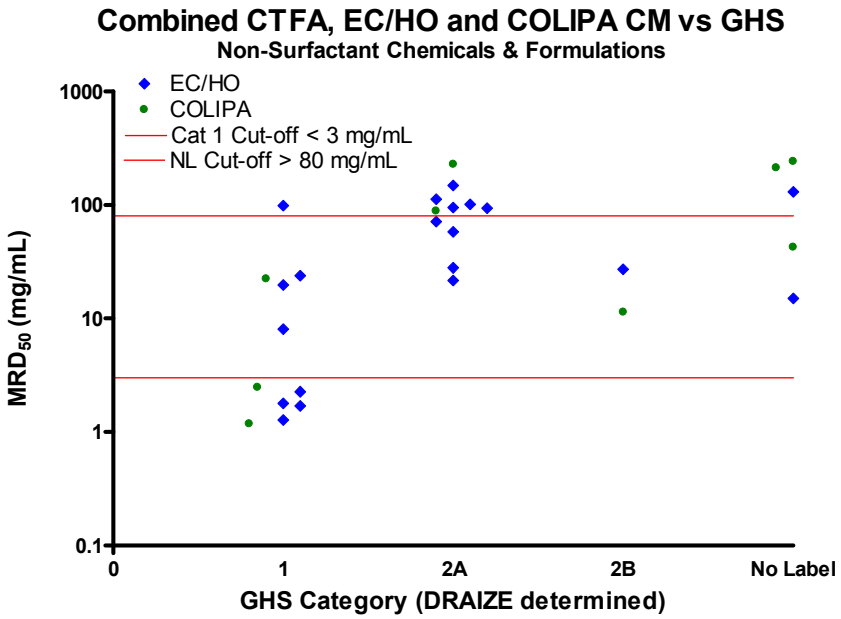

Figure 6.1.3.4.2.a Combined data for all non-surfactant containing chemicals and formulations from the CTFA, EC/HO and COLIPA studies graphed against GHS hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define No Label materials and <3 mg/ml to define Category 1 materials are shown. There are 29 total materials; 23 unique chemical:concentration combinations, and 21 totally unique chemicals.

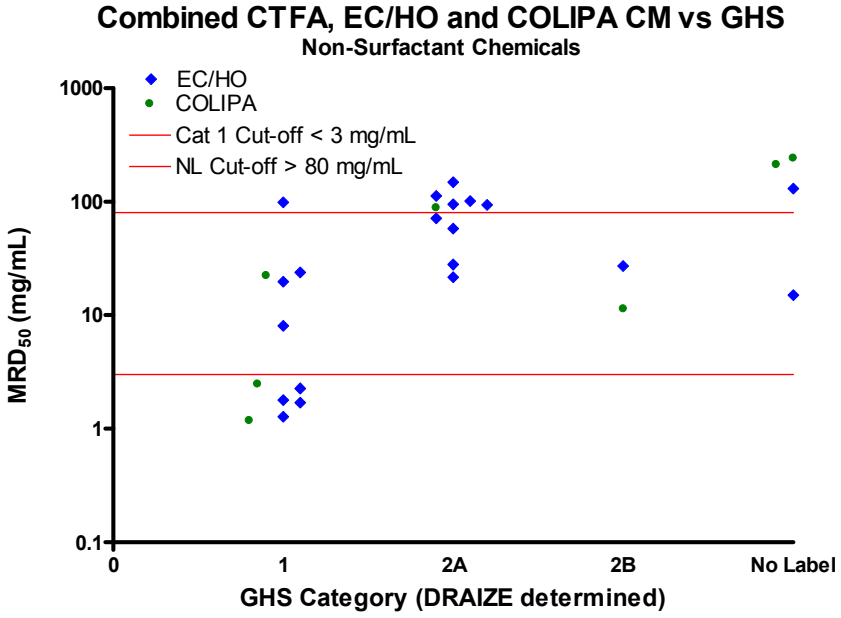

Figure 6.1.3.4.2.b Combined data for non-surfactant containing chemicals from the CTFA, EC/HO and COLIPA studies graphed against GHS hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define No Label materials and <3 mg/ml to define Category 1 materials are shown. There are 27 total materials; 21 unique chemical:concentration combinations, and 19 totally unique chemicals.

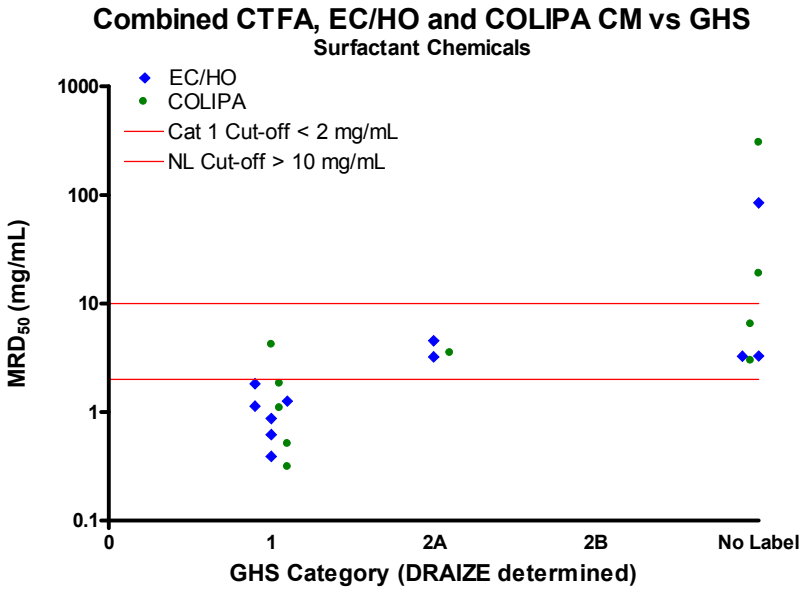

Figure 6.1.3.4.2.c Combined data for surfactant containing chemicals from the CTFA, EC/HO and COLIPA studies graphed against GHS hazard classifications. Proposed MRD_{50} cut-off values of >10 mg/ml to define No Label materials and <2 mg/ml to define Category 1 materials are shown. There are 21 total materials; 13 unique chemical:concentration combinations, and 6 totally unique chemicals.

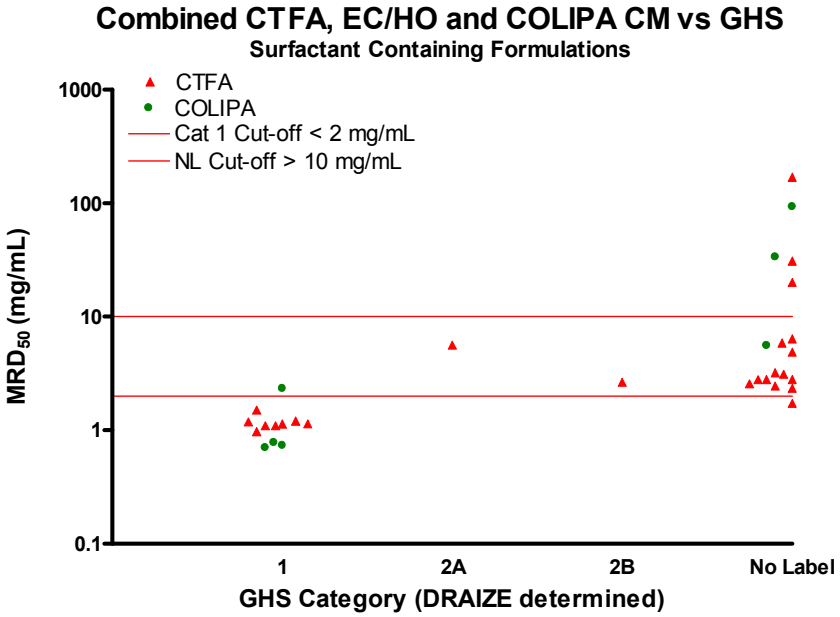

Figure 6.1.3.4.2.d Combined data for surfactant containing formulations from the CTFA, EC/HO and COLIPA studies graphed against GHS hazard classifications. Proposed MRD_{50} cut-off values of >10 mg/ml to define No Label materials and <2 mg/ml to define Category 1 materials are shown. There are 32 totally unique chemicals. The assumption is that the six reformulated test materials for the COLIPA study are different from similar test materials used in the CTFA study.

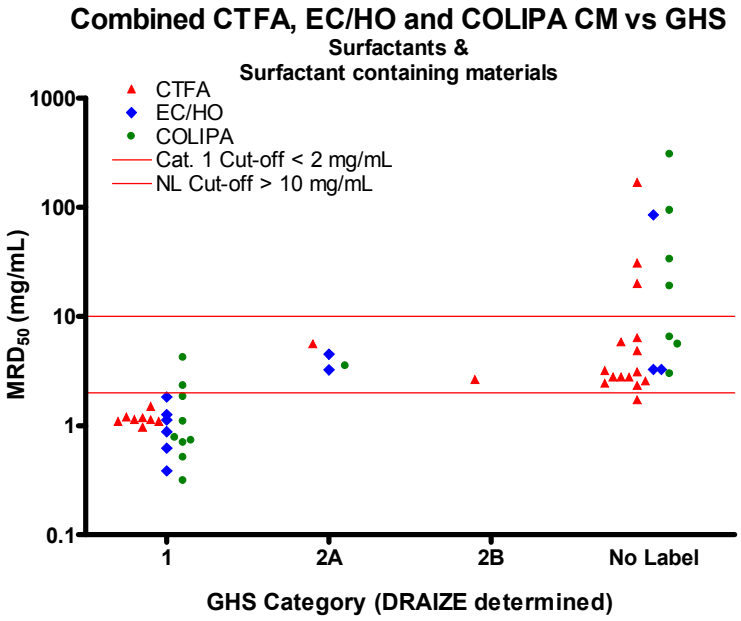

Figure 6.1.3.4.2.e Combined data from the CTFA, EC/HO and COLIPA studies graphed against GHS hazard classifications. Proposed MRD_{50} cut-off values of >10 mg/ml to define No Label materials and <2 mg/ml to define category 1 materials are shown. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. There are 53 total materials; 45 unique chemical:concentration combinations, and 38 totally unique chemicals/formulations.

Table 6.1.3.4.2.a Combined studies – Non-surfactant chemicals and formulations - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values from Figure 6.1.3.4.2.a are applied. N = 29 non-surfactant materials.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	6	4		1	11	55%	NA	45%
2A	0	4		7	11	36%	0%	64%
2B	0	2		0	2	100%	0%	0%
No Label	0	2		3	5	60%	40%	NA
Total	6	12		11	29	52%		
Predictivity	100%	50%		27%				
Category Under predicted	NA	33%		73%				
Category Overpredicted	0%	17%		NA				

Table 6.1.3.4.2.b Combined studies – Non-surfactant chemicals - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values from Figure 6.1.3.4.2.b are applied. N = 27 non-surfactant chemicals.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	6	4		1	11	55%	NA	45%
2A	0	4		6	10	40%	0%	60%
2B	0	2		0	2	100%	0%	0%
No Label	0	1		3	4	75%	25%	NA
Total	6	11		10	27	56%		
Predictivity	100%	55%		30%				
Category Under predicted	NA	36%		70%				
Category Overpredicted	0%	9%		NA				

Table 6.1.3.4.2.c Combined studies – Surfactant chemicals - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values from Figure 6.1.3.4.2.c are applied. N = 21 surfactant chemicals.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	10	1		0	11	91%	NA	9%
2A	0	3		0	3	100%	0%	0%
2B	0	0		0	0	NA	NA	NA
No Label	0	4		3	7	43%	57%	NA
Total	10	8		3	21	76%		
Predictivity	100%	38%		100%				
Category Under predicted	NA	13%		0%				
Category Overpredicted	0%	50%		NA				

Table 6.1.3.4.2.d Combined studies – Surfactant containing formulations - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values from Figure 6.1.3.4.2.d are applied. N = 32 surfactant formulations.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	11	1		0	12	92%	NA	8%
2A	0	1		0	1	100%	0%	0%
2B	0	1		0	1	100%	0%	0%
No Label	1	12		5	18	28%	73%	NA
Total	12	15		5	32	56%		
Predictivity	92%	13%		100%				
Category Under predicted	NA	7%		0%				
Category Overpredicted	8%	80%		NA				

Table 6.1.3.4.2.e Combined studies – Surfactants and surfactant containing materials - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values from Figure 6.1.3.4.2.e are applied. N = 53 surfactant materials.

Draize Determined GHS Category	GHS Category Predicted By CM					Concordance	Toxicity Overpredicted	Toxicity Underpredicted
	1	2A	2B	No Label	Total			
1	21	2		0	23	91.3%	NA	8.7%
2A	0	4		0	4	100.0%	0%	0%
2B	0	1		0	1	100.0%	0%	0%
No Label	1	16		8	25	32.0%	68.0%	NA
Total	22	23		8	53	64.2%		
Predictivity	95.5%	21.7%		100.0%				
Category Under predicted	NA	8.7%		0%				
Category Overpredicted	4.5%	69.6%		NA				

6.1.3.4.3 Analysis of the combined CTFA, EC/HO, and COLIPA data with EPA category

Possible prediction models for the EPA hazard classification were determined similarly to that described for the EU classification system described in Section 6.1.3.4.1 by setting cut-off values empirically in a conservative manner that kept under predictions to a minimum. For the combined non-surfactant chemicals and formulations, and non-surfactant chemicals (Figures 6.1.3.4.3.a and 6.1.3.4.3.b), this seemed to be done by cut-off values at 80 mg/ml and 3 mg/ml such that EPA I predictions would be made for materials with MRD_{50} values ≤ 3.0 mg/ml, EPA II predictions would be made for materials where their MRD_{50} was between 3 and 10 mg/ml (it was not possible to discriminate between EPA II and EPA III categories) and EPA IV for materials with MRD_{50} ≥ 80 mg/ml. Cut-off values are represented in the figures by horizontal lines. Some data points which overlapped have been displaced horizontally along the x-axis so that they are more easily visible. The predictive capacity for this prediction model for non-surfactant chemicals (there were only two non-surfactant formulations) is shown in Table 6.1.3.4.3.b. It can be seen from both the Table and the Figure 6.1.3.4.3.b that there is a high level of under prediction of both EPA I (43%) and EPA II (60%) categories.

A similar approach was taken for the surfactants and surfactant containing formulations. Figures 6.1.3.4.3.c and 6.1.2.4.3.d show that for these materials a slightly lower cut-off (2 mg/ml) was selected to identify the most irritating EPA I materials. Materials with a MRD_{50} score of between 2 mg/ml and 80 mg/ml were considered category II materials, and materials with an MRD_{50} >80 mg/ml were considered to be EPA IV. One main difficulty with choosing these cut-off values was that there was only one surfactant chemical that carried an EPA IV classification.

Analyses of the performance statistics that result from using this prediction model are shown in Table 6.1.3.4.3.c (surfactant chemicals) and Table 6.1.3.4.3.d (surfactant containing formulations). These statistics are better than for the non-surfactant materials.

There was some under prediction of EPA I surfactants or surfactant containing materials (22% and 21%, respectively), and also some over prediction of EPA IV materials (0% for surfactants [but only one material was in that category] and 60% for surfactant containing materials). The positive predictive values for the extremes of the irritation categories were in general good for both surfactants and surfactant containing formulations. For surfactants, the predictivity was 78% for EPA I's and 50% for EPA IV's. For surfactant containing formulations, the predictivity was 92% for EPA I's and 100% for EPA IV's.

For the combination of surfactants and surfactant containing materials the predictivity was 85.7% for EPA I's (18/21) and 75% for EPA IV's (3/4).

A further analysis of the ability of the CM to predict either EPA I versus the other categories or EPA IV versus the other categories is shown in Table 6.1.3.4.h. It can be seen that the PPV is moderate for the non-surfactant chemicals, surfactants and surfactant containing materials (80% for severes vs. the rest [non-surfactant chemicals], 78% for severes vs. the rest [surfactants] and 92% severes vs. the rest [surfactant formulations]. The NPV for the non-surfactants is 83%, while it is 82% for the surfactant chemicals and 85% for the surfactant formulations.

Because the number of materials that were analyzed either as surfactants or as surfactant-containing materials was still small for each group, we combined these two classes to obtain a data set of 52 test materials of which 49 have unique chemical:concentration combinations and 38 are completely unique chemicals or formulations. For these materials the PPV for severe irritants versus the rest was 86% and the NPV was 84%. For EU non irritants versus the rest, the PPV was 94% and the NPV was 75%.

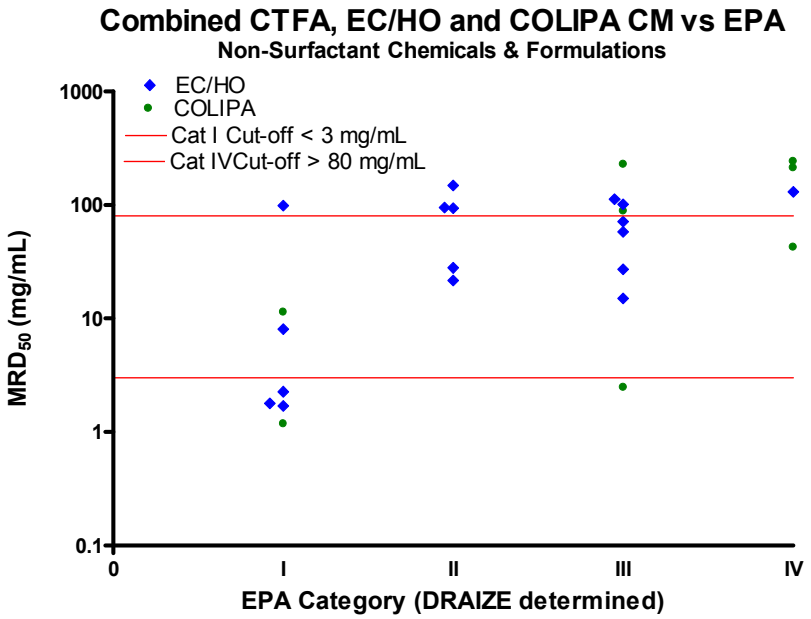

Figure 6.1.3.4.3.a Combined data for all non-surfactant chemicals and formulations from the CTFA, EC/HO and COLIPA studies graphed against EPA hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define category IV materials and <3 mg/ml to define category I materials are shown. There are 25 total materials; 20 unique chemical:concentration combinations, and 18 totally unique chemicals.

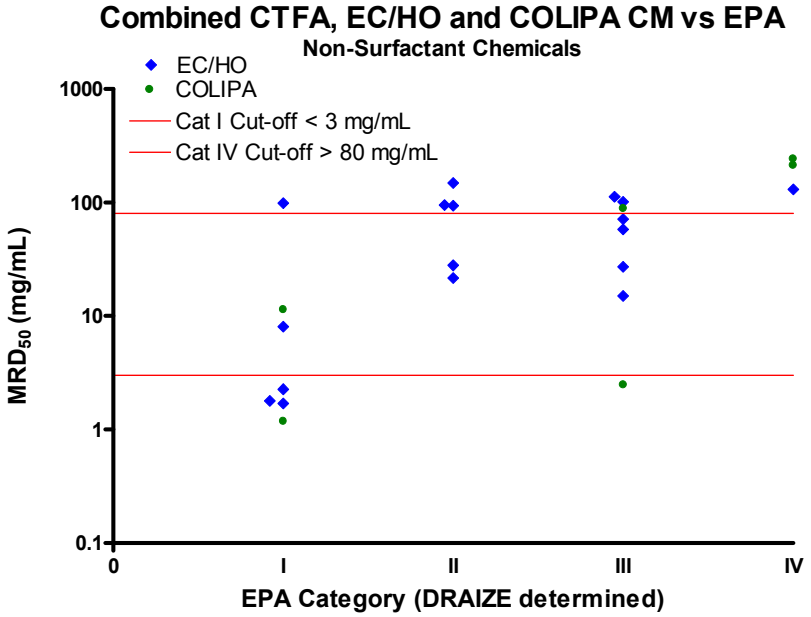

Figure 6.1.3.4.3.b Combined data for non-surfactant chemicals from the CTFA, EC/HO and COLIPA studies graphed against EPA hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define category IV materials and <3 mg/ml to define category I materials are shown. There are 23 total materials; 18 unique chemical:concentration combinations, and 16 totally unique chemicals.

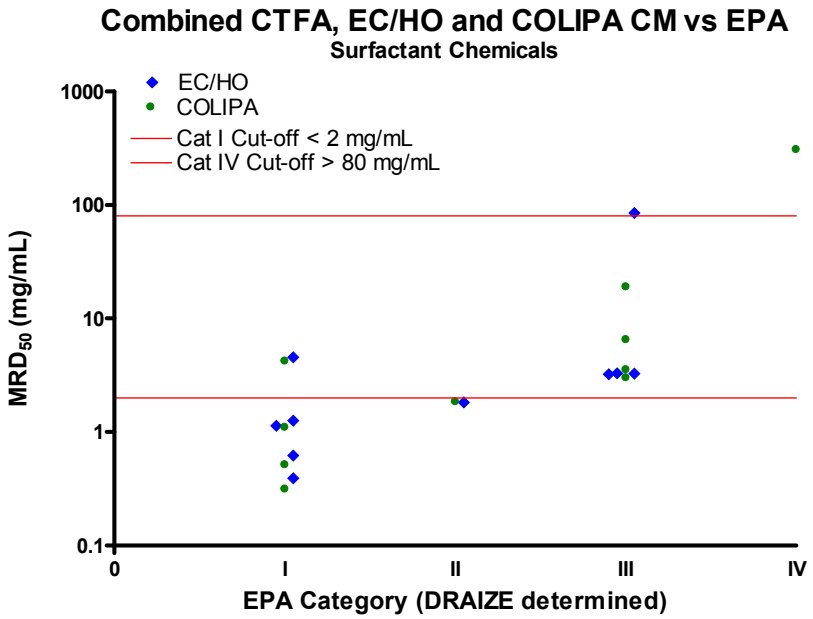

Figure 6.1.3.4.3.c Combined data for surfactant chemicals from the CTFA, EC/HO and COLIPA studies graphed against EPA hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define category IV materials and <2 mg/ml to define category I materials are shown. There are 20 total materials; 12 unique chemical:concentration combinations, and 6 totally unique chemicals.

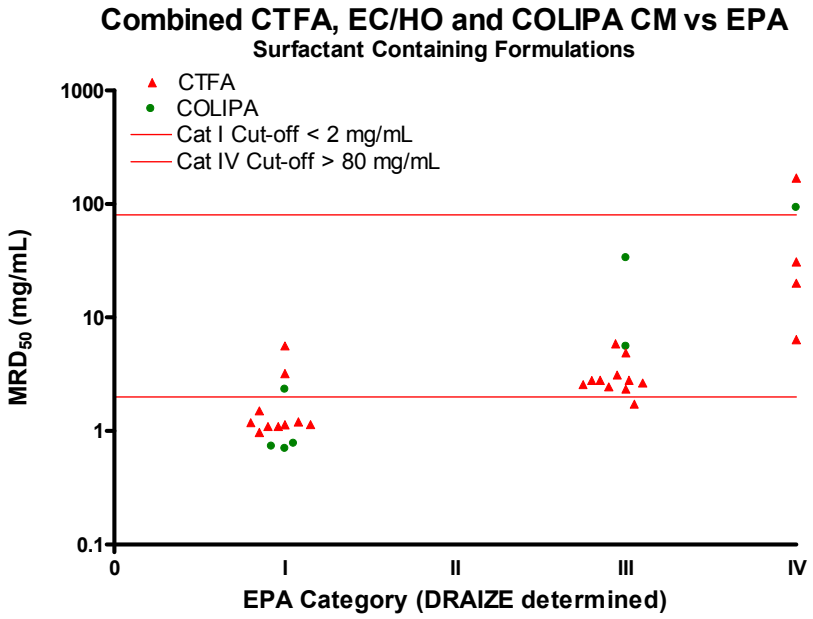

Figure 6.1.3.4.3.d Combined data for surfactant formulations from the CTFA, EC/HO and COLIPA studies graphed against EPA hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define category IV materials and <2 mg/ml to define category I materials are shown. There are 32 totally unique chemicals. The assumption is that the six reformulated test materials for the COLIPA study are different from similar test materials used in the CTFA study.

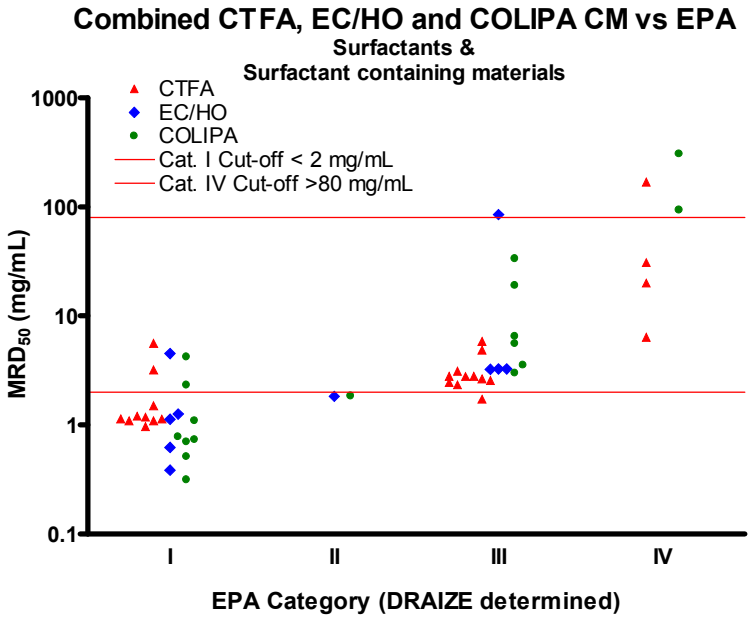

Figure 6.1.3.4.3.e Combined data for all surfactants and surfactant containing materials from the CTFA, EC/HO and COLIPA studies graphed against EPA hazard classifications. Proposed MRD_{50} cut-off values of >80 mg/ml to define Category IV materials and <2 mg/ml to define Category I materials are shown. In some cases data points have been slightly offset along the X-axis in order to clearly separate them from data of similar magnitude. There are 52 total materials; 44 unique chemical:concentration combinations, and 38 totally unique chemicals/formulations.

Table 6.1.3.4.3.a Combined studies – Non-surfactant chemicals and formulations - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values from Figure 6.1.3.4.3.a are applied. N = 25 non-surfactant materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	4	2		1	7	57%	NA	43%
II	0	2		3	5	40%	0%	60%
III	1	4		4	9	44%	11%	44%
IV	0	1		3	4	75%	25%	NA
Total	5	9		11	25	52%		
Predictivity	80%	67%		27%				
Category Underpredicted	NA	22%		73%				
Category Overpredicted	20%	11%		NA				

Table 6.1.3.4.3.b Combined studies – Non-surfactant chemicals - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values from Figure 6.1.3.4.3.b are applied. N = 23 non-surfactant chemicals.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	4	2		1	7	57%	NA	43%
II	0	2		3	5	40%	0%	60%
III	1	4		3	8	50%	13%	37%
IV	0	0		3	3	100%	0%	NA
Total	5	8		10	23	57%		
Predictivity	80%	75%		30%				
Category Underpredicted	NA	25%		70%				
Category Overpredicted	20%	0%		NA				

Table 6.1.3.4.3.c Combined studies – Surfactant chemicals - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values from Figure 6.1.3.4.3.c are applied. N = 20 surfactant chemicals.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	7	2		0	9	78%	NA	22%
II	2	0		0	2	0%	100%	0%
III	0	7		1	8	88%	0%	13%
IV	0	0		1	1	100%	0%	NA
Total	9	9		2	20	75%		
Predictivity	78%	78%		50%				
Category Underpredicted	NA	22%		50%				
Category Overpredicted	22%	0%		NA				

Table 6.1.3.4.3.d Combined studies – Surfactant containing formulations - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values from Figure 6.1.3.4.3.d are applied. N = 32 surfactant formulations.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	11	3		0	14	79%	NA	21%
II	0	0		0	0	NA	NA	NA
III	1	12		0	13	92%	8%	0%
IV	0	3		2	5	40%	60%	NA
Total	12	18		2	32	78%		
Predictivity	92%	67%		100%				
Category Underpredicted	NA	17%		0%				
Category Overpredicted	8%	17%		NA				

Table 6.1.3.4.3.e Combined studies - Surfactant Chemicals and Formulations - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values from Figure 6.1.3.4.3.e are applied. N = 52 surfactant materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	18	5		0	23	78.3%	NA	21.7%
II	2	0		0	2	0.0%	100.0%	0%
III	1	19		1	21	90.5%	4.8%	4.8%
IV	0	3		3	6	50.0%	50.0%	NA
Total	21	27		4	52	76.9%		
Predictivity	85.7%	70.4%		75.0%				
Category Underpredicted	NA	18.5%		25.0%				
Category Overpredicted	14.3%	11.1%		NA				

An analysis of the ability of the CM assay to test surfactants and surfactant-based materials and separate severe irritants from "the rest", and non-irritants from "the rest" is shown in Table 6.1.3.4.m. Using the larger data set of materials combined from the HO/EC, CTFA Phase III and COLIPA data sets, the CM assay seemed to perform reasonably well when separating EU and GHS severe irritants from "the rest" (both had positive predictivities of 100% and negative predictivities of 78%). However, this analysis is confounded some what by the uncertainty of how many truly "unique" materials are represented in overlapping data set of 53 (or 52 EPA) materials.

Table 6.1.3.4.f Combined studies - Concordance for severe irritants versus the rest and non-irritants versus the rest for the combined studies. Cut-off values from Figures 6.1.3.4.1.a – 6.1.3.4.1.e were applied. N = 29 non-surfactant materials, 27 non-surfactant chemicals, 21 surfactant chemicals, 32 surfactant formulations, and 53 surfactant materials.

		Vs Rest													
		Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Non-Surfactant Chemicals and Formulations	EU Severe irritants	76%	22/29	45%	5/11	94%	17/18	83%	5/6	74%	17/23	6%	1/18	55%	6/11
	EU non-irritants	76%	22/29	76%	16/21	75%	6/8	89%	16/18	55%	6/11	25%	2/8	24%	5/21
Non-Surfactant Chemicals	EU Severe irritants	74%	20/27	45%	5/11	94%	15/16	83%	5/6	71%	15/21	6%	1/16	55%	6/11
	EU non-irritants	81%	22/27	80%	16/20	86%	6/7	94%	16/17	60%	6/10	14%	1/7	20%	4/20
Surfactant Chemicals	EU Severe irritants	90%	19/21	83%	10/12	100%	9/9	100%	10/10	82%	9/11	0%	0/9	17%	2/12
	EU non-irritants	76%	16/21	100%	13/13	38%	3/8	72%	13/18	100%	3/3	63%	5/8	0%	0/13
Surfactant Formulations	EU Severe irritants	88%	28/32	79%	11/14	94%	17/18	92%	11/12	85%	17/20	6%	1/18	21%	3/14
	EU non-irritants	59%	19/32	100%	14/14	28%	5/18	52%	14/27	100%	5/5	72%	13/18	0%	0/14
Surfactant Chemicals and Formulations	EU Severe irritants	89%	47/53	81%	21/26	96%	26/27	95%	21/22	84%	26/31	4%	1/27	19%	5/26
	EU non-irritants	66%	35/53	100%	27/27	31%	8/26	60%	27/45	100%	8/8	69%	18/26	0%	0/27

Table 6.1.3.4.g Combined studies - Concordance for severe irritants versus the rest and non-irritants versus the rest for the combined studies. Cut-off values from Figures 6.1.3.4.2.a – 6.1.3.4.2.e were applied. N = 29 non-surfactant materials, 27 non-surfactant chemicals, 21 surfactant chemicals, 32 surfactant formulations, and 53 surfactant materials.

		Vs Rest													
		Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Non-Surfactant Chemicals and Formulations	GHS Severe irritants	83%	24/29	55%	6/11	100%	18/18	100%	6/6	78%	18/23	0%	0/18	45%	5/11
	GHS non-irritants	66%	19/29	67%	16/24	60%	3/5	89%	16/18	27%	3/11	40%	2/5	33%	8/24
Non-Surfactant Chemicals	GHS Severe irritants	81%	22/27	55%	6/11	100%	16/16	100%	6/6	76%	16/21	0%	0/16	45%	5/11
	GHS non-irritants	70%	19/27	70%	16/23	75%	3/4	94%	16/17	30%	3/10	25%	1/4	30%	7/23
Surfactant Chemicals	GHS Severe irritants	95%	20/21	91%	10/11	100%	10/10	100%	10/10	91%	10/11	0%	0/10	9%	1/11
	GHS non-irritants	81%	17/21	100%	14/14	43%	3/7	78%	14/18	100%	3/3	57%	4/7	0%	0/14
Surfactant Formulations	GHS Severe irritants	94%	30/32	92%	11/12	95%	19/20	92%	11/12	95%	19/20	5%	1/20	8%	1/12
	GHS non-irritants	59%	19/32	100%	14/14	28%	5/18	52%	14/27	100%	5/5	72%	13/18	0%	0/14
Surfactant Chemicals and Formulations	GHS Severe irritants	94%	50/53	91%	21/23	97%	29/30	95%	21/22	94%	29/31	3%	1/30	9%	2/23
	GHS non-irritants	68%	36/53	100%	28/28	32%	8/25	62%	28/45	100%	8/8	68%	17/25	0%	0/28

Table 6.1.3.4.h Combined studies - Concordance for severe irritants versus the rest and non-irritants versus the rest for the combined studies. Cut-off values from Figures 6.1.3.4.3.a – 6.1.3.4.3.e were applied. N = 25 non-surfactant materials, 23 non-surfactant chemicals, 20 surfactant chemicals, 32 surfactant formulations, and 52 surfactant materials.

		Vs Rest													
		Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Non-Surfactant Chemicals and Formulations	EPA Severe irritants	84%	21/25	57%	4/7	94%	17/18	80%	4/5	85%	17/20	6%	1/18	43%	3/7
	EPA non-irritants	64%	16/25	62%	13/21	75%	3/4	93%	13/14	27%	3/11	25%	1/4	38%	8/21
Non-Surfactant Chemicals	EPA Severe irritants	83%	19/23	57%	4/7	94%	15/16	80%	4/5	83%	15/18	6%	1/16	43%	3/7
	EPA non-irritants	70%	16/23	65%	13/20	100%	3/3	100%	13/13	30%	3/10	0%	0/3	35%	7/20
Surfactant Chemicals	EPA Severe irritants	80%	16/20	78%	7/9	82%	9/11	78%	7/9	82%	9/11	18%	2/11	22%	2/9
	EPA non-irritants	95%	19/20	95%	18/19	100%	1/1	100%	18/18	50%	1/2	0%	0/1	5%	1/19
Surfactant Formulations	EPA Severe irritants	88%	28/32	79%	11/14	94%	17/18	92%	11/12	85%	17/20	6%	1/18	21%	3/14
	EPA non-irritants	91%	29/32	100%	27/27	40%	2/5	90%	27/30	100%	2/2	60%	3/5	0%	0/27
Surfactant Chemicals and Formulations	EPA Severe irritants	85%	44/52	78%	18/23	90%	26/29	86%	18/21	84%	26/31	10%	3/29	22%	5/23
	EPA non-irritants	92%	48/52	98%	45/46	50%	3/6	94%	45/48	75%	3/4	50%	3/6	2%	1/46

6.1.3.5 Analysis of Company # 1 unpublished data

The reference data for the Company # 1 studies (Annex K) are all from LVET studies. We were informed by Company # 1 that these data were gathered over a number of years from studies carried out at various animal facilities. Some of the studies were done with full GLP compliance and some were not; however, there is no indication in the data received from Company # 1 which did have GLP compliance. The reference chemicals are mainly raw surfactants and surfactant-containing formulations. There were 76 materials which had unique CM values although 80 materials are listed in Table 6.1.3.5.a. The four additional listings represent a second animal test for 4 separate materials. The results of the second animal test for these materials were listed to give an indication of the variability of the animal test. It should be noted that in each of the four sets of duplicate animal tests at least one of the EU, GHS or EPA categories differed in one test from what was found in a presumably duplicate test. The variability of the LVET has been shown to be similar to that of the traditional Draize test (Cormier, Parker et al. 1996).

The 76 materials which make of this data set are highly biased towards mildness; the vast majority of the tests (61 out of 80 tests; 76%) are Not Classified when using the EU system, or have No Category (39 out of 80 tests; 49%) when using the GHS system. However, when using the EPA scoring system a much smaller fraction (13 out of 80 tests; 16%) are classified as Category IV materials. Specifically under the EU system 61 are Not Classified, 10 (12%) are R36 and 9 (11%) are R41. Under the GHS system 39 have No Category, 17 (21%) are Category 2B, 16 (20%) are Category 2A, and 8 (10%) are Category 1. Under the EPA system 13 are Category IV, 47 (59%) are Category III, 11 (14%) are Category II and 9 (11%) are Category I.

The *in vitro* data come from primarily two sources – internal Company # 1 laboratories and Company # 4 (later the Company # 3). *In vitro* data were generated by both the SM and CM. For the purposes of this analysis, all SM MRD_{50} values have been transformed to equivalent CM MRD_{50} values by an algorithm described earlier in this BRD in section 2.2.1.1. Raw data (including the original SM MRD_{50} values) for these studies can be found in Annex F.

Because the Company # 1 materials were being tested for commercial use when the data were generated, the identities of the materials will not be made publicly available for this BRD. However, the materials have been characterized by Company # 1 staff with respect to the type of ingredient being tested or to the primary components of the formulation. Only those materials which are described as surfactants or as formulations which are surfactant based are analyzed in the following section.

Table 6.1.3.5.a Summary of Company # 1 study. N = 76 materials.

In Vivo Eye Irritation Classifications – Company 1										
BRD chemical number	Substance	CASRN[1]	Concentration Tested	Purity (%)	n. of animals	*In Vivo* EU[2,3]	*In Vivo* GHS[4,5]	*In Vivo* EPA[6,7]	ECETOC MMAS Score[8]	MRD_{50} (mg/mL)
1001					6	Not classified	No Category	Category III		0.435
1002					6	Not classified	No Category	Category III		0.535
1003					6	Not classified	Category 2A	Category II		0.44
1004					3	Not classified	Category 2B	Category III		0.421
1005					3	Not classified	Category 2A	Category II		0.411
1006					3	Not classified	No Category	Category III		0.443
1007					3	Not classified	Category 2B	Category III		0.428
1008					3	Not classified	Category 2B	Category III		0.272
1009					6	Not classified	No Category	Category III		0.465
1010					3	R41	Category 1	Category I		0.456
1011					3	R41	Category 1	Category I		0.44
1012					3	Not classified	Category 2A	Category II		0.415
1013					3	Not classified	Category 2B	Category III		0.426
1014					3	Not classified	No Category	Category III		0.444
1015					3	Not classified	No Category	Category III		0.412
1016					3	Not classified	Category 2B	Category III		0.272
1017					3	Not classified	Category 2B	Category III		0.432
1018					3	R36	Category 2B	Category III		0.465
1019					3	R41	Category 1	Category I		0.276
1020					3	R41	Category 1	Category I		0.296
1021					6	Not classified	No Category	Category III		0.19
1022					6	R41	Category 2A	Category I		0.51
1023	**Same as 1022; 2nd animal study***				6	R36	Category 2A	Category III		
1024					3	R41	Category 1	Category I		0.2
1025					6	Not classified	No Category	Category III		0.829
1026					6	Not classified	Category 2B	Category III		0.434
1027					6	Not classified	No Category	Category III		0.44
1028					3	Not classified	Category 2A	Category III		0.46
1029					3	R36	Category 2B	Category III		0.45
1030					3	Not classified	Category 2B	Category III		0.6
1031					3	R36	Category 2B	Category III		0.5
1032					3	R36	Category 2A	Category III		0.96
1033	**Same as 1032; 2nd animal study***				3	Not classified	Category 2B	Category III		
1034					3	Not classified	No Category	Category III		0.67
1035					3	Not classified	No Category	Category IV		63.9
1036					3	Not classified	No Category	Category III		0.79

1039		3	R41	Category 1	Category I	0.26
1040		3	R36	Category 2A	Category II	0.76
1041		3	R36	Category 2A	Category II	0.22
1043		6	Not classified	Category 2A	Category II	0.407
1044		6	Not classified	Category 2A	Category II	0.428
1045		6	Not classified	Category 2A	Category III	0.344
1046		3	R36	Category 2A	Category II	0.264
1047		3	Not classified	No Category	Category III	0.286
1051		3	Not classified	Category 2B	Category III	7.103
1052		3	Not classified	No Category	Category III	1.354
1053		3	Not classified	Category 2B	Category III	0.0808
1054		3	Not classified	Category 2B	Category III	0.0773
1055		3	R36	Category 2A	Category II	0.638
1056		3	R36	Category 2A	Category II	0.817
1057	**Same as 1056; 2nd animal study***	3	R41	Category 1	Category I	
1058		3	Not classified	Category 2A	Category II	0.81
1059		3	Not classified	No Category	Category IV	0.787
1060		3	Not classified	No Category	Category III	0.9
1061		3	Not classified	No Category	Category III	26.733
1062		3	Not classified	No Category	Category IV	46.5
1063		3	Not classified	No Category	Category III	43.1
1064		3	Not classified	No Category	Category III	0.501
1065		3	Not classified	No Category	Category IV	300
1066		3	Not classified	No Category	Category III	3.8
1067		3	Not classified	No Category	Category III	2.573
1068		3	Not classified	No Category	Category III	4.308
1069		3	Not classified	No Category	Category III	0.556
1070		3	Not classified	No Category	Category III	1.96
1071		3	Not classified	No Category	Category IV	0.66
1072		3	Not classified	No Category	Category IV	3.718
1074		3	Not classified	No Category	Category III	4.19
1075		6	Not classified	No Category	Category IV	10.96
1076		6	Not classified	No Category	Category III	0.63
1077			R41	Category 1	Category I	0.63
1078		6	Not classified	No Category	Category III	0.49
1079		3	Not classified	Category 2B	Category III	0.71
1080	**Same as 1079; 2nd animal study***	3	Not classified	No Category	Category IV	
1081		3	Not classified	No Category	Category IV	0.72
1082		3	Not classified	No Category	Category IV	2.02
1083		3	Not classified	No Category	Category III	1.43

1084	3	Not classified	No Category	Category IV	3.86
1085	3	Not classified	No Category	Category IV	15.18
1086	6	Not classified	Category 2B	Category III	0.93
1087	3	Not classified	No Category	Category IV	2.49

[1]CASRN=Chemical Abstract Services Registry Number

[2]EU=European Union (EU [2001]).

[3]Risk phrase R41 = risk of serious damage to the eyes; R36 = irritating to the eyes; not classified.

[4]GHS=Globally Harmonized System (UN [2003])

[5]Eye Irritant Category 1 = irreversible effects on the eye/serious damage to the eye; Category 2A = reversible effects on the eye/irritating to the eyes; Category 2B = reversible effects on the eye/mildly irritating to the eyes; No category

[6]EPA=U.S. Environmental Protection Agency (EPA [1996]).

[7]Toxicity Category I for the Primary Eye Irritation Study = Corrosive, or corneal involvement or irritation not reversible within 21 days; Category II = Corneal involvement or irritation clearing in 8-21 days; Category III = Corneal involvement or irritation clearing in 1-7 daysl; Category IV: minimal effects clearing in less than 24 hr

[9]MMAS scores reported in Harbell et al. (1999)

[9]SCNM=Study Crtieria Not Met

[10]n.p.=not provided

* - Most severe animal data used for graphing purposes.

In order to visualize the relationship between the CM MRD_{50} data and the EU, GHS, and EPA hazard classification system, scatter plot graphs were constructed of MRD_{50} versus hazard category. Figure 6.1.3.5.a depicts the EU categories, Figure 6.1.3.5.b depicts the GHS categories, and Figure 6.1.3.5.c depicts the EPA categories. Starting with EU hazard classification, a possible prediction model was determined empirically by trying to balance over predictions and under predictions in a conservative manner that kept under predictions to a minimum. However, the grouping of the data points showed that there was no obvious difference in MRD_{50} values between R36 and R41 materials. Therefore, we focused just on separating the No Label materials from the combination of R41 and R36 materials. A conservative cut-off value seemed to be at MRD_{50} >2 mg/ml for the No Label materials. This is somewhat above the highest R36 value and seems to be a reasonable approach since there are only eight R36 materials in the analysis. Thus an MRD_{50} score >2 mg/ml would be considered to have a No Label category. Materials with values ≤2 mg/ml would be considered to be either R41 or R36 with the default being the most severe R41 category. The exact designation would have to be determined by a second *in vitro* test which was validated to be able to differentiate between R36 and R41 materials. Figure 6.1.3.5.a shows the 76 data points and the cut-off value represented by a horizontal line.

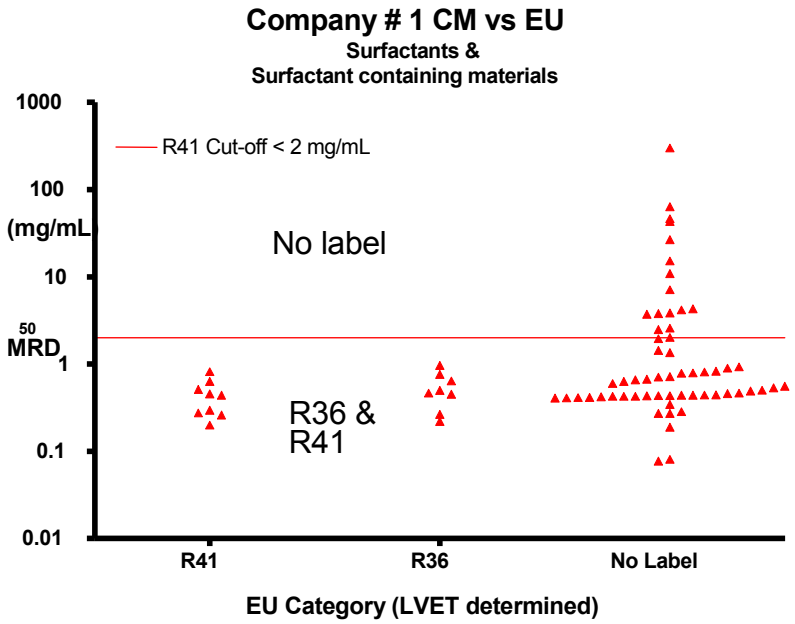

Figure 6.1.3.5.a Results from 76 materials from Company # 1 related to EU classification. For the four materials in Table 6.1.3.5.a which had 2 independent animal tests only the most irritating animal scores were used in this figure.

When the above proposed prediction model (MRD_{50} >2 mg/ml) is applied to the data in Figure 6.1.3.5.a, 16 of 59 (27.1%) of the materials determined to be Not Classified by the animal test would be correctly identified by the CM. This means that 72.9% of the animal designated Not Classified materials would have to be tested in a second level test. None of the materials identified as R41 or R36 by the animal test would be under predicted.

Applying a slightly more aggressive cut-off of MRD_{50} > 1 mg/ml = Not Classified, would raise the number of correctly identified Not Classified materials to 19 out of 59 (32.2%), again with no under prediction of R41 or R36 materials. However, since the 1 mg/ml cut-off is just above the lowest score of the R36 materials, it would seem that more R36 materials should be tested to determine if the range of MRD_{50} values found in this study holds for a larger data set before accepting a cut-off of 1 mg/ml.

An assay of this type could be very useful as a screening assay for products which are very mild. Any products or materials which are below the cut-off might be excluded because of their toxicity, but materials scoring above the cut-off could be reasonably assured to be a non-irritant.

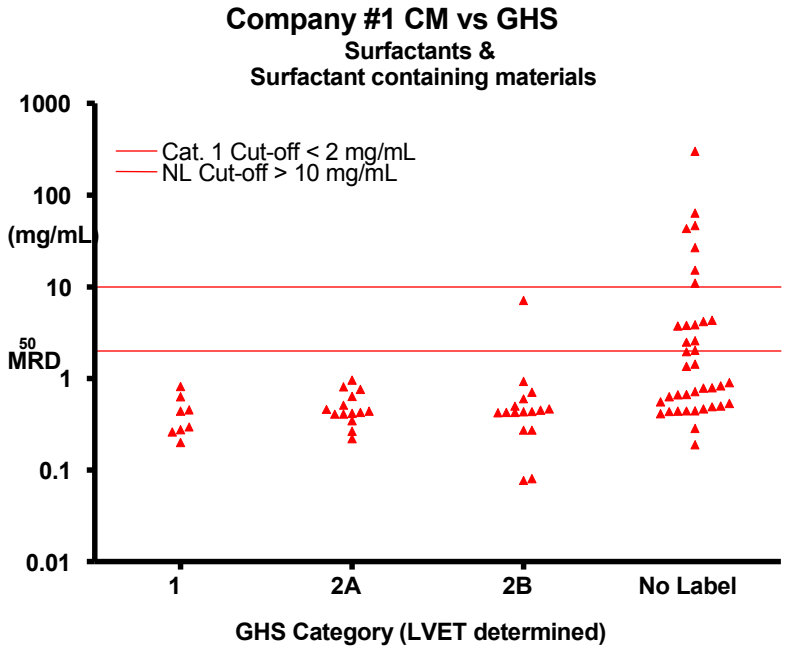

Figure 6.1.3.5.b Results from 76 materials from Company # 1 related to GHS classification. For the four materials in Table 6.1.3.5.b which had 2 independent animal tests only the most irritating animal scores were used in this figure.

As with the EU analysis, the GHS hazard classification analysis began with an attempt to set cut-offs which balanced over predictions and under predictions in a conservative manner that kept under predictions to a minimum. However, the grouping of the data points (Figure 6.1.3.5.b) showed that there was no obvious difference in MRD_{50} values between Category 1, 2A and 2B materials. Again we focused just on separating the No Category materials from the combination of Category 1, 2A and 2B materials. A conservative cut-off value seemed to be at MRD_{50} >10 mg/ml. This is slightly above the highest Category 2B value which seems to be a reasonable approach since there only 16 Category 2B materials in the analysis and it would reasonably assure that no under predictions would occur. Thus an MRD_{50} score >10 mg/ml would be considered to have No Category. Materials with values ≤2 mg/ml would have to be considered to be Category 1 materials unless there was a second *in vitro* test which was validated to successfully categorize them as either 1, 2A, or 2B. Figure 6.1.3.5.b shows the 76 data points and the cut-off values represented by horizontal lines at MRD_{50} = 10 mg/ml and MRD_{50} =2 mg/ml.

An assay of this type could be very useful as a screening assay for products which are very mild. Any products or materials which are below the cut-off might be excluded because of their toxicity, but materials scoring above the cut-off could be reasonably assured to be non-irritant.

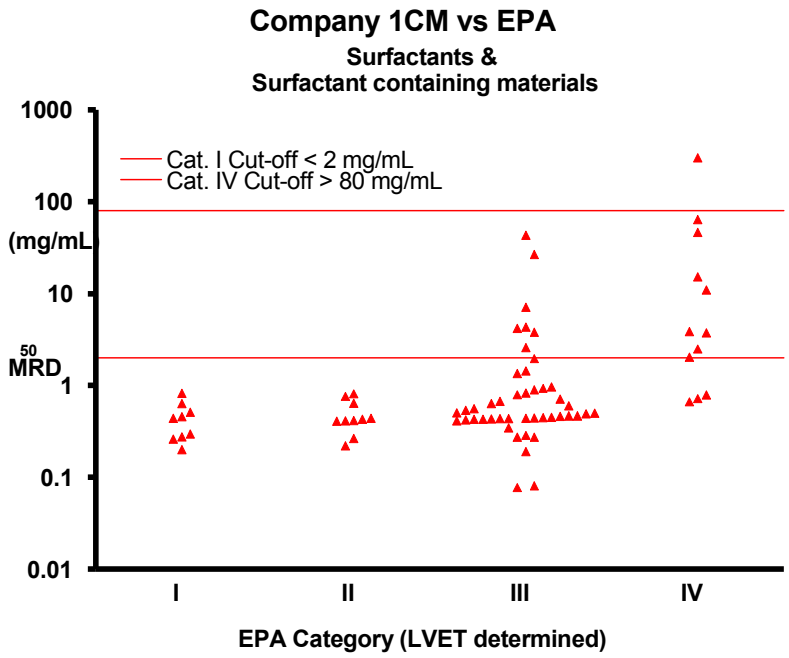

Figure 6.1.3.5.c Results from 76 materials from Company 1 related to EPA classification. For the four materials in Table 6.1.3.5.c which had 2 independent animal tests only the most irritating animal scores were used in this figure.

For the analysis according to the EPA scoring system (Figure 6.1.3.5.c) we again began with an attempt to set cut-offs which balanced over predictions and under predictions in a conservative manner that kept under predictions to a minimum. However, the grouping of the data points (Figure 6.1.3.5.c) showed that there was no obvious difference in MRD_{50} values between Category I, II and a majority of the Category III materials. Therefore, we focused just on separating the highly irritating materials from a portion of the Category III materials and then from the Category IV materials. A conservative cut-off value between the least irritating of the Category III materials and the Category IV materials seemed to be at MRD_{50} >80 mg/ml. This is slightly above the highest Category III value and in fact covers a significant number of the Category IV materials. Thus a material with an MRD_{50} score >80 mg/ml would be considered to have Category IV (there is only one example in this data set; all the other Category IV's would be overpredicted). Materials with MRD_{50} values between 80 mg/ml and 2 mg/ml would be considered to be Category III materials. Materials with MRD_{50} values ≤2 mg/ml would be considered to be Category I materials unless there was a second *in vitro* test which validated to categorize these materials into a I, II, or III category. Figure 6.1.3.5.c shows the 76 data points and the cut-off values represented by horizontal lines at MRD_{50} = 80 mg/ml and MRD_{50} =2 mg/ml.

For all three of the scoring systems the data distribution of the Company # 1 materials seems slightly different from that observed with data from the EC/HO, CTFA and COLIPA validation studies (Section 6.1.3.4). In general, the distribution is shifted to the milder end of the irritation scale. This is likely due to a combination of two phenomena: 1)

the Company #1 data set contains many marketed or prototype products which would likely be in the lower range of irritancy than materials in a validation study which are chosen to cover a wide range of the irritancy spectrum, and 2) the Company # 1 data have hazard classifications assigned based to ocular changes observed during a Low Volume Eye Test assay (10 μl of test material instilled in the eye) as opposed to the use of the traditional Draize test (100 μl of test material instilled in the eye) for hazard category determination that was used in the validation studies. It is known that the LVET does give somewhat lower irritation scores than does the traditional Draize test, although the LVET response is still generally more severe than the human response (Freeberg, Nixon et al. 1986) (Cormier, Parker et al. 1996).

The results of using the suggested cut-offs for the EU, GHS and EPA scoring systems are depicted in the contingency tables 6.1.3.5.b, 6.1.3.5.c and 6.1.3.5.d, respectively.

Table 6.1.3.5.b Company 1 - Contingency table depicting the concordance and predictivity of the CM assay for EU hazard classifications when the cut-off values from Figure 6.1.3.5.a are applied. N = 76 materials.

Draize Determined EU Category	EU Category Predicted by CM						
	R41	R36	Not Classified	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
R41	9		0	9	100.0%	NA	0%
R36	8		0	8	100.0%	NA	0%
Not Classified	43		16	59	27.1%	72.9%	NA
Total	60		16	76	43.4%		
Predictivity	28.3%		100.0%				
Category Underpredicted	NA		0%				
Category Overpredicted	71.7%		NA				

Table 6.1.3.5.c Company 1 - Contingency table depicting the concordance and predictivity of the CM assay for GHS hazard classifications when the cut-off values from Figure 6.1.3.5.b are applied. N = 76 materials.

Draize Determined GHS Category	GHS Category Predicted By CM							
	1	2A	2B	No Label	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
1	8		0	0	8	100.0%	NA	0%
2A	14		0	0	14	0%	100.0%	0%
2B	15		1	0	16	6.3%	93.8%	0%
No Label	23		8	7	38	18.4%	81.6%	NA
Total	60		9	7	76	39.5%		
Predictivity	36.7%		11.1%	100.0%				
Category Underpredicted	NA		0%	0%				
Category Overpredicted	63.3%		88.9%	NA				

Table 6.1.3.5.d Company 1 - Contingency table depicting the concordance and predictivity of the CM assay for EPA hazard classifications when the cut-off values from Figure 6.1.3.5.c are applied. N = 76 materials.

Draize Determined EPA Category	EPA Category Predicted By CM							
	I	II	III	IV	Total	Concordance	Toxicity Overpredicted	Toxicity Underpredicted
I	9		0	0	9	100.0%	NA	0%
II	10		0	0	10	100.0%	NA	0%
III	38		7	0	45	15.6%	84.4%	0%
IV	3		8	1	12	8.3%	91.7%	NA
Total	60		15	1	76	35.5%		
Predictivity	31.7%		46.7%	100.0%				
Category Underpredicted	NA		0%	0%				
Category Overpredicted	68.3%		53.3%	NA				

Table 6.1.3.5.e Company 1 - Concordance table for surfactant and surfactant containing materials for severe irritants versus the rest and non-irritants versus the rest for Company # 1 study. Cut-off values from Figures 6.1.3.5.a, 6.1.3.5.b, and 6.1.3.5.c were applied. N = 76 materials.

	Vs Rest													
	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
EU Severe irritants	33%	25/76	100%	9/9	24%	16/67	15%	9/60	100%	16/16	76%	51/67	0%	0/9
EU non-irritants	43%	33/76	100%	17/17	27%	16/59	28%	17/60	100%	16/16	73%	43/59	0%	0/17
GHS Severe irritants	32%	24/76	100%	8/8	24%	16/68	13%	8/60	100%	16/16	76%	52/68	0%	0/8
GHS non-irritants	59%	45/76	100%	38/38	18%	7/38	55%	38/69	100%	7/7	82%	31/38	0%	0/38
EPA Severe irritants	33%	25/76	100%	9/9	24%	16/67	15%	9/60	100%	16/16	76%	51/67	0%	0/9
EPA non-irritants	86%	65/76	100%	64/64	8%	1/12	85%	64/75	100%	1/1	92%	11/12	0%	0/64

An analysis of the ability of the CM assay to test surfactants and surfactant-based materials and separate severe irritants from “the rest”, and non-irritants from “the rest” is shown in Table 6.1.3.5.e. The CM test seemed to perform best when identifying non-irritants – the negative predictive value was 100% for all the analysis conditions with only the EPA non-irritant analysis being somewhat suspect since there was only one negative material in the data set.

6.1.4 Compilation of data on predictive capacity of the test method

6.1.4.1 Description & rational for the prediction model(s) applied and statistical approaches used

EC/HO (Balls, Botham et al. 1995)

The original analysis of this study was conducted independently by BIBRA (Lovell) with guidance from the management team. There was no prediction model proposed before the start of the study. Post hoc analysis was conducted mainly by determining the Pearson product moment correlation between the *in vitro* score (MRD_{50}) and the Draize Modified Maximum Average Score (MMAS). A second comparison was made to chemicals having MMAS >59 and <59 and chemicals having MMAS >25 and <25 using the Mann-Whitney test. Subgroups of chemicals (solids, liquids, surfactants) were also analyzed. The subsets into which the chemicals fell were determined by the Chemicals Selection Committee. No special measures were taken to account for the uncertain solubility of some of the materials observed by some of the laboratories. The statistical analysis on these data examined the between-laboratory reproducibility of the method and the relationship between the SM with the transwell MRD_{50} values and *in vivo* data (presented as the MMAS score). After the analysis was concluded, it was determined that the CM with the transwell was not a valid assay for predicting the MMAS of either the universe of chemicals or any of the subgroups. In fact, none of the assays participating in the study were found to be valid for predicting the MMAS.

In the analysis conducted in this BRD, the CM with the transwell data were investigated for their ability to predict EU, GHS, or EPA hazard classifications. It appears that many R41, GHS Category 1, and EPA Category I materials – especially surfactants - can be distinguished from less irritating materials by using a cut-off or MRD_{50} < 2 mg/ml. The final sample size (11 materials) was very small, however.

CTFA Phase III (Gettings, Lordo et al. 1996)

The original analysis for this study was conducted by an independent contractor (Battelle Laboratories) using the SAS system for personal computers. The system performed a statistical analysis in three areas. The first area examined was the distributional characterization of the Draize and *in vitro* results. Next, the data obtained was analyzed for concordance with the Draize MAS (although only continuously distributed endpoints were measured). Regressional modeling of *in vitro* vs. MAS data was the ultimate method of analysis for this study since the main goal of this work was to predict Draize results based on *in vitro* information. The variability of the Draize results was estimated during this study to enable evaluation of the degree of separation between pairs of test materials. No attempt was made at the conclusion of the study to determine if any of the assays were actually valid for the purpose of predicting Draize MAS scores.

However, the CM was one of the better performing assays for these surfactants and surfactant-based formulations.

For this BRD, the relationship between MRD_{50} values and EU, GHS, and EPA hazard classifications was investigated. It was found that a cut-off value of >80 mg/ml might identify some Not Classified or No Category materials. However, there was much overlap between the other hazard categories.

COLIPA (Brantom, Bruner et al. 1997)

An aim of this study was the development and validation of a prediction model (PM) for each assay. Parallel *in vitro* and *in vivo* data generated in house, much of which had previously been voluntary submitted to IRAG (US Interagency Regulatory Alternatives Group), were evaluated to create a semi-logarithmic plot of the MRD_{50} vs. Draize MMAS for 133 surfactant products. A mathematical description of this plot served as the prediction model. BIBRA International performed a quality check on the data submitted by independent laboratories to determine if data generated in the study matched the prediction model. The CM PM for this study was a 3 parameter logistic model.

The formula proposed to relate *in vitro* MRD_{50} volumes and *in vivo* MMAS scores was:

$$MMAS = \frac{A}{1 + eB^* (\log 10\ MRD_{50} - G)}$$

where A =148.0, B =1.813 and G =2.329

In addition, an *in vitro* cut-off score was proposed to provide a means of assessing the *in vitro* test data independent of any assumptions regarding the fit of the statistical methods to historical data. After the analysis, the CM using the proposed prediction model was not considered to be valid for the purpose of predicting Draize MAS scores.

In this BRD, a comparison was made to EU, GHS, and EPA hazard categories. It was clear from observations of the scatter plots that it would not be possible to set any cut-offs that would correctly identify test materials of intermediate, i.e. R36 or Category 2A or 2B materials. Therefore, an attempt was made to determine cut-off values that would identify either the more highly irritating materials or the non-irritating materials. Although It did not seem to be possible to adequately separate the non-irritating materials from the mid-level materials, a prediction model is proposed where, for surfactants, an MRD_{50} >10 mg/ml (for the GHS and EU systems), or an MRD_{50} > 80 mg/ml (for the EPA system) can identify Not Classified, No Category, or category IV materials, and an MRD_{50} <2 mg/ml can identify R41 or Category 1 materials. The prediction model was empirically determined by applying various cut-off values and calculating the number of either highly toxic or non-toxic substances that would be either under or overpredicted. A conservative approach was applied that attempted to minimize the number of under predictions.

Combined Studies

Data for surfactants and surfactant-containing products from the three studies mentioned above were combined to provide a much larger data set for analysis. This was done with knowledge of the fact that some of the materials used in the studies might be identical; however, it is not certain that the materials are identical. It was clear from observations of the scatter plots that it would not be possible to set any cut-offs that would correctly identify test materials of intermediate, i.e. R36 or Category 2A or 2B materials. Therefore, an attempt was made to determine cut-off values that would identify either the more highly irritating materials or the non-irritating materials. It was found that a prediction model could be proposed where, for surfactants and surfactant-containing materials, an MRD_{50} >10 mg/ml (for the GHS and EU systems), or an MRD_{50} >80 mg/ml (for the EPA system) can identify Not Classified or No Category materials, and an MRD_{50} <2 mg/ml can identify R41 or Category 1 or EPA Category IV materials.

Company # 1 Unpublished

The Company # 1 unpublished data analysis was slightly different than the preceding studies in that the Low Volume Eye Test was used for a standard rather than the traditional Draize test. It appeared from the scatter plots (Figures 6.1.3.5.a, 6.1.3.5.b and 6.1.3.5.c) that it was impossible to determine a cut-off value that would separate either the Category 1, 2A, or 2B categories from each other, the R41 and R36 categories from each other, or the EPA Category I, II, or III from each other. Therefore, an attempt was made just to identify a cut-off that would separate a high proportion of the No Label or No Category materials from the irritating materials. A conservative approach was taken where we sought to have as few under predictions as possible. The data in these studies suggested that – for surfactants and surfactant-containing products - a prediction model of MRD_{50} ≥10 mg/ml can be used to identify EU Not Classified or GHS No Category materials, and a prediction model of MRD_{50} ≥ 80 mg/ml can be used to identify EPA Category IV materials. These cut-offs were chosen to be somewhat higher than any of the MRD_{50} scores obtained by the R36, 2B or category III materials to give a higher probability that subsequent testing of a larger number of materials would not uncover any R36 or Category 2B or EPA category III materials that exceeded these limits.

However, more aggressive cut-off values of MRD_{50} ≥ 1 mg/ml for identifying EU No Label materials and GHS No Category materials were also suggested even though these values are very close to the less toxic R36 and Category 2B materials.

The distribution of the Company # 1 unpublished data points appears different than what was found for the other sets of data presented. This is likely due to the fact that the EU, GHS and EPA hazard categories were determined using the LVET rather than the traditional Draize assay. It is known that the LVET assay is gives somewhat lower classifications than does the Draize test, although both still over predict the human response.

6.1.4.2 Description of performance compared to reference and eventually, to the human situation for each study

EC/HO (Balls, Botham et al. 1995)

An analysis of the EC/HO data indicated that cut-offs could be applied which would have value in separating several of the hazard classes. The prediction model proposed was MRD_{50} > 80 mg/mL = Not Classified, No Label, and Category IV and MRD_{50} < 2 mg/mL = R41, Category 1, and Category I for the EU, GHS, and EPA classification systems respectively. The results of applying these prediction models to the data are shown in Tables 6.1.3.1.c, 6.1.3.1.d, and 6.1.3.1.e. The analysis of the surfactant materials only lead to the same conclusion for the prediction model as stated above. The results of applying these prediction models to the surfactant data are shown in Tables 6.1.3.1.g, 6.1.3.1.h, and 6.1.3.1.i.

CTFA Phase III (Gettings, Lordo et al. 1996)

An analysis of the CTFA data indicated that cut-offs could be applied which would have value in separating several of the hazard classes. The prediction model proposed was MRD_{50} > 80 mg/mL = Not Classified, No Label, and Category IV and MRD_{50} < 2 mg/mL = R41, Category 1, and Category I for the EU, GHS, and EPA classification systems respectively. The results of applying these prediction models to the data are shown in Tables 6.1.3.2.b, 6.1.3.2.c, and 6.1.3.2.d.

COLIPA (Brantom, Bruner et al. 1997)

An analysis of the COLIPA data indicated that cut-offs could be applied which would have value in separating several of the hazard classes. The prediction model proposed was MRD_{50} > 80 mg/mL = Not Classified, No Label, and Category IV and MRD_{50} < 2 mg/mL = R41, Category 1, and Category I for the EU, GHS, and EPA classification systems respectively. The results of applying these prediction models to the data are shown in Tables 6.1.3.3.b, 6.1.3.3.c, and 6.1.3.3.d. The prediction model proposed for the surfactant and surfactant containing materials was MRD_{50} > 10 mg/mL = Not Classified, and No Label, MRD_{50} > 80 mg/mL = Category IV and MRD_{50} < 2 mg/mL = R41, Category 1, and Category I for the EU, GHS, and EPA classification systems respectively. The results of applying these prediction models to the surfactant and surfactant containing materials are shown in Tables 6.1.3.3.e, 6.1.3.3.f, and 6.1.3.3.g.

Combined Studies

An analysis of the data for surfactant and surfactant-containing materials from the combined studies (EC/HO, CTFA Phase III and COLIPA) indicated that cut-offs could be applied which would have value in separating several of the hazard classes. The

prediction model proposed was MRD_{50} >10 mg/ml = Not Classified and No Label, and MRD_{50} <2 mg/ml = R41 and Category 1for the EU and GHS classification system respectively. The proposed prediction model for the EPA classification system was MRD_{50} > 80 mg/mL = Category IV and MRD_{50} < 2mg/mL = Category I. The results of applying these prediction models to the data are shown in the following contingency tables 6.1.3.4.b, 6.1.3.4.c, and 6.1.3.4.d

Because these contingency tables for the combined studies are based on the results with 53 test materials, they are a much better representation of the performance of the CM test than contingency tables based on any of the individual studies alone where far fewer materials were tested.

Company 1 Unpublished

Applying the prediction model of MRD_{50} > 10 mg/ml to the actual data in Figure 6.1.3.5.b reveals that 7 of the 38 materials (18.4%) identified as No Label by the animal test would be identified as No Label by the CM. This means that 81.6% of the animal designated No Category materials would have to pass on to a second level test.

If a more aggressive prediction model is used, e.g. MRD_{50} >1 is applied to Figure 6.1.3.5.b, then there would be significant improvement of the performance. Eighteen of the 38 materials (47.4%) identified as No Category by the animal test would be identified as No Label by the CM. However, one 2B material would now be under predicted by the model. Again, the remaining materials, 38 in this case, would have to pass on to a second level test in order to be correctly categorized.

An analysis of the data for surfactant and surfactant-containing materials from the Company # 1 study indicated that cut-offs could be applied which would have value in separating several of the hazard classes. The prediction model proposed was MRD_{50} >2 mg/ml = Not Classified and MRD_{50} <2 mg/ml = R41 for the EU classification system. The prediction model proposed was MRD_{50} >10 mg/ml = No Label and MRD_{50} <2 mg/ml = Category 1 for the GHS classification system. The prediction model proposed was MRD_{50} >80 mg/ml = Category IV and MRD_{50} <2 mg/ml = Category I for the EPA classification system. The results of applying these prediction models to the data are shown in the following contingency tables 6.1.3.5.b, 6.1.3.5.c, and 6.1.3.5.d

The above contingency tables indicate that a prediction model can be developed which allows little under prediction of the hazard categories of any of the classification systems, but it does not allow discrimination between the higher categories of irritancy.

A further analysis of the ability of the CM assay to categorize materials into two binary classification systems (severe irritants vs. the rest and non-irritants vs. the rest) is shown in Table 6.1.3.5.e.

6.1.4.3 Discussions

Description of the limitations of the test method (applicability domain based on the results of the data compilation)

Indications of the limitations of the test method first appeared in the EC/HO study. As seen in Tables 6.1.3.1.a and 6.1.3.1.b, R41 and Category I substances had a very wide range of CM MRD_{50} scores, ranging from approximately 0.3 mg/ml to nearly 100 mg/ml. The higher MRD_{50} values in this range overlapped significantly the No Label or No Category substances. These high scoring materials were of mixed chemistry – some with high pH (10 % sodium hydroxide), others of mixed chemistry (cyclohexanol, imidazole, pyridine, and 2,5-dimethyllohexanediol). By limiting the analysis to surfactants, Figures 6.1.3.1.d, 6.1.3.1.e, and 6.1.3.1.f show that most of the R41 and R36 with high MRD_{50} values as well as the Category 1 and 2A substances with high MRD_{50}'s were removed such that almost all the R41 and Category 1 materials had MRD_{50}'s <2 mg/ml. This was good evidence that the CM methodology was not very accurate for a broad range of chemical classes.

Subsequently the CTFA Phase III study showed that the 10 R41 surfactants or surfactant-containing materials and 8 Category 1 materials also had MRD_{50} values in the lower range, generally < 3 mg/ml.

Finally in the COLIPA there were several 2A and R36 non-surfactants which had relatively non-toxic MRD_{50} values. When the analysis was limited to surfactants, the separation by the cut off values seemed much clearer.

Thus, there is relatively strong evidence that the applicability domain of the CM assay is surfactants and surfactant-containing formulations. Although the data from Company # 1 do not add evidence against testing non-surfactant materials with the CM (since all the Company # 1 materials were surfactants) the studies do support the conclusion that non-irritant surfactant materials can be identified by their relatively high MRD_{50} values.

An additional part of the applicability domain was not really proven by the data of these studies but instead by the physical constraints of the test system, i.e. a certain number of materials were automatically excluded by the limitations of the machine itself. Any materials which are not completely water-soluble cannot be properly delivered by the pumping mechanisms of the machine and hence cannot even enter into the testing phase.

Possible rational(s) for differences observed

Apparent differences in the performance of the assay over the three major studies reviewed above are likely due to differences in the spectrum of surfactants or surfactant-containing materials that were tested, rather than differences in the CM's response to surfactants in general. This spectrum had an effect in the individual studies because so

few materials were tested in each one. A small variation in the irritation spectrum of the materials would likely make it appear that the prediction model might be different among the studies, yet when the studies were combined it was obvious that the hazard categories of the surfactants could be reasonably identified (at least the most toxic and least toxic of them) by high and low cut-off values.

6.2 Additional studies where raw data are not available

6.2.1 Relevant information for each study where raw data are not available

Table 6.2.1 Table presenting the relevant information for each study where raw data are not available

Studies	No. of chemicals	No. of products	Coded	No. of labs	No. of exps	Data format (raw, summary)	Chemical classes covered	Ranges of toxicity covered	Physico-chemical properties covered
Company # 1[1](Bruner, Miller et al. 1991)	4	13	Unknown	1	1	Summary	Household and personal cleaning products and ingredients	MAS 0-44.7	Liquids and granular materials
Several companies[2] (Bagley, Bruner et al. 1992)	12	20	Unknown	2	1	Summary	Cleaning product ingredients- almost all surfactants	MAS 0.3–43 for analysis; 0–44.7 total	See Annex A and B
Company # 2[3] (Catroux, Rougier et al. 1993)	21	32	Unknown	1	3	Summary	Surfactants and formulations	MAS 1.67-54	Liquids and water-soluble materials

[1] Performed using the silicon microphysiometer protocol with a 320-second exposures and normal human epidermal keratinocytes (see (Bruner, Miller et al. 1991)

[2] Performed using the silicon microphysiometer with glass coverslip standard protocol (as in 3) and the silicon microphysiometer with transwells with a 500-second exposure cycle

[3] Performed using the silicon microphysiometer instrument (transwell) with the 400-second exposure cycle

6.2.2 Relevant information for *in vivo* study where raw data are available

Table 6.2.2 Table presenting the *in vivo* reference data available

Studies	Species and protocols used as reference data	Sources of information	No. of labs	Quality of the reference data	Format of data available
Company # 1[1] (Bruner, Miller et al. 1991)	Albino rabbits using Low Volume Eye Test. No anesthesia. Whether every chemical was tested this way could not be determined	Previous testing	multiple	Not stated	Final LVET score only; 3-12 rabbits per study
Several companies[2] (Bagley, Bruner et al. 1992)	Albino rabbits using three different protocols. 1) Draize protocol with ocular anesthesia. 2) Draize protocol without ocular anesthesia, and 3) LVET protocol without anesthesia.	Not stated	At least 5.	Not stated	Single MAS value per chemcial
Company # 2[3] (Catroux, Rougier et al. 1993)	Rabbits tested using the Journal Officiel, 24 October 1984	Internal data bank	Unknown	Not stated	Summary MAS and MRD_{50} values in publication. More complete data for 20 of the 53 materials obtained from Company 2 for this BRD

[1] Performed using the silicon microphysiometer protocol with a 320-second exposures and normal human epidermal keratinocytes (see (Bruner, Miller et al. 1991)

[2] Performed using the silicon microphysiometer with glass coverslip standard protocol and the silicon microphysiometer with transwells with a 500-second exposure cycle

[3] Performed using the silicon microphysiometer instrument (transwell) with the 400-second exposure cycle

6.2.3. Brief description of the studies without available raw data

There were three studies which did not have raw data available. A comparison among the studies is somewhat difficult because different *in vitro* protocols were used in each study.

6.2.3.1 Company # 1 study (Bruner, Miller et al., 1991)

This study used the original SM instrument with normal human epidermal keratinocytes grown on a glass coverslip. Exposure was for approximately 300 seconds per dose. Seventeen materials were tested, 3 pure surfactants and 14 surfactant-containing formulations. LVET MAS scores were provided for each test substance along with the SEM and the number of rabbits used in the study (3-12).

6.2.3.2 Multiple lab study (Bagley, Brunner et al., 1992)

This study used two different SM protocols, one with L929 cells plated on glass coverslips and exposed for approximately 500 sec to each dose of test substance. The second protocol used the SM fitted with a transwell. L929 cells were grown on the membrane which formed the base of the transwell, and they were exposed to each dose for approximately 500 seconds, as in the first protocol. The purpose for using the two protocols was to compare data from the original SM with data from a new machine configuration that was to form the basis of a new commercial instrument, the CM.

Although there were 32 test materials studied in the project, only 17 were used for *in vivo*/*in vitro* comparisons since they all were tested with a traditional Draize test. The other materials were not compared because they were tested with the LVET. The materials were mostly surfactants and surfactant-containing formulations although a few were other ingredients often found in personal care or household cleaning products.

6.2.3.3 Company # 2 study (Catroux, Rougier et al. 1993)

This study used a SM fitted with a transwell chamber. L929 cells were grown in the chamber and exposed to test material for approximately 400 seconds per dose. Fifty-three materials were tested – 21 surfactants and 32 surfactant-based formulations. Draize tests were conducted on each material according to French legislation (Journal Officiel, 24 October 1984). Only summary Draize MAS scores were given in the manuscript, but more detailed mean date for rabbit groups was obtained from Company # 2 for some of the materials. However, since these additional data were group averages only and not individual animal scores, EH or GHS hazard categories could not be calculated.

6.2.4 Compilation of data on predictive capacity of the test method

6.2.4.1 Description & rational for the Prediction Model(s) applied and statistical approaches used

Company # 1 study

To analyze the performance of the SM, a semilog scatter plot was constructed between the mean and SEM of the MRD_{50} (in g/ml) and the LVET MAS plus SEM. Spearman's rank correlation test was then used to compare the relationship between the two values. No prediction model was proposed before the study was started.

Multiple lab study

Similar to the Company # 1 study, semilog scatter plots were constructed from the MRD_{50}'s from each SM protocol and LVET MAS values. Pearson and Spearman correlation coefficients were then generated to determine the strength of the relationship between the SM values and the LVET data. No prediction model was proposed before the study was started.

Company # 2 study

Several types of analyses were conducted to describe the performance of the SM instrument. One method was to construct a semilog scatter plot of MAS scores and MRD_{50} values for each of the test materials. A second method was to divide the MAS scale in to three classes (0-20, 20-40, and 40-60) and the log MRD_{50} into three similar classes (2.5-3.5, 3.5-5.0, and 5.0-6.5). Data points falling within each corresponding "box" were considered to be correct predictions. Data points falling outside their respective "boxes" were considered to be under or over predictions as appropriate. No prediction model was proposed before the study was started.

6.2.4.2 Description of performance compared to reference and eventually, to the human situation for each study

Company # 1 study

The SM data were compared to the LVET MAS scores on a semilog plot. Spearman's rank correlation test showed an r value of 0.89 between the two values indicating a relatively good predictive value of the SM method for the rabbit score. Only one material – Hard Surface Cleaner B – was a clear outlier, and it was under predicted.

Multiple lab study

Both the SM with a glass coverslip and the SM with a transwell chamber were compared to the LVET MAS for 17 substances. For the glass coverslip method the

Pearson/Spearman correlation coefficients were -0.71 and -0.65, respectively. For the SM with transwell chamber the coefficients were -0.72 and -0.61, respectively. Thus both methods predicted the animal score reasonably well. An examination of the scatterplots shows that MAS values <10 had log $1/MRD_{50}$ values which extended over at least two logs. The curve then rose steeply over the next two logs to the maximum tested LVET scores of slightly less than 50.

Company # 2 study

Comparisons were first made directly between the MAS score and the log MRD_{50}. Pearson and Spearman coefficients were calculated for the 53 test materials and were found to be 0.91 and 0.89 respectively. This indicates a very good predictive capacity of the SM for the Draize MAS scores up to about MAS=54.

A second analysis was conducted by dividing the scores into three classes each. This analysis showed that there were 8 false positives and no false negatives. This type of analysis is not very common since there is little chance for a correct prediction at points where “correct” boxes are adjacent. This is because the boxes only touch at a corner (essentially a point) and any scatter of points at all would create many outliers. None the less, this analysis indicates that there is quite a good predictive power to the SM assay as used here.

Several types of analyses were conducted to describe the performance of the SM instrument. One method was to construct a semilog scatter plot of MAS scores and MRD_{50} values for each of the test materials. A second method was to divide the MAS scale in to three classes (0-20, 20-40, and 40-60) and the log MRD_{50} into three similar classes (2.5-3.5, 3.5-5.0, and 5.0-6.5). Data points falling within each corresponding “box” were considered to be correct predictions. Data points falling outside their respective “boxes” were considered to be under or over predictions as appropriate.

6.2.4.3 Discussions

All three of the studies used similar test substances – almost all surfactants or surfactant-containing formulations. Even though the protocols varied somewhat – mouse cells versus human cells, cells grown on cover slips versus cells grown in transwell chambers, and different exposure times all studies showed good prediction of the ocular irritation level (as measured by the MAS score) of the test substances.

6.3 Attempt to combine the data using weigh-of-evidence approaches

All three studies reported here were very similar in their construction, even though details of their protocols varied. Each tested very similar materials (surfactants and surfactant containing formulations). Only one test (Bagley, Bruner et al. 1992) was reported as being conducted on coded materials, but the automated data recording attributes of the SM tend to eliminate most bias that might occur by knowing the identity of

the test material. Since the reported Pearson's correlation coefficient for all of the four methods was >0.82 (individual r values were 0.97, 0.82, 0.86, and 0.91) the three studies support each other very well and give weight to the conclusion from these studies that the SM (using any of a number of different protocols) predicts the rabbit MAS very well, between 0 and ~50.

7. Applicability Domain (Module 6)

Much of the definition of the applicability domain has been influenced by the physical constraints imposed by the SM or CM machine itself, or by in-house studies that have not been reported in the open literature. Although some of the information shown in this BRD supports the existing feeling that the domain of the SM or CM is for surfactants and water soluble surfactant-containing products, there is no overwhelming data to support this point.

Clearly the physical properties of the machine which require it to expose the cells by pumping test material through a small diameter tube, and then wash the cells by pumping fresh media across the cells and out the chamber through another small diameter tube, dictate that no solids or suspensions be used. Materials of this physical state would tend to clog the machine or not be washed out once they had reached the exposure chamber. Thus test substances should be limited to water-soluble materials.

Personal communication with users of the CM over the last decade indicate that their experience is that most non-surfactant substances are not well predicted by the CM. Conversely they feel that surfactants and surfactant containing materials are well predicted. There are not strong data in this BRD to support the view that non-surfactants are not well predicted since very few non-surfactant materials were tested. Only the rather poor predictive results from the EC/HO study which used many non-surfactant materials would support this view.

The general class of materials (and the irritancy level) that does seem to be reasonably predicted is mild surfactants and surfactant-containing formulations as shown by the COLIPA study, the CTFA study, and the internal Company # 1 data.

This highlights an important role for a test like the CM. It can be used to immediately identify very mild surfactant-containing materials which may be a useful feature for a cosmetics or personal care product company that desires to produce products at the very low end of the irritancy scale. If the material is more irritating than the cut-off level chosen for the very mild products, a second type of *in vitro* test having a more robust nature, e.g a three dimensional tissue or an ex vivo eye model, could be used to properly classify the material in one of the higher hazard classification levels.

8. Supporting materials

8.1 Relevant publications, other scientific reports and reviews

Araki, K., Y. Ohahsi, S. Knoshita, K. Hayahsi, Y. Kuwayama and Y. Tano (1994). "Epithelial Wound Healing in the Denervated Cornea." Current Eye Research **13**: 203-211.

Bagley, D. M., L. H. Bruner, O. De Silva, M. Cottin, K. A. O'Brien, M. Uttley, et al. (1992). "An Evaluation of Five Potential Alternatives In Vitro to the Rabbit Eye Irritation Test In Vivo." Toxicology In Vitro **6**(4): 275-284.

Baldwin, H. A., T. O. McDonald and C. H. Beasley (1973). "Slit examination of experimental animals eyes. II Grading scales and photographic evaluation of induced pathological conditions." J Society of Cosmetic Chemistry **24**: 181-195.

Balls, M., P. A. Botham, L. H. Bruner and H. Spielmann (1995). "The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test." Toxicology In Vitro **9**(6): 871-929.

Botham, P., R. Osborne, K. Atkinson, G. Carr, M. Cottin and R. G. Van Buskirk (1997). "IRAG Working Group 3. Cell Function-based Assays. Interagency Regulatory Alternatives Group." Food and Chemical Toxicology **35**(1): 67-77.

Brantom, P. G., L. H. Bruner, M. Chamberlain, O. De Silva, J. Dupuis, L. K. Earl, et al. (1997). "A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test." Toxicology In Vitro **11**: 141-179.

Bruner, L. H., D. J. Kain, D. A. Roberts and R. D. Parker (1991). "Evaluation of Seven In Vitro Alternatives for Ocular Saftey Testing." Fundam Appl Toxicol **17**(1): 136-149.

Bruner, L. H., K. M. Miller, J. C. Owicki, J. W. Parce and V. C. Muir (1991). "Testing ocular irritancy in vitro with the silicon microphysiometer." Toxicology In Vitro **5**: 277-284.

Butler, J. M. and B. R. Hammond (1980). "The Effects of Sensory Denervation on the Response of the Rabbit Eye to prostaglandin E1, bradykinin and substance P." British Journal of Pharmacology **69**: 495-502.

Catroux, P., A. Rougier, K. G. Dossou and M. Cottin (1993). "The Silicon Microphysiometer for Testing Ocular Toxicity In Vitro." Toxicol In Vitro **7**(4): 465-469.

Chan, P. K. and A. W. Hayes (1994). "Acute Toxicity and Eye Irritancy." In Principles and Methods of Toxciology, 3rd Edition (ed Hayes A.W.): 579-647.

Cormier, E. M., R. D. Parker, C. Henson, L. W. Cruze, A. K. Merritt, R. D. Bruce, et al. (1996). "Determination of the Intra- and Inter-Laboratory Reproducibility of the Low Volume Eye Test and its Statistical Relationship to the Draize Test." Reg. Tox. Pharmac. **23**: 156-161.

Dearman, R. J., M. Cumberbatch and I. Kimber (2003). "Cutaneous Cytokine Expression: Induction by Chemical Allergen and Paracrine Regulation." Journal of Toxicology-Cutaneous and Ocular Toxicology **22**: 69-86.

DeSousa, D. J., A. A. Rouse and W. J. Smolon (1984). "Statistical Consequences of Reducing the Number of Rabbits Utilized in Eye Irritation Testing: Data on 67 Petrochemicals." Toxicol Appl Pharmacol **76**: 234-242.

Dua, H. S., J. A. P. Gomes and A. Singh (1994). "Corneal Epithelium Wound Healing." British Journal of Ophthamology **78**: 401-408.

ECETOC (1992). "ECETOC Technical Report No 48."

ECETOC (1998). "ECETOC Technical Report No. 48 (2)."

Eurell, T. E., J. M. Sinn, P. A. Gerding and C. L. Alden (1991). "In Vitro Evaluation of Ocular Irritatns Using Corneal Protein Profiles." Toxicology and Applied Pharmacology **108**: 374-378.

Farquhar, M. G. and G. E. Palade (1963). "Junctional complexes in Various Epithelia. ." J Cell Biology **17**: 375-412.

Feder, P. I., R. A. Lordo, D. M. B. Dipasquale, M. Chudkowski, J. L. Demetrulias, K. L. Hintze, et al. (1991). "The CTFA Evaluation of Alternatives Program: An Evaluation of *In Vitro* Alternatives to the Draize Primary Eye Irritation Test (Phase I) Hydro-alcoholic Formulations; (Part 1) Statistical Methods." In Vitro Toxicology **4**(4): 231-246.

Freeberg, F. E., G. A. Nixon, P. J. Reer, J. E. Weaver, R. D. Bruce, J. F. Griffith, et al. (1986). "Human and Rabbit Eye Responses to Chemical Insult." Fund. Appl. Toxic. **7**: 626-634.

Gautheron, P., J. Giroux, M. Cottin, A. Audegond, A. Morilla, L. Mayordomo-Blanco, et al. (1994). "Interlaboratory Assessment of the Bovine Corneal Opacity and Permeability (BCOP) Assay." Toxicology In Vitro **8**(3): 381-392.

Gettings, S. D., L. C. Dipasquale, D. M. Bagley, P. L. Casterton, M. Chudkowski, R. D. Curren, et al. (1994). "The CTFA Evaluation of Aternatives Program: An Evaluation of *In Vitro* Alternatives to the Draize Primary Eye Irritation Test. (Phase II) Oil/Water Emulsions." Fund. Chem. Toxic. **32**(10): 943-976.

Gettings, S. D., R. A. Lordo, K. L. Hintze, D. M. Bagley, P. L. Casterton, M. Chudkowski, et al. (1996). "The CTFA Evaluation of Alternatives Program: an evaluation of in vitro alternatives to the Draize primary eye irritation test. (Phase III) surfactant-based formulations." Food Chem Toxicol **34**(1): 79-117.

Griffith, J. F., G. A. Nixon, R. D. Bruce, P. J. Reer and E. Bannan (1980). "Dose response studies with chemical irritants in albino rabbits as a basis for selecting optimum testing conditions for predicting hazard to the human eye." Toxicology and Applied Pharmacology **55**: 501-513.

Hackett, R. B. and T. O. McDonald (1994). "Mechanisms of Ocular Response to Irritants." Dermatotoxicology, 5th Editions, (ed Marzulli F.N, Maibach H.I.): 299-306.

Harbell, J., S. W. Koontz, R. W. Lewis, D. Lovell and D. Acosta (1997). "IRAG Working Group 4: Cell Cytotoxicity Assays." Food and Chemical Toxicology **35**: 79-126.

Harbell, J. W., R. Osborne, G. J. Carr and A. Peterson (1999). "Assessment of the Cytosensor™ Microphysiometer Assay in the COLIPA In Vitro Eye Irritation Validation Study." Toxicology In Vitro **13**: 313-323.

Hartung, T., S. Bremer, S. Casati, S. Coecke, R. Corvi, S. Fortaner, et al. (2004). "A Modular Approach to the ECVAM Principles on Test Validity " Atern Lab Anim **32**: 467-472.

Hogan, M. J. and L. E. Zimmerman (1962). "Ophthalmic Pathology: An Atlas and Textbook, 2nd edition." Philadelphia: W.B. Saunders.

Hubert, F. (1992). "The Eye (Rabbit/Human): Parameters to be Measured in the Field of Ocular Irritation." ATLA **20**: 476-479.

Jacobs, G. A. and M. A. Martens (1990). "Quantification of Eye Irritation Based Upon In Vitro Changes of Corneal Thickness." ATLA **17**: 255-262.

Jester, J. V., L. Li, A. Molai and J. K. Maurer (2001). "Extent of initial corneal injury as a basis for alternative eye irritation tests." Toxicol In Vitro **15**(2): 115-30.

Jester, J. V., W. M. Petroll, J. Bean, R. D. Parker, G. J. Carr, H. D. Cavanagh, et al. (1998). "Area and depth of surfactant-induced corneal injury predicts extent of subsequent ocular responses." Invest Ophthalmol Vis Sci **39**(13): 2610-25.

Katahira, J. H., H. Sugiyama, N. Inoue, Y. Horiguchi, M. Matsuda and N. Sugimoto (1997). "Clostridium Perfringens Enterotoxin Utilizes Two Structurally Related Membrane Proteins as Functional Receptors in Vivo." J Biological Chemistry **272**: 26652-26658.

Klyce, S. D. and R. W. Beuerman (1988). "Structure and Function of the Cornea." The Cornea: 3-23.

Lovell, D. "EC/HO International Study on Alternatives to the Draize Eye Test." BIBRA Project No. 1367/1.

Maurer, J. K. and R. D. Parker (1996). "Light Microscope Comparision of Surfactant Induced Eye Irritation i Rabbits and Rats at Three Hours and Recovery/day 35." Toxicologic Pathology **24**: 403-411.

Maurer, J. K., R. D. Parker and J. V. Jester (2002). "Extent of initial corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays." Regul Toxicol Pharmacol **36**(1): 106-117.

Maurer, J. L. and R. D. Parker (1996). "LIght Microscope Comparison of Surfactant-induced Eye Irritation in Rabbits and Rats at Three Hours and Recovery/day 35." Toxicologic Pathology **6**: 464-477.

McConnell, H. M., J. C. Owicki, J. W. Parce, D. L. Miller, G. T. Baxter, H. G. Wada, et al. (1992). "The Cytosensor Microphysiometer: Biological Applications of Silicon Technology." Science **257**(5078): 1906-12.

Parce, J. W., J. C. Owicki, K. M. Kercso, G. B. Sigal, H. G. Wada, V. C. Muir, et al. (1989). "Detection of Cell-affecting Agents with a Silicon Biosensor." Science **246**: 243-247.

Prinsen, M. K. (2006). "The Draize Eye Test and in vitro alternatives; a left-handed marriage?" Toxicol In Vitro **20**(1): 78-81.

Van Meer, G., W. van Hof and I. nan Genderen (1992). "Tight Junctions and Polarity of Lipids " Tight Junctions (ed Cereijido M.): 187-201.

Wilhelmus, K. R. (2001). "The Draize Eye Test." Surveys in Ophthalmology **45**: 493-515.