Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

20-Jul-2009

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/MONO(2009)9
Unclassified

SERIES ON TESTING AND ASSESSMENT
Number 105

REPORT ON BIOSTATISTICAL PERFORMANCE ASSESSMENT OF THE DRAFT TG 436 ACUTE TOXIC CLASS TESTING METHOD FOR ACUTE INHALATION TOXICITY

English - Or. English

JT03268107

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 105

REPORT ON BIOSTATISTICAL PERFORMANCE ASSESSMENT OF THE DRAFT TEST GUIDELINE 436 ACUTE TOXIC CLASS TESTING METHOD FOR ACUTE INHALATION TOXICITY

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among UNEP, ILO, FAO, WHO, UNIDO, UNITAR and OECD

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2009

Also published in the Series on Testing and Assessment:

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, revised 2006)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (2007)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the use of Multimedia Models for estimating overall Environmental Persistence and long range Transport in the context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances Which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (2009)*

No. 40, *Detailed Review Document on Classification in*

OECD Member Countries of Substances and Mixtures Which Cause Respiratory Tract Irritation and Corrosion (2003)

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (2008)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work Sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of emission estimation methods used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case study of pulp and paper and textile sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater*

Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60, *Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61, *Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62, *Final OECD Report of the Initial Work Towards the Validation of the Rat Hershberger Assay: Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63, *Guidance Document on the Definition of Residue (2006)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65, *OECD Report of the Initial Work Towards the Validation of the Rodent Utertrophic Assay - Phase 1 (2006)*

No. 66, *OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2. Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67, *Additional data supporting the Test Guideline on the Uterotrophic Bioassay in rodents (2007)*

No. 68, *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the follow up of this report (2006)*

No. 69, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70, *Report on the Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71, *Guidance Document on the Uterotrophic Bioassay - Procedure to Test for Antioestrogenicity (2007)*

No. 72, *Guidance Document on Pesticide Residue Analytical Methods (2007)*

No. 73, *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded Testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74, *Detailed Review Paper for Avian Two-generation Toxicity Testing (2007)*

No. 75, *Guidance Document on the Honey Bee (Apis Mellifera L.) Brood test Under Semi-field Conditions (2007)*

No. 76, *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 - Optimisation of the Test Protocol (2007)*

No. 77, *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 - Multi-chemical Interlaboratory Study (2007)*

No. 78, *Final Report of the Validation of the 21-day Fish Screening Assay for the Detection of Endocrine Active*

Substances. Phase 2: Testing Negative Substances (2007)

No. 79, *Validation Report of the Full Life-cycle Test with the Harpacticoid Copepods Nitocra Spinipes and Amphiascus Tenuiremis and the Calanoid Copepod Acartia Tonsa - Phase 1 (2007)*

No. 80, *Guidance on Grouping of Chemicals (2007)*

No. 81, *Summary Report of the Validation Peer Review for the Updated Test Guideline 407, and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up of this report (2007)*

No. 82, *Guidance Document on Amphibian Thyroid Histology (2007)*

No. 83, *Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 84, *Report on the Workshop on the Application of the GHS Classification Criteria to HPV Chemicals, 5-6 July Bern Switzerland (2007)*

No. 85, *Report of the Validation Peer Review for the Hershberger Bioassay, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 86, *Report of the OECD Validation of the Rodent Hershberger Bioassay: Phase 2: Testing of Androgen Agonists, Androgen Antagonists and a 5 α-Reductase Inhibitor in Dose Response Studies by Multiple Laboratories (2008)*

No. 87, *Report of the Ring Test and Statistical Analysis of Performance of the Guidance on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (Transformation/ Dissolution Protocol) (2008)*

No.88 *Workshop on Integrated Approaches to Testing and Assessment (2008)*

No.89 *Retrospective Performance Assessment of the Test*

Guideline 426 on Developmental Neurotoxicity (2008)

No.90 *Background Review Document on the Rodent Hershberger Bioassay (2008)*

No.91 *Report of the Validation of the Amphibian Metamorphosis Assay (Phase 3) (2008)*

No.92 *Report of the Validation Peer Review for the Amphibian Metamorphosis Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-Up of this Report (2008)*

No.93 *Report of the Validation of an Enhancement of OECD TG 211: Daphnia Magna Reproduction Test (2008)*

No.94 *Report of the Validation Peer Review for the 21-Day Fish Endocrine Screening Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2008)*

No.95 *Detailed Review Paper on Fish Life-Cycle Tests (2008)*

No.96 *Guidance Document on Magnitude of Pesticide Residues in Processed Commodities (2008)*

No.97 *Detailed Review Paper on the use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors (2008)*

No. 98 *Considerations Regarding Applicability of the Guidance on Transformation/Dissolution of Metals Compounds in Aqueous Media (Transformation/Dissolution Protocol) (2008)*

No. 99 *Comparison between OECD Test Guidelines and ISO Standards in the Areas of Ecotoxicology and Health Effects (2008)*

No.100 *Report of the Second Survey on Available Omics Tools (2009)*

No.101 *Report on the Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox (2009)*

No.102 *Guidance Document for using the OECD (Q)SAR*

Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals (2009)

No.103 *Detailed Review Paper on Transgenic Rodent Mutation Assays (2009)*

No.104 *Performance Assessment: Conparsion of 403 and CxT Protocols via Simulation and for Selected Real Data Sets (2009)*

No. 105 *Report on Biostatistical Performance Assessment of the draft TG 436 Acute Toxic Class Testing Method for Acute Inhalation Toxicity (2009)*

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 30 industrialised countries in North America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides and Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and the Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/ehs/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The participating organisations are FAO, ILO, OECD, UNEP, UNIDO, UNITAR and WHO. The World Bank and UNDP are observers. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/ehs/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division

2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80
E-mail: ehscont@oecd.org

FOREWORD

This document includes a Retrospective Biostatistical Performance Assessment of the "Acute Toxic Class Method", which the draft Test Guideline (TG) 436 is based upon. The development of this report was part of an acute inhalation package, also including the draft TG 436, the updated TG 403, the draft Guidance Document No. 39 and the draft Performance Assessment reports of TG 403 (C x t approach).

The project for developing this report was led by Germany and the work of the Performance Assessment Group (PAG) was initiated in February 2006 at the 1st acute inhalation expert meeting, held in Berlin (Germany). The Biostatistical Assessment Report was prepared by Matthias Greiner (BfR, Germany). The acute inhalation package was then discussed at a meeting of the expert group, held in Washington DC (United States) in November 2006. The results of the PAG were presented at the last acute inhalation expert meeting held in Washington DC in April 2008, where the report was thoroughly discussed. The meeting generally approved the biostatistical approach taken by the PAG and also endorsed that the PAG Report could be used as a proof-of-concept for the TG 436 Acute Toxic Class Method.

Comments on the draft report were requested from the WNT in November 2008. The final draft report was prepared based on only minor comments received in the commenting round, basically only a change in the wording not to include the word 'retrospective validation', but rather use "retrospective performance assessment". The Working Group of National Coordinators of the Test Guidelines Programme endorsed the draft report at its meeting held on 31 March- 2 April 2009.

This document is published on the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

Contact for further details:
Environment, Health and Safety Division
Environment Directorate
Organisation for Economic Co-Operation and Development
2, rue André Pascal
75775 Paris Cedex 16, France

Tel : 33-1-45-24-16-74
E.mail : env.edcontact@oecd.org

TABLE OF CONTENTS

Report on Biostatistical performance assessment of the Draft TG 436 acute toxic class testing method for acute inhalation toxicity

11 February 2009

ABSTRACT

The draft TG 436 (acute toxic class method) has been developed as an alternative testing method for acute inhalation toxicity for the classification and labeling of chemicals according to the Globally Harmonized System of Classification and Labeling of Chemicals (GHS). The OECD, on an expert consultation meeting held in February 2006 in Berlin, commissioned a study for a statistical performance assessment of this draft guideline and the fixed concentration procedure (draft TG 433) and initiated the Performance Assessment Group (PAG) to pursue this task. This draft report is based on a version dated November 7, 2006, in which results for both draft test guidelines were presented. The PAG has been informed that the UK withdrew the draft TG 433 from the work plan. Therefore, this document reports only the results pertaining to the draft TG 436. As regards the biometrical performance characterization conducted according to a study protocol developed by the PAG, no new results have been added since the version of November 7, 2006. However, the presentation of the approach and the discussion of the results have been modified in response to comments given in the position papers by the UK (June 5, 2007) and the US. Additions that directly address issues raised in the US Position paper are indicated by footnotes. The results of the retrospective performance assessment study, after discussion of the comments from the US Position papers, might be considered as proof-of-concept of the draft TG 436.

1. INTRODUCTION

Alternative methods for acute inhalation toxicity testing have been developed, which allow the classification and labeling (C&L) of chemical substances according to the Globally Harmonized Classification System (GHS) using a smaller number of animals compared to the Testing Guideline (TG) 403. The GHS uses five categories of decreasing toxicity level, labeled 1 to 5 and one unclassified category, labeled "5/unclassified". The latter category will be denoted with "6" in this report. The testing method considered here is the acute toxic class (ATC) method (OECD draft guideline TG 436, see Figure 1 where starting concentration 1mg/L is shown as an example). The biostatistical performance characterization of this method has been described [2]. This performance characterization was based on a set of hypothetical substances, which have been defined by their physical class (vapors, aerosols and gases), LC50 and slope value of the probit concentration-response function. Using a simulation approach, the probabilities of classification into each of the GHS toxicity classes and the expected numbers of animals and deaths were established. At the OECD Expert Meeting on Acute Inhalation Toxicity, 22-24 February 2006 in Berlin, a three-stage process for a re-assessment of the performance and validation status of TG 436 (and TG 433) was adopted. The three stages were (1) a review of the model concepts and assumptions, (2) development of a harmonized approach for a biostatistical performance characterization and (3) a reassessment of the performance using the harmonized biostatistical study protocol. To facilitate this process, the formation of Performance Assessment Group (PAG) was commissioned. This report describes the assessment results obtained in 2006 and discusses the findings in the light of the comments given in the Position Papers presented by the UK and the US. All references to the TG 436 refer to the version of February 24, 2006.

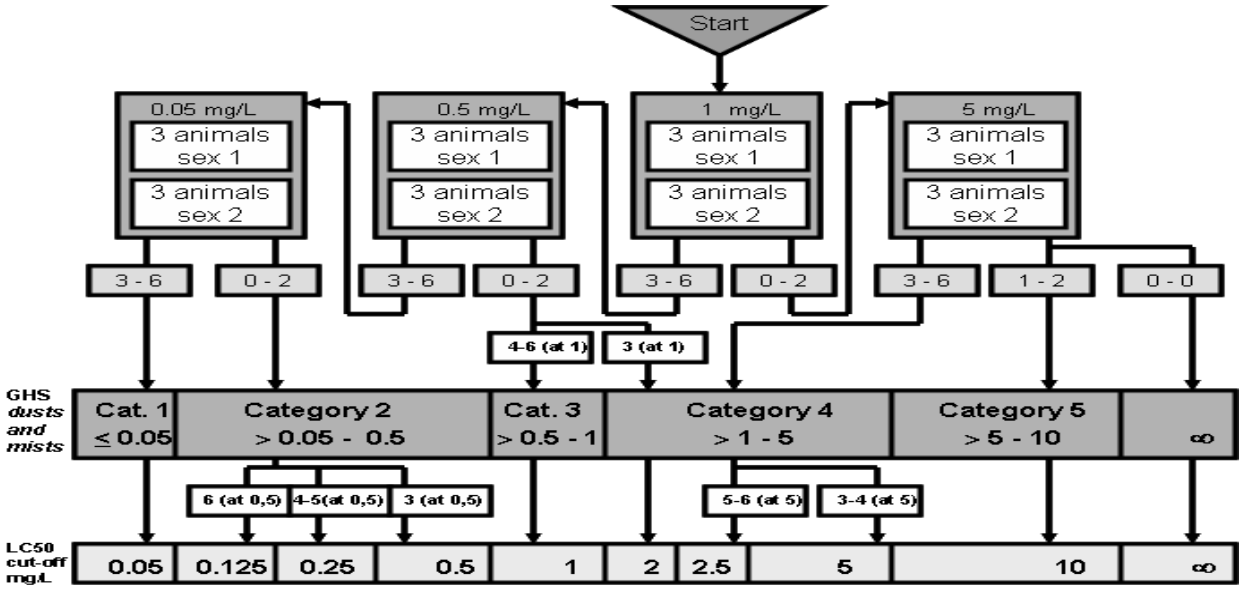

Figure 1. Classification scheme of the TG 436 for dusts and mists with starting concentration 1mg/L.

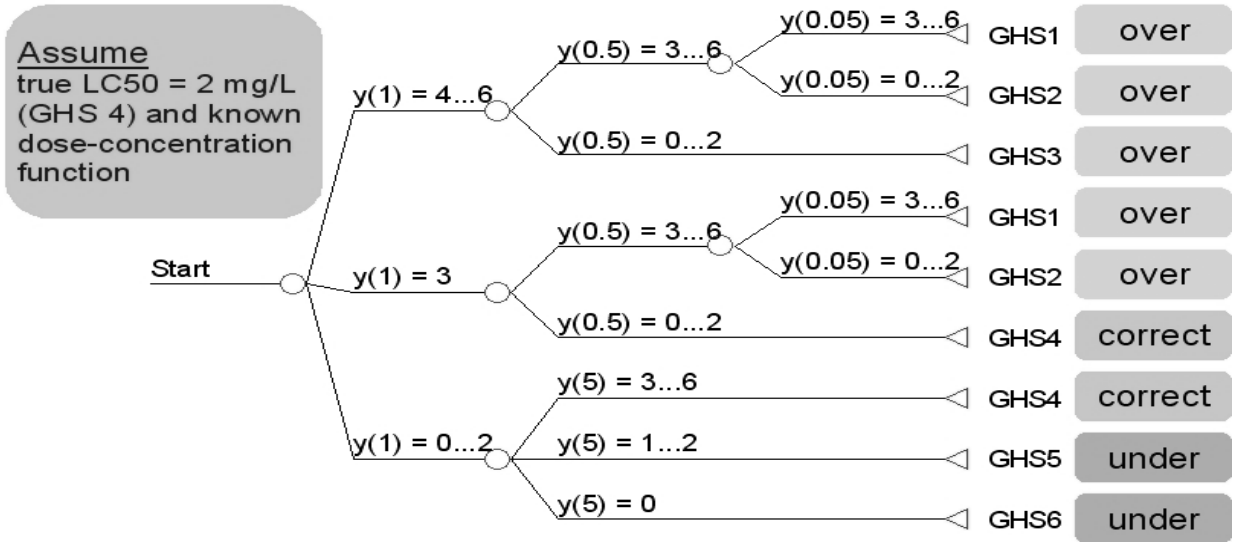

Figure 2. Same example as Fig. 1 represented as probability tree (see text for notation). Branches emerging from probability nodes (circles) can be parameterized using Binomial probabilities using a constant sample size of six and outcome probabilities derived from the probit concentration-response function. The classification probabilities are then obtained as sums of all those conditional (given an LC50 value within a given GHS class considered as "true" class) probabilities referring to all possible pathways from an assigned true class to each of the six different GHS classifications (triangles). The classification probabilities can be summarized in terms of correct classification, under- and overclassification (by one or more classes) as shown in the result tables.

2. MATERIAL AND METHODS

2.1 Description of the mathematical model

The theoretical mathematical model underlying the performance assessment of the TG 436 has been described [2]. Similar biostatistical performance assessments have been described for inhalation, oral and dermal toxicity [7-15]. TG 436 does not comprise a sighting study, only two outcomes for individual animals (death/moribund status or survival) are considered and $n = 6$ (three of each sex) animals are used at each concentration step. Let x_i, i = 1,…,4, denote four different concentrations of testing article. For example, when testing a dust or a mist, these concentrations are $x_1 = 0.05$ mg/L, $x_2 = 0.5$ mg/L, $x_3 = 1$ mg/L and $x_4 = 5$ mg/L (Figure 1). The foreseen experimental pathway and decision rules can be represented as probability tree (Figure 2) It is assumed that a particular substance has a probit concentration-response function with some LC50 value denoted by d, and some slope value, denoted by β. The probability of death for a single animal exposed at concentration xi is as given by

$$p_i = \Phi[\beta(\log x_i - \log d)] \qquad (1)$$

where Φ denotes the standard normal distribution function and the logarithm is to the base 10. Furthermore, let y_i denote the outcome of an experiment at concentration x_i, i.e. the number of dead or

moribund animals observed at concentration x_i. According to the experimentation rules of TG 436, decisions for the next dosing or final classification of the testing article are taken depending on the outcomes y_i. For example, if one animal dies at a starting concentration of 1mg/L in a test for dusts or mists (y_3=1), the decision is taken to continue testing at concentration 5mg/L (Figure 1). If then three animals die at concentration of 5mg/L (y_4=3), the decision is taken to classify the article into GHS class 4.

Consider the lower (r) and upper (s) limits of the outcome y_i that trigger the same decision. For the example above at the starting concentration x_3 = 1 mg/L, the decision "Continue at concentration 5 mg/L" is taken if the outcome y_3 falls into the [r,s]-interval [0,2] and else the decision "Continue at concentration 0.5 mg/L" is taken. At the concentration x_4 = 5 mg/L three possible decisions are taken (classification into GHS class 4, 5 or 6) depending on the outcome of y_4 falling into the [r,s]-intervals [3,6], [1,2] and [0,0], respectively. The probability (Pr) for each decision at the concentration x_i, given the rules for the [r,s]-interval, can be derived from the binomial probability distribution function with the form

$$\Pr(y \text{ is in the interval } [i,j]) = \sum_{y=r}^{s} \binom{n}{y} p_i^{y} (1-p_i)^{n-y}. \qquad (2)$$

Thus, the mathematical model is a precise representation of the experimentation rules (as shown at the example of Figure 1) as a probability tree (Figure 2). The probabilities of each branch can be given as shown in eq. (2). Each possible starting concentration is represented as a separate probability tree, which is evaluated in separate analyses. The probability of classifying an article into the correct GHS class or into a more toxic class (overclassification) or less toxic class (underclassification) class depends on the articles true LC50 and probit slope value. For example, a chemical with true LC50 value in the middle of a GHS class and large slope value will have high probability of correct classification while a chemical with LC50 value close to a GHS class boundary and/or low slope value will have a low probability of correct classification. Therefore, the use of the model for performance assessment requires that realistic distributions for LC50 and slope values are used. The parameterization of the model, which accounts for variability (actual variability present in chemicals to be tested) and uncertainty (due to limited sample size of empirical data) of the LC50 and probit slopes for real-world chemicals is described in section 2.2. The biological variability of the response to exposure concentrations is captured by the probabilistic simulation approach described in section 2.3.

2.2 Parameterization of the model

The model is used in a stochastic ("probabilistic") way. That means, for evaluation of the classification performance, distributions of LC50 and slope values are used rather than fixed values. The distributions of LC50 and slope values were defined using the following data sources.

LC50 values. The European data base for notification of new substances according to the Chemicals Act (by courtesy of Dr. Thiele, Federal Institute for Occupational Safety and Health, Germany; below referred to as the "data on new substances") has been used for modelling the LC50 values. The data set included variables that describe the processing identification number, the article identification number, the LC50 and 95% confidence interval of the LC50 (all metrics in mg/L), the probit slope and the physical state of the article with three different levels (G=gas, L=liquid, S=solid). The inclusion criterion for the data set was non-missing information on the LC50. The numbers of records and available LC 50 values that met the inclusion criteria were:

G 45 records 41 LC50 results
L 116 records 112 LC50 results
S 191 records 183 LC50 results

Only the LC50 values for solid articles were considered because this physical status could be matched with the definition of "dusts" used in the draft TG 436. Also different metrics for gases (mg/L in the data base and ppm in the TG 436 and GHS) made it impossible to use the data for gases. The molecular weight was not available from the data. If it was, conversion between ppm and mg/L for gases and mists would have introduced additional uncertainty (air pressure and temperature)[1]. Only for 6 records of probit slope values referred to solid articles The model that fitted the LC50 values for solids best was a mixture of distributions with two components. One component of the mixture represents tests of toxic dusts (HighTox scenario) and the other component represents limit tests (LowTox scenario). The parameters of the mixture of distributions have been estimated by minimizing the mean squared error of the empirical cumulative distribution function (Solver routine add-in from MS-Excel). From these data, the log(LC50) values for the HighTox scenario were assumed to follow a smallest extreme value distribution with probability density function given by

$$f(x) = \frac{1}{b}\exp\left[\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right],$$

where $x = \log(\text{LC50})$, $a = 1.145$ is the location parameter and $b = 0.836$ is the scale parameter. This is equivalent to an assumption that the LC50 values follow a Weibull distribution with density, $f(x) = \lambda\gamma x^{\gamma-1}\exp(-\lambda x^{\gamma})$, with $\lambda = \exp(-a/b)$ and $\gamma = 1/b$. The LC50 values for the LowTox scenario were assumed to have a log-normal distribution, with log(LC50) assumed to be normally distributed with mean 1.638 and standard deviation 0.038 (logarithms being taken to base e). For the LowTox and HighTox scenario, a range of LC50 values for dusts was considered which covers GHS classes 4-6 and 1-3, respectively. Note that for the classification probabilities evaluated in this study, the distribution of LC50 values within rather than across GHS classes is relevant because classification probabilities are conditional on true GHS class.

Slope values for probit model. As the data on new substances did not contain sufficient information on the probit slope β, published data from Fowles et al. (1999) [1] and a data compilation of the US Environmental Protection Agency (EPA/RTP, by courtesy of Dr. George Woodall) have been used for this purpose. Fowles et al. (1999) provide the following description of their data: In all, 47 chemicals and 120 animal lethality data sets were used to determine benchmark concentrations using probit and Weibull models. Species tested included rats (65 studies), mice (42 studies), guinea pigs (8 studies), hamsters (3 studies), and rabbits and dogs (1 study each). The studies were included as they became available through literature searches on the acute health effects of airborne chemicals. The exposure duration for all studies ranged from 5 min to 8 h. The greatest range of durations for a given chemical and test species was 5 min to 4 h (crotonaldehyde). Each data set used contained at least four treatment groups.

Ron Crosier (US Army; retired) used 100 acute inhalation studies for 41 chemicals to come up with a median probit slope of 11.27, chemical GSD of 1.38 and experiment GSD of 1.82 by analyses of variance on log (probit slope) of the Fowles data.. Twenty studies were excluded from Ron's analyses (exclusion criteria are not known). The Fowles et al. 1990 data set covered studies from different species (rats, mice, guinea pigs, hamsters, rabbits and dogs) for exposure times ranging from 5 min to 8 hrs, whereas TG 436 preferably test rats for 4 hrs. As the mean slope and SD were derived from a wider range of test species and exposure times instead of 4-hr rat studies, the effect on classification probabilities if the mean slope and SD were derived from 4hr rat studies is not known.

The other database used for empirical slope values was provided by Dr. George Woodall (US EPA/RTP).

[1] Paragraph has been modified in response to issue #4, US Position Paper 2.

Ron Crosier used 74 experiments for 24 chemicals and the median probit slope was 9.45, chemical GSD= 1.71 and experiment GSD=1.91. It remains to be clarified whether only 4-hr rat acute data has been used and what the final size of retained sample of probit slope was. In fact, the database went through numerous revisions after the parameterization was conducted, and this might have changed the input value for probit slopes.

Using data from the EPA database, the slope of the concentration-response curve in terms of log10(concentration) was assumed to follow an identical log-normal distribution for both the LowTox and HighTox scenario, with slope assumed to be independent of LC50. The logarithm taken to base e of the slope was assumed to be normally distributed with mean log(10) = 2.30 and standard deviation log(1.5) = 0.405. It is noted that empirical slopes for tests conducted on rats ranged from 4.3 to 34.7 (n=28) and from 5 to 95 (n=26) for chemicals included in both the Fowles study and the EPA data base and those occurring only in the EPA data base, respectively (EXCEL spreadsheet *Analysis of refs Fowles and EPA db_042308.xls*). Using this data set, the mean log slope value can be shown to be significantly different between species (ANOVA, $P<0.01$), whereby rats (n=54 studies) had a greater mean log slope compared to guinea pig (n=7 studies), hamster (n=3 studies) and mouse (n=34 studies).

2.3 Implementation of the simulation model

The mathematical models were implemented in different environments and operated by independent teams (Nigel Stallard and Matthias Greiner). The results for one example scenario (LC50 = 2.5mg/L, probit slope = 1.5, starting concentration 0.05mg/L) were confirmed by a third, independent team (Ron Crosier).The results reported here (and in previous versions) were generated by Nigel Stallard. The principle of the numerical integration was to use calculated classification probabilities and expected numbers of animals used (n) and dead (d) for each true GHS class averaged over the distributions of LC50 and slope given above. Therefore, the impact of LC50 and probit slope on classification probabilities is captured in the model in a way that allows inference be made to the general expected classification probabilities. The model could also be used to estimate classification probabilities for selected LC50 and slope values. However, this application of the model would not be adequate for estimating general expected classification probabilities. The classification performance is numerically summarized in form of a classification matrix. The diagonal entries in this matrix contain the probabilities for correct classification whereas cells above and below the diagonal entries contain the probabilities of underclassification and overclassification, respectively. The classification probabilities express both variability and uncertainty and are therefore given without a confidence interval.

The results presented in this report were obtained by integration over LC50 conducted separately for the ranges 0.003 to 0.05, 0.05 to 0.5, 0.5 to 1, 1 to 5 and 5 to 100 mg/L. Each range was divided into intervals equally spaced on a logarithmic scale and the integral over the range was estimated by the sum of the integrand evaluated at the midpoints of these intervals. For the HighTox scenario, 50 intervals were used for each range. For the LowTox scenario, as the distribution is considerably more peaked a finer grid of 500 intervals per range was used. Integration over the concentration-response curve slope β was conducted in a similar fashion using 50 intervals equally spaced on a logarithmic scale over the range from 1.32 to 75.9 (that is log(10) ± 5 standard deviations on the logarithmic scale). The evaluation of classification probabilities was conditional on the true LC50 being in one of the (true) GHS class intervals. Therefore, for each calculation of classification probabilities (rows in Tables 1-8) only the grid points corresponding to the true GHS class were considered.

3. RESULTS OF THE CASE STUDY

3.1 LowTox scenario

The results for limit test chemical substances and starting values of 5 mg/L are shown in Table 4. The probability of correct classification are 0.725, 0.455 and 0.94 for classes 4, 5 and 6, respectively. Category 5 has the lowest probability of correct classification. The probability of classifying into a too high class is 0.274 for true GHS 4 and less than 0.05 for true GHS 5 and 6. These probabilities are not affected by the choice of other starting concentrations ranging from 0.05 to 1 mg/L (Tables 1 to 3). The expected number of animals enrolled in the experiments clearly depends on the choice of the starting concentration. This number is smallest when starting at 5 mg/L (mean 9.3 animals) and largest when starting at 0.05 mg/L (mean 24). On the other hand, the expected number of deaths appears to be independent of the starting concentration. As could be expected, this number increases with the true toxicity level of the chemical and ranges from 0.1 to 3.2 (mean 2.7).

3.2. HighTox scenario

The results for non-limit test chemical substances and starting values ranging from 0.05 to 5 mg/L are shown in Tables 5 to 8. The probability of correct classification ranges from 0.39 (GHS 5) to 0.97 (GHS 4) and appears to be not affected by the choice of different starting concentrations. Chemicals belonging the true GHS 3 category appear to have the second worse classification probability. The mean (across GHS classes) expected number of animals enrolled in the experiments was 21.8, 16.6, 12.7 and 13.1 for the starting concentrations 0.05, 0.5, 1 and 5 mg/L. The mean numbers of expected deaths for same starting concentrations were 4.8, 4.9, 5.5 and 7.0 and this showed a slight increasing trend from low to high starting concentrations. In all cases, as expected, the expected number of deaths is related to the true GHS level of the tested chemical.

4. DISCUSSION

The biostatistical performance assessment of the TG 436 described in the literature [2] has been critically reviewed by the PAG in terms of the underlying model assumptions and implementation of the biostatistical assessment. The goal was to achieve a consensus about a suitable study design for re-assessment of the classification performance.

Data available for modeling. LC50 and slope values were modeled from two distinct empirical data sets. The model did not account for any possible correlation (i.e. parameter covariance) between LC50 and slope. A further study on the TG 403 suggests that there is no correlation between these two parameters[2].

Observation of endpoint. The model assumes that the biological responses can be observed

[2] Performance assessment: comparison of 403 and CxT protocols via simulation and for selected real data sets (prepared by Bruce Allan; February 29, 2008).

unambiguously. The number of dead animals or moribund animals are counted in each batch of the experiment. These numbers determine the pathway through the experiment and thus the classification of the chemical. Observational misclassification could result in a classification bias towards less toxic categories. The PAG concluded that observational errors are not an issue because the TG require a 14-day observation period.

Sample sizes and decision rules. Batches of six animals at each concentration are tested. These sample sizes and the decision rules, which define all possible pathways from a starting concentration to the last concentration tested and the final classification affect the classification performance. It was agreed in the PAG that the sample sizes and decision rules should be kept as they are. Although there is a potential benefit of an optimization of sample sizes and decision rules, a refinement of these rules require specification of the utilities of outcomes (and misclassifications). This would be a research project beyond the scope of the PAG.

Modelling of the LC50 values. The use of the Weibull model provided the best fit as a mixture component representing highly toxic chemicals using the available empirical data. It is possible that other distribution models would provide a better fit if more data became available.

Modelling of the probit slope values. Including empirical non-rat data in the fitting of a distribution for the probit slope has likely introduced a slight bias towards smaller slope values as suggested by an analysis of slope values for a possible species effect. Thus, the inclusion of non-rat data has more likely resulted in an underestimation rather than overestimation of correct classification probabilities.

Gases and vapors were not included in the performance assessment. The reason for the omission is explained in section 2.2. It is not known that gases and vapors have different input distributions (LC50 and slope values) compared to dusts. If such information is available the model can easily be run for modified distributions reflecting those classes of articles. According to the model, the classification probabilities for any article depend only on the two quantities LC50 and slope and therefore the biometrical evaluation approach is also valid for gases and vapors. The PAG has considered that the results obtained for dusts (given the LC50 and slope values) are in principle also valid for gases and vapors.

The effect of the probit slope on the classification probability was not formally assessed. The effect of the probit slope is captured in the analysis by using distributions of this parameter for the stochastic simulations. It should be possible to conduct a formal sensitivity analysis to assess the impact of this parameter on the classification probability although the gain of information is probably limited. It is already known that the biological characteristic expressed by low probit slope values affects the precision of estimates of LC50 as it will affect the probability of correct classification using the TG 436.

Independence of observations. The implicit assumption is made that the responses from animals within one batch of the study (i.e., tested in one concentration step) are independent and identically distributed (iid) and that responses of all animals in one experiment are independent. Several processes may lead to overdispersion of responses and thus violation of the assumption. These include auto-correlation among the animals of one litter or parameter heterogeneity. The first of these is unlikely to lead to substantial problems, and any the auto-correlation could be reduced by randomisation, the provision for which is given in both TG. The PAG concluded that the randomisation of animals is a matter for the test house to consider. A special and important of parameter heterogeneity is due to the experimental error (concentrations being random rather than a fixed effect) or the error due to induced (by the chemical) reduction of breath volume. A variance greater than that predicted by the model will most likely affect the misclassification probabilities. Although we have no direct evidence for the absence of extrabinomial variation in a “typical” acute inhalation experiment, it was concluded that the assumption of independence is acceptable. Other accepted guidelines make explicit or implicit use of a similar assumption without

proof. It has been shown for the draft TG 433 that extrabinomial variation has led to less predictable classification of substances with steep concentration-response curves [3].

Advantages of the biometrical performance assessment. The approach is well documented in the scientific literature [2,7-15], fully model-based (all conditions are under control), transparent and reproducible and no animals are used. The advantage of the biometrical performance assessment approach compared to animal tests is not only the saving of experimental animals. The usefulness of a direct comparison of experimental LC50 results and experimental GHS classifications is also limited by experimental errors of the LC50 estimation that would inevitably occur unless very large sample sizes were used.

Disadvantages of the biometrical performance assessment. The validity of the results relies on the quality of the data (Is the data representative for articles to be tested in practice?) and on the model assumptions, which are discussed in this section. A limitation is that gases and vapors could not be considered in the assessment. It is not known whether the results (classification probabilities) obtained for dusts will be similar to those for gases and vapors. The potential instability of the test concentrations is not covered by the model. Modeling overdispersion due to experimental error would involve additional model parameters and therefore increase uncertainty.

Advantages of the TG 436. The maximum number of animals used to classify a chemical is 24, which is advantageous under the 3R's principle. It can be seen as an advantage of the TG 436 compared to the TG 403 that the experimental pathway (sequential choice of concentrations) is defined by the procedure. The defined testing procedure provides the fundament on which a biometrical performance assessment of the TG 436 is possible. The PAG identified the uncertainty about the actual procedure of a TG 403 experiment as the main obstacle to a biostatistical performance assessment of the TG 403. In fact, even if experimental pathways of a TG 403 were known with certainty, all model assumptions and limitations discussed in this section would equally apply to a performance assessment of TG 403. The standardized procedure of the TG 436 and the minimized use of animals is an advantage, which could be especially relevant when testing less toxic substances, where C&L rather than LC50 estimation is a priority from a management point of view. When testing toxic substances, the TG 436 results could be very valuable to select appropriate concentration ranges for subsequent TG 403 studies. Under certain circumstances, additional testing (i.e. traditional LC50 study or Cxt protocol) is required for chemicals which have already been subjected to TG 436 testing. This situation may occur either after completion of TG 436 (i.e. when a GHS classification has been obtained) or before completion of TG 436 testing (i.e. no GHS classification has been obtained). If a TG 436 test has been completed, the GHS classification could be used for range finding in the TG 403 or CxT study. Moreover, if the outcomes (i.e. the number of dead/moribund out of six animals tested at all 2, 3 or 4 concentrations) of TG 436 are available and the experiment has been conducted under comparable conditions, they should be used as data points in the calculation of all relevant parameters. This will lead to an augmented data set available the TG 403 (and CxT) data analysis and thus more precise estimates. Investigators may choose to refrain from the pooled analysis if there is evidence that the experimental conditions (e.g., physical state of exposure atmospheres, particle size etc.) of the TG 436 and the TG 403 are not comparable and the data points derived from TG 436 have an appreciable impact on parameter estimates using appropriate statistical methods. When results for all four concentrations are available from a TG 436 experiment a crude estimate of LC50 may be obtained. If a TG 436 test has not been completed but any outcome of the TG 436 is available, this information may be used for range finding in the TG 403 (and CxT) study.

Disadvantages of the TG 436. The fixed concentrations of the sequential pathway may be problematic if the inhalation study cannot achieve a stable chamber atmosphere. Although using GHS seems an advantage now in light of harmonizing classification and labeling, changes to GHS bands in the future mean that the biometrical performance of TG 436 has to be reassessed; this would not occur with TG 403

as the concentrations are not fixed to GHS cutoff values. TG 436 was not designed to get a lethality point estimate and concentration-response relationship. However, TG 436 data would be useful to select concentrations for TG 403 studies.

Overclassification and underclassification. The evaluation of the utilities of correct and incorrect classifications was beyond the remit of the study. The two types of misclassification are associated with different adverse consequences. Misclassification of a less-toxic chemical into a toxic class (overclassification) can lead to economic losses for the chemical and transport industry to consider while misclassification of a toxic chemical into a less toxic class (underclassification) is a public health concern as it carries risk of increased exposure to humans and adverse health effects. The assessment of the relative importance of such consequences is a management issue. There is an inverse relationship between correct classification by TG 436 and closeness of the LC50 to the GHS boundaries. However, this effect will apply too when chemicals are classified according to LC50 values estimated by TG 403.Misclassification is a general problem and not specific to TG 436.

Assessment of classification for articles with medium toxicity. The use of two scenarios for toxicity levels (LowTox and HighTox) was based on an explorative analysis of available LC50 data. This analysis suggested that a 2-component mixed model fitted the data best. If more data were available, this analysis could be repeated. However, it would be inconsistent with our empirical approach to presuppose that there is a subpopulation of articles with medium toxicity if this is not supported by the data.

Predictive value of classification. The predictive value of the classification, i.e. the probability that a given classification outcome corresponds to the true class has not been considered in the agreed study plan. An evaluation of the predictive value of the classification requires the knowledge of the distribution of the true toxicological parameters (LC50 values and probit slopes) in the real world population of articles that can be subjected to the testing. If the available data on LC50 and probit slope is considered an unbiased sample of the population parameters, the results can be interpreted in terms of predictive values using the Bayes theorem.

Probit response function. The mild assumption is made that the probability of death in acute inhalation toxicology follows a probit concentration-response function with concentration measured on a logarithmic scale. This choice is well supported by the scientific literature [4-6]. The probit model is preferred to the logit model on conceptual grounds. The probit slope is the number of standard deviations per a factor of 10 change in the concentration whereas the logit slope does not have this interpretation. The PAG concluded that the probit is a valid model in this context and that it is acceptable to use this model for evaluating the outcome probabilities in the performance characterization. If the true concentration-response function would follow a logit or another function, this would slightly change the classification probabilities at each concentration step. As the number of animals tested is small ($n = 6$), the classification outcomes would likely not be affected. The difference in binomial random outcomes for a small sample size of six animals predicted by a probit versus an alternative concentration-response model seems negligible. Thus, the model is robust with regard to the choice of the response function.

Mathematical modeling approach. The performance of the classification is estimated based on mathematical models. This entails that chemical substances with assigned true levels of toxicity are subjected to a virtual classification. The results are presented for a set of given parameters (LC50, probit slopes). The performance is estimated as classification probabilities, conditional on a true category defined by the assigned toxicological parameters. The usefulness of these classification probabilities for predicting GHS categories of unknown compounds will be limited to those parameters that occur in real life. It was agreed that a random effects model to account for experimental error is not required. It was concluded that the mathematical evaluation approach would be useful to obtain valid and reliable predictions of the

classification performance if the modeling is conducted using realistic parameters. The validity of the results of this performance assessment depends on whether the data is representative and the model assumptions are met. However, it is important to consider both the level of uncertainty about a model assumption and the robustness of the model in the presences of a violation of an assumption. The detailed interpretations, assumptions and limitations presented in this section do not indicate that the assumption with regard to modeling of empirical LC50 and probit slope values and the choice of the probit response function is very uncertain and influential at the same time.

GHS system. The PAG has been requested to investigate the classification probabilities using the GHS classification system. Chemicals falling within the true GHS class 5 had the lowest probability of correct classification. The second worse classification probabilities were observed for chemicals belonging to the GHS 3 class. Similar results were previously reported [2]. Given the uniformity of this result for all scenarios under investigation and considering the narrow interval of the GHS classes 3 and 5, it can be concluded that this misclassification is a problem inherent in the GHS classification system. There are not only difficulties in classification due to some narrow categories, but potential legal problems due to the inconsistency of the boundaries (and units) for gases and vapors. The distinction between gas and vapor is somewhat arbitrary. If the normal status of a chemical at room temperature and atmospheric pressure is not unanimously defined, some manufacturer may decide to declare a new chemical as gas or vapor to get a better classification. The gas and vapor classification systems should be equivalent so that no manipulation of the overall system is possible. Likewise for mists (aerosols). For some chemicals, it may be difficult to generate vapor without forming a mist, or generate a mist without forming a vapor. A modification of the GHS system was not considered in this study. At the expert meeting (Washington, April 2008) the hypothetical effect of using 10 vs. 12.5mg/L as limit for GHS cat. 5/5+ for aerosols/dusts was discussed. It was concluded that a wider interval for the GHS 5 category than the one used in the simulation would be anticipated to give a larger probability of correct classification of chemicals.

Factors affecting the classification probabilities. The misclassification due to unequal and narrow intervals of the GHS has been described above. The results suggest that the performance of TG 436 is not markedly influenced by the choice of the starting concentration. The results clearly show that the misclassification probabilities are highest near the boundaries between the GHS classes for small beta slope values. It is noted that the misclassification probabilities are asymmetric since misclassification into a more toxic class is much more likely than misclassification into a less toxic one. The results did not show a dependency of classification probabilities on the choice of the starting concentration. In a biometrical performance assessment of the ATC method for acute dermal toxicity it was found that, since estimated slopes for most of the substances were greater than 1, the results were in fact independent of the starting dose [7]. The choice of the LC50 distribution has an effect on the classification probabilities. This can be derived from noticeable differences between the LowTox and HighTox scenario in classification of the true GHS classes 4, 5 and 6.

5. CONCLUSIONS

1. Considering the Holzhütter paper [2], other publications on the biostatistical performance assessments using a similar approach, the documentation (tables) provided on the Expert Consultation Meeting 2006 in Berlin and the present report, the biostatistical model for performance characterization of the acute toxic class method (TG 436) is documented in sufficient detail to allow a critical review and reproduction of the results.

2. The model assumptions for biostatistical performance characterization are valid within acceptable limits of uncertainty. Violations of model assumptions (e.g. concentration-response function) are not likely to have a large effect on classification results.

3. Empirical data (mainly obtained from GLP compliant tests in the process of notification of New Chemicals to meet requirements of the Dangerous Substances Directive 67/548/EEC) were available and have been used to support the model assumptions for the performance characterization. While the limitation of the available data has been acknowledged, it is plausible to assume that results for scenarios not considered in this study (e.g., other physical classes) would not be contradictory. Moreover, in the absence of classification probabilities for the TG 403 that could be used for comparison, no additional analyses were carried out using other data sources..

4. Irregular spacing of GHS categories across the physical states might affect the results in other physical states (e.g., gases and vapors) suggesting that the results (classification probabilities) obtained with dusts may not be similar to those predicted for other physical states. GHS spacing of the classification bands is one of the factors affecting the classification probabilities. Irregular spacing is a general problem with the current GHS system and not specific to TG 436.

5. The extension of the PAG evaluation for dusts to other physical classes is currently hampered by inconsistencies in the physical characterization of substances across the TG, the GHS rules and available data bases for chemicals.

6. The biostatistical performance assessment reported here allows estimation of the probabilities for correct classifications and classification into less or more lenient classes for dusts as a case study. A performance assessment for gases and vapors could be based on the same biostatistical approach and therefore would be expected to yield valid results if similar input information was available as is the case for dusts.

7. The results suggest that the TG436 can be used for classification and labeling. Chemicals falling into the GHS categories 3 and 5 generally have a lower probability of correct classification compared to other chemicals. This is due to the unequal and narrow band width of these categories in combination with the handling of decisions at the limit concentration.

8. When classification and labeling is the intention, TG 436 has an advantage in that it uses fewer animals than TG 403. When testing toxic substances, the TG 436 results provide valuable information for concentration range finding of TG 403 experiments. The pre-specification of experimental pathways for TG 436 facilitates the execution of the protocol in the laboratory.

6. ACKNOWLEDGEMENTS

This report was prepared by Matthias Greiner (BfR, Germany) with support of the Performance Assessment Group (PAG). Members of the PAG were (in alphabetical order by country) Elke Genschow, Matthias Greiner, Manfred Liebsch, Christine Müller-Graf, Hans-B. Richter-Reichhelm and Horst Spielmann from Germany, Josje Arts and Marcel van Raaij from the Netherlands, Robert Combes and Nigel Stallard from the UK and Iris Camacho, Finis Cavender, Ronald Crosier, Allan H. Marcus, Harry Salem, Douglas Sommerville and George Woodall from the USA. The PAG through their members had the privilege of access to data collected at the US Environmental Protection Agency (EPA). The PAG also acknowledges extract data from the data base for notification of new substances according to the Chemicals Act from Herbert Thiele (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Germany).

REFERENCES

[1] Fowles JR, Alexeeff GV and Dodge D, 1999. The use of benchmark dose methodology with acute inhalation lethality data. Regul. Toxicol. Pharmacol. 29: 262-278.
[2] Holzhutter HG, Genschow E, Diener W, Schlede E, 2003. Dermal and inhalation acute toxic class methods: test procedures and biometric evaluations for the Globally Harmonized Classification System. Arch. Toxicol. 77: 243-254.
[3] Stallard N, Whitehead A, and Indans I, 2003. Statistical evaluation of the fixed concentration procedure for acute inhalation toxicity assessment. Hum. Exp. Toxicol. 22: 575-585.
[4] Unkelbach HD, Wolf T (1985) Qualitative Dosis-Wirkungs-Analysen. Gustav Fischer Verlag, Stuttgart
[5] Bliss CI (1934) The methods of probits. Science 79:38–39
[6] Finney DJ (1971) Probit analysis, 3rd en. Cambridge University Press, Cambridge
[7] Diener W (1998) Biometric evaluation of the ATC method for the determination of the acute dermal toxicity of chemicals. Biometrical Journal, 40: 979-991.
[8] Diener W and E. Schlede. Acute toxic class methods: Alternatives to LD/LC50 tests. Altex-Alternativen zu Tierexperimenten 16 (3):129-134, 1999.
[9] Diener W. Biometric evaluation of the ATC method for the determination of the acute dermal toxicity of chemicals. Biometrical Journal 40 (8):979-991, 1998.
[10] Diener W. and E. Schlede. FDP and ATC method: A mathematical comparison. Human & Experimental Toxicology 15 (10):855-856, 1996.
[11] Diener W, L. Siccha, U. Mischke, D. Kayser, and E. Schlede. The Biometric Evaluation of the Acute-Toxic-Class Method (Oral). Archives of Toxicology 68 (10):599-610, 1994.
[12] Diener W, U. Mischke, E. Schlede, and D. Kayser. The Biometrical Evaluation of the Oecd Modified Version of the Acute Toxic Class Method (Oral). Archives of Toxicology 69 (10):729-734, 1995.
[13] Diener W, D. Kayser, and E. Schlede. The dermal acute toxic class method: test procedures and biometric evaluations. Archives of Toxicology 72 (12):751-762, 1998.
[14] Diener W, D. Kayser, and E. Schlede. The inhalation acute toxic class method: test procedures and biometric evaluations. Archives of Toxicology 71 (9):537-549, 1997.
[15] Schlede E, U. Mischke, W. Diener, and D. Kayser. The International Validation-Study of the Acute Toxic Class Method (Oral). Archives of Toxicology 69 (10):659-670, 1995.

Annex : Classification tables

Table 1. TG 436 classification results in the LowTox scenario with starting concentration 0.05 mg/L.

True	Classification								
class	1	2	3	4	5	6	Total	E(*n*)	E(*d*)
4	0.000	0.000	0.000	**0.725**	0.264	0.010	0.226	24.0	3.2
5	0.000	0.000	0.000	0.499	**0.455**	0.046	0.774	24.0	2.5
6	0.000	0.000	0.000	0.001	0.059	**0.940**	0.000	24.0	0.1
Mean	0.000	0.000	0.000	0.550	0.412	0.038	1.000	24.0	2.7

LC50 values typical for limit test chemicals were used for the assessment (LowTox scenario). Total = proportion of chemicals falling into this GHS category. E(*n*) = expected number of animals, E(*d*) = expected number of deaths. Mean = mean value over GHS categories. Correct classification probabilities are printed in bold.

Table 2. TG 436 classification results in the LowTox scenario with starting concentration 0.5 mg/L (legend: see Table 1).

True	Classification								
class	1	2	3	4	5	6	Total	E(*n*)	E(*d*)
4	0.000	0.000	0.000	**0.725**	0.264	0.010	0.226	18.0	3.2
5	0.000	0.000	0.000	0.499	**0.455**	0.046	0.774	18.0	2.5
6	0.000	0.000	0.000	0.001	0.059	**0.940**	0.000	18.0	0.1
Mean	0.000	0.000	0.000	0.550	0.412	0.038	1.000	18.0	2.7

Table 3. TG 436 classification results in the LowTox scenario with starting concentration 1 mg/L (legend: see Table 1).

True	Classification								
class	1	2	3	4	5	6	Total	E(*n*)	E(*d*)
4	0.000	0.000	0.000	**0.725**	0.264	0.010	0.226	12.0	3.2
5	0.000	0.000	0.000	0.499	**0.455**	0.046	0.774	12.0	2.5
6	0.000	0.000	0.000	0.001	0.059	**0.940**	0.000	12.0	0.1
Mean	0.000	0.000	0.000	0.550	0.412	0.038	1.000	12.0	2.7

Table 4. TG 436 classification results in the LowTox scenario with starting concentration 5 mg/L (legend: see Table 1).

True	Classification								
class	1	2	3	4	5	6	Total	E(*n*)	E(*d*)
4	0.000	0.000	0.000	**0.725**	0.264	0.010	0.226	10.4	3.2
5	0.000	0.000	0.000	0.499	**0.455**	0.046	0.774	9.0	2.5
6	0.000	0.000	0.000	0.001	0.059	**0.940**	0.000	6.0	0.1
Mean	0.000	0.000	0.000	0.550	0.412	0.038	1.000	9.3	2.7

Table 5. TG 436 classification results in the HighTox scenario with starting concentration 0.05 mg/L.

True class	Classification 1	2	3	4	5	6	Total	E(*n*)	E(*d*)
1	**0.969**	0.031	0.000	0.000	0.000	0.000	0.007	6.2	5.6
2	0.008	**0.961**	0.031	0.000	0.000	0.000	0.098	12.1	5.6
3	0.000	0.087	**0.773**	0.139	0.000	0.000	0.119	17.8	5.5

LC50 values typical for non-limit test chemicals were used for the assessment (HighTox scenario). Total = proportion of chemicals falling into this GHS category. E(*n*) = expected number of animals, E(*d*) = expected number of deaths. Mean = mean value over GHS categories. Correct classification probabilities are printed in bold.

Table 6. TG 436 classification results in the HighTox scenario with starting concentration 0.5 mg/L (legend: see Table5).

True class	Classification 1	2	3	4	5	6	Total	E(*n*)	E(*d*)
1	**0.969**	0.031	0.000	0.000	0.000	0.000	0.007	12.0	11.4
2	0.008	**0.961**	0.031	0.000	0.000	0.000	0.098	12.0	5.6
3	0.000	0.087	**0.773**	0.139	0.000	0.000	0.119	12.3	5.5

.

Table 7. TG 436 classification results in the HighTox scenario with starting concentration 1 mg/L (legend: see Table5).

True class	Classification 1	2	3	4	5	6	Total	E(*n*)	E(*d*)
1	**0.969**	0.031	0.000	0.000	0.000	0.000	0.007	18.0	17.4
2	0.008	**0.961**	0.031	0.000	0.000	0.000	0.098	17.8	11.4
3	0.000	0.087	**0.773**	0.140	0.000	0.000	0.119	12.5	6.0

.

Table 8. TG 436 classification results in the HighTox scenario with starting concentration 5 mg/L (legend: see Table5).

True class	Classification 1	2	3	4	5	6	Total	E(*n*)	E(*d*)
1	**0.969**	0.031	0.000	0.000	0.000	0.000	0.007	24.0	23.4
2	0.008	**0.961**	0.031	0.000	0.000	0.000	0.098	23.8	17.4
3	0.000	0.087	**0.773**	0.140	0.000	0.000	0.119	18.2	11.7

.