



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Environmental and Molecular Mutagenesis

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa12708>

Paper:

Gollapudi, B., Johnson, G., Hernandez, L., Pottenger, L., Dearfield, K., Jeffrey, A., Julien, E., Kim, J., Lovell, D., et. al. (2013). Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. *Environmental and Molecular Mutagenesis*, 54(1), n/a-18.
<http://dx.doi.org/10.1002/em.21727>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

Research Article

Quantitative Approaches for Assessing Dose–Response Relationships in Genetic Toxicology Studies

B.B. Gollapudi,¹ G.E. Johnson,² L.G. Hernandez,³ L.H. Pottenger,¹ K.L. Dearfield,⁴
A.M. Jeffrey,⁵ E. Julien,⁶ J.H. Kim,^{7*} D.P. Lovell,⁸ J.T. MacGregor,⁹
M.M. Moore,¹⁰ J. van Benthem,³ P.A. White,¹¹ E. Zeiger,¹² and V. Thybaud¹³

¹The Dow Chemical Company, Midland, Michigan

²Institute of Life Science, College of Medicine, Swansea University, SA2 8PP
Wales, United Kingdom

³Laboratory for Health Protection Research, National Institute for Public Health
and the Environment (RIVM), Bilthoven 3720 BA, The Netherlands

⁴U.S. Department of Agriculture, Food Safety and Inspection Service,
Washington, District of Columbia

⁵Department of Pathology, New York Medical College, Valhalla, New York

⁶LifeSci Research Service, Rockville, Maryland

⁷ILSI Health and Environmental Sciences Institute, Washington, District of Columbia

⁸Division of Biomedical Sciences, St. George's University of London, Cranmer
Terrace, London SW17 0RE, United Kingdom

⁹Toxicology Consulting Services, Arnold, Maryland

¹⁰Division of Genetic and Reproductive Toxicology, National Center for
Toxicological Research, Jefferson, Arkansas

¹¹Environmental Health Sciences and Research Bureau, Healthy Environments
and Consumer Safety Branch, Health Canada, Tunney's Pasture 0803A,
Ottawa, Ontario, Canada K1A 0K9

¹²Errol Zeiger Consulting, Chapel Hill, North Carolina

¹³Sanofi, Vitry-Alfortville Research Center, Drug Disposition, Preclinical Safety,
and Animal Research, Vitry-sur-Seine, France

Genetic toxicology studies are required for the safety assessment of chemicals. Data from these studies have historically been interpreted in a qualitative, dichotomous “yes” or “no” manner without analysis of dose–response relationships. This article is based upon the work of an international multi-sector group that examined how quantitative dose–response relationships for *in vitro* and *in vivo* genetic toxicology data might be used to improve human risk assessment. The group examined three quantitative approaches for analyzing dose–response curves and deriving point-of-departure (POD) metrics (i.e., the no-observed-genotoxic-effect-level (NOGEL), the threshold effect level (Td), and the benchmark dose (BMD)), using data for the induction of micronuclei and gene mutations by methyl methanesulfonate or ethyl methanesulfonate *in vitro* and *in vivo*. These results suggest that the POD descriptors obtained

using the different approaches are within the same order of magnitude, with more variability observed for the *in vivo* assays. The different approaches were found to be complementary as each has advantages and limitations. The results further indicate that the lower confidence limit of a benchmark response rate of 10% (BMDL₁₀) could be considered a satisfactory POD when analyzing genotoxicity data using the BMD approach. The models described permit the identification of POD values that could be combined with mode of action analysis to determine whether exposure(s) below a particular level constitutes a significant human risk. Subsequent analyses will expand the number of substances and endpoints investigated, and continue to evaluate the utility of quantitative approaches for analysis of genetic toxicity dose–response data. Environ. Mol. Mutagen. 54:8–18, 2013. © 2012 Wiley Periodicals, Inc.

Key words: mutagenicity; risk assessment; low-dose; benchmark dose; point of departure; no-observed-genotoxic-effect-level; dose–response

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: J.H. Kim, ILSI Health and Environmental Sciences Institute, Washington, DC 20005.

E-mail: jkim@hesiglobal.org

Received 1 February 2012; provisionally accepted 19 July 2012; and in final form 23 July 2012

DOI 10.1002/em.21727

Published online 18 September 2012 in

Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Evaluation of genetic damage is a pivotal component of the safety assessment of chemicals. Historically, genetic toxicologists have relied on a battery of tests to screen for various types of genetic damage in an attempt to cast a wide net to identify potential mutagens. The typical battery of *in vitro* and *in vivo* hazard screening tools includes: (1) an *in vitro* test for gene mutations in bacteria, (2) an *in vitro* test for chromosomal damage and/or gene mutations in cultured mammalian cells, and (3) an *in vivo* test for cytogenetic effects in rodent bone marrow cells. This testing paradigm has been effective in preventing the introduction of potent genotoxic agents, and genotoxic carcinogens, into the environment and it is widely agreed that this approach has served the regulatory community well.

The need to assess *in vivo* dose–response relationships as a follow-up to hazard screening when evaluating the risks associated with human exposure was recognized at the time regulatory genetic toxicology testing was first introduced [Flamm et al., 1977; Thybaud et al., 2007], but the lack of suitable *in vivo* methods for evaluating mutations and other genetic damage available at the time, coupled with initial enthusiasm about the apparent excellent correlation between bacterial mutagenicity and rodent carcinogenicity [McCann and Ames, 1976], led to the establishment of a qualitative screening battery designed to categorize agents as either non-mutagens or demonstrated mutagens (e.g., see [Dearfield et al., 1991]). In contrast to other toxicology disciplines, decisions about genotoxic risk continue to be based on qualitative factors that classify an agent as “positive” or “negative” in the above test battery and supplementary tests. Genetic toxicologists, not unlike other toxicologists, have embraced the concept of testing up to a maximum tolerated dose (MTD) for *in vivo* tests, and to maximum cytotoxicity and/or concentration limit for *in vitro* tests. The use of relatively high doses/concentrations for genotoxicity tests is based on limited sensitivity and the finite number of animals/cells used in an assay, and geared toward maximizing the potential to detect an effect. This is appropriate for hazard identification, but when testing does not provide the data required for a dose–response curve that covers the range from background to maximal response, that data cannot be used for the determination of a no-observed-genotoxic-effect-level (NOGEL), and moreover, the data cannot be used for estimation of risk at realistic exposure levels.

Positive findings for genotoxicity have been encountered in experiments using *in vitro* test systems with chemicals that do not appear to have genotoxic activity in *in vivo* assays [Thybaud et al., 2007]. This is perhaps not surprising, considering the much higher exposure concentrations that can be achieved in *in vitro* assays relative to

actual levels of exposure achievable *in vivo*, and considering the differences in metabolism, pharmacokinetics, and target cell distribution between *in vivo* and *in vitro* systems. Nevertheless, even in the presence of other toxicology data indicating negligible concern for human safety, positive results from *in vitro* genotoxicity tests have often led to the prohibition of use and/or cessation of development of compounds that may have had substantial societal value. Thus, it is important that genetic toxicity testing move toward a more quantitative approach for the evaluation of genetic toxicology data.

In an effort to improve the prevailing paradigm in genetic toxicology, experts from North America, Europe, and Japan initiated three workgroups to address the aforementioned issues, as well as other related issues, under the auspices of the Health and Environmental Sciences Institute (HESI) of the International Life Sciences Institute (ILSI) [Gollapudi et al., 2011]. Two workgroups have published critical evaluations of emerging technologies to assist genetic toxicologists in improving the scientific basis of sound *in vitro* genetic toxicity testing for more accurate human risk assessment [Lynch et al., 2011], and a follow-up strategy for determining the relevance of *in vitro* test results to human health [Dearfield et al., 2011], respectively. The present report is based on the work of the third workgroup (the “Quantitative Analysis Workgroup”) whose primary objectives are to develop strategies for the quantitative analysis of *in vitro* and *in vivo* genotoxicity dose–response data, and moreover, to consider critically how quantitative analyses might be used to assess human risk.

This initial work examined different quantitative dose–response modeling approaches for derivation of point-of-departure (POD) metrics such as NOGEL, benchmark dose (BMD), and threshold effect levels (Td) for the well-studied alkylating agents methyl methanesulfonate (MMS) and ethyl methanesulfonate (EMS). Subsequent work will expand the analyses to include additional compounds and endpoints, develop approaches for quantitative comparisons across endpoints and extrapolations from *in vitro* responses to *in vivo* exposure situations, and critically evaluate the ability to use quantitative data from experimental models and human exposure scenarios to determine acceptable exposure levels.

BACKGROUND INFORMATION

Quantitative Analysis Workgroup

The Quantitative Analysis Workgroup (QAW), which includes scientists from multiple sectors (i.e., government, industry, and academia), was charged with collecting and analyzing genotoxicity data from publicly available sources, as well as unpublished results available from the HESI committee’s member organizations. The QAW, in

collaboration with Health Canada, created a Microsoft-ACCESS database to store detailed dose–response information for a variety of genotoxicity and carcinogenicity end points for four pilot chemicals (i.e., the G4 chemicals)—EMS, MMS, ethyl nitrosourea (ENU), and methyl nitrosourea (MNU). Data were collected from studies examining the following endpoints: in vitro and in vivo micronucleus, in vitro mammalian gene mutations, in vivo and in vitro chromosomal aberrations, in vivo and in vitro DNA strand breaks (i.e., comet), in vivo Pig-a mutations, mutations in transgenic rodents (e.g., lacZ, cII), and carcinogenicity. Publicly available, peer-reviewed studies indexed in the National Library of Medicine’s PubMed and TOXNET databases, including the Chemical Carcinogenesis Research Information System (CCRIS) and GENETOX databases (<http://toxnet.nlm.nih.gov/index.html>) were consulted, along with the Carcinogenic Potency Database (CPDB) (<http://potency.berkeley.edu/index.html>), and US National Toxicology Program (http://ntp-apps.niehs.nih.gov/ntp_tox/index.cfm) databases.

Data Characteristics

Ideally, investigations of dose–response relationships are conducted using data from experimental designs that include more dose levels than normally used in traditional hazard assessment (i.e., 3 or 4 dose groups and a negative control). For quantitative analysis, it is preferable to include several doses not only in the effect zone, but also 3–5 at the lower end of the dose–response curve where no apparent increase over the background is expected for the endpoint of interest. The more dose groups in the range where the response is expected to be minimal, the more precise will be the ability to describe effectively the dose–response relationship. Uncertainty can also be reduced with the inclusion of more replicates at each dose level. Table I summarizes the key characteristics of the datasets that were analyzed. Only studies that tested a minimum of three dose levels, with a concurrent negative control, were included in the database.

When analyzing the collected data, due attention has been given to the fact that visual inspection and interpretation of the dose–response relationship can be distorted by the format used to plot the values [Jeffrey, 2009]. It is well known that the use of different graph scales with the same data can affect the perception of the dose–response relationship, and can lead to misinterpretation of the actual shape of the response. For example, linear-linear graphs of mutation dose–response data are often stated to be preferable depictions; however, these can give the mistaken impression that the underlying data are linear when in fact they are not. This mistake can happen when data from the critical low-dose range are compressed to the point that they cannot be adequately distinguished and visualized as to whether they are linear or not (Fig. 1a).

Further, a log-linear scale allows visualization of low dose values but gives the visual impression of nonlinearity with linear data (Fig. 1b), and a log-log plot will show an apparent threshold below a dose that induces a response that is smaller than an existing spontaneous value (Fig. 1c). Thus, the QAW recognized that various plots have different merits for displaying data for particular purposes, and it is important to define clearly “threshold responses” other than by their visual appearance on a Cartesian plot, and to use great care when reaching conclusions as to whether a dose–response displays a segmented or “hockey stick” response with a threshold dose (i.e., Td). For the aforementioned reasons, both linear and nonlinear responses need to be distinguished using statistical procedures to determine the best fit of the data.

METHODS

For this proof-of-concept project, dose–response data for induction of gene mutations (in vitro and in vivo) and micronuclei (in vitro and in vivo) by EMS and MMS were evaluated. Although these chemicals have been widely used as positive control test chemicals by genetic toxicologists for decades, experiments designed to examine their dose–response patterns, especially at low doses, have only become available during the last few years (e.g., see [Doak et al., 2007; Gocke and Wall, 2009; Pottinger et al., 2009; Bryce et al., 2010]).

Determination of POD Metrics

After identifying dose–response data that are useful for this exercise, it was necessary to identify the desirable dose–response descriptors that can be derived. The desired outputs included descriptors of the dose below the point at which a response cannot be detected, the initial slope of the dose–response curve, the maximum mutagenic response observed, and an index of cellular toxicity at conditions that produce the observed responses. These descriptors included the POD metrics outlined below.

No-Observed-Genotoxic-Effect-Level

NOGEL is defined as the highest tested dose for which no statistically significant increase in the incidence of the genotoxic effect is observed relative to an appropriate untreated control (i.e., background). Ideally, this would include specification of the statistical power of the test used to define the NOGEL; such a power calculation was considered for several of the datasets, but not reported for the analyses conducted here. Datasets were identified that included a suitable number of replicates (preferably three or more), with a suitable number of cells/targets (e.g., mostly >10,000 cells analyzed per dose for the analysis of MN) and with several doses resulting in genotoxic event frequencies similar to the NOGEL and solvent control. In fact, the in vitro micronucleus datasets from Doak et al. [2007] had several additional replicates at four doses surrounding the NOGEL, in order to reach 10,000 or more cells at this critical region, while LeBaron et al. [2008] evaluated >10,000 cells per replicate. Experimental datasets that fit these requirements were very limited and are presented in Tables I and II. Accordingly, one can argue that a NOGEL should in principle determine a point on the dose–response curve where the response is indistinguishable from background, and thus, doses below that point, may represent negligible concern. However, since the NOGEL is dependent on the experimental design (i.e., the selected doses), it has a measure of uncertainty associated with it.

TABLE I. Characteristics of datasets analyzed for NOGEL, and by dose-response models, and details on data-handling

Chemical	Effect	Gene target	VT or VV	Cell type / tissue	Units	# Dose levels	# Replicates	Treatment regime	Expression time ^a	Log-Transformed (NOGEL)	All doses in hockey-stick model	Reference
EMS	Gene Mutation	HPRT	<i>In vitro</i>	AHH-1	µg/ml	7 + control	3	24 h	13 d	Yes	Yes	Doak et al., 2007; Johnson et al., 2009
	Gene Mutation	LacZ	<i>In vivo</i>	Bone Marrow	mg/kg/d	7 + control	7	28 d/gavage	3 d	No	Yes	Gocke and Wall, 2009
	Gene Mutation	LacZ	<i>In vivo</i>	Liver	mg/kg/d	7 + control	7	28 d/gavage	3 d	No	Yes	Gocke and Wall, 2009
	Gene Mutation	LacZ	<i>In vivo</i>	GI Tract	mg/kg/d	7 + control	7	28 d/gavage	3 d	No	Yes	Gocke and Wall, 2009
	Micronucleus		<i>In vitro</i>	TK6	µg/ml	22 + control	4 + 8 control	24–30 h	0 h	Yes	No (supralinearity/saturation)	Bryce et al., 2010
	Micronucleus		<i>In vitro</i>	TK6	µg/ml	22 + control	4 + 8 control	24–30 h	0 h	Yes	No (supralinearity/saturation)	Bryce et al., 2010
MMS	Micronucleus		<i>In vitro</i>	AHH-1	µg/ml	9 + control	2 -5 + 4 control ^b	18 h	0 h	No	Yes	Doak et al., 2007; Johnson et al., 2009
	Micronucleus		<i>In vivo</i>	Bone Marrow	mg/kg/d	8 + control	6	7 d/gavage	1 d	No	No (supralinearity/saturation)	Gocke and Wall, 2009
	Gene Mutation	HPRT	<i>In vitro</i>	AHH-1	µg/ml	7 + control	3	24 h	13 d	Yes	Yes	Doak et al., 2007; Johnson et al., 2009
	Gene Mutation	T k	<i>In vitro</i>	L5178Y	µg/ml	8 + control	5	4 h	48 h	Yes	Yes	Pottenger et al., 2009
	Micronucleus		<i>In vitro</i>	AHH-1	µg/ml	12 + control	3 -5 + 6 control	18 h	0 h	No	Yes	Doak et al., 2007; Johnson et al., 2009
	Micronucleus		<i>In vitro</i>	TK6	µg/ml	22 + control	4 + 8 control	24–30 h	0 h	Yes	Yes	Bryce et al., 2010
	Micronucleus		<i>In vitro</i>	TK6	µg/ml	22 + control	4 + 8 control	24–30 h	0 h	No ^c	No (supralinearity/saturation)	Bryce et al., 2010
	Micronucleus		<i>In vivo</i>	Blood	mg/kg/d	5 + control	6 rats/dose	4 d/gavage	1 d	Yes	Yes	LeBaron et al., 2008

^aExpression time refers to time elapsed post-treatment before selection was initiated.

^bReplicate numbers varied, with 4 replicates collected for doses around the NOGEL, including the controls.

^cDataset still demonstrated heterogeneity after log-transformation, based on Bartlett's statistic.

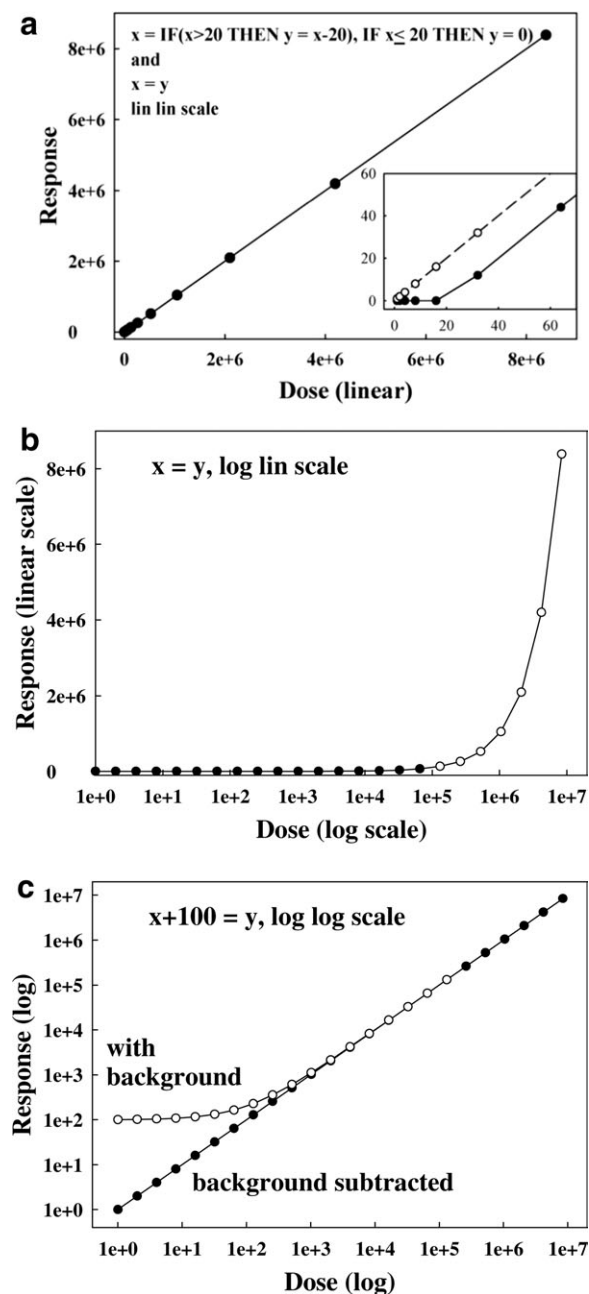


Fig. 1. Visual display of dose–response data leading to potential distortions. (a) Linear–linear scale. Inset shows expanded view of the low dose data and the differences between a linear response (open circles) and one in which there is a threshold (filled circles). (b) Log–linear scale. (c) Log–log scale. Lines of linear plots of data are indistinguishable when all the data are displayed.

Standard approaches were used to determine the NOGEL value. Data were evaluated using analysis of variance (ANOVA) followed by a one-sided Dunnett’s test at $\alpha = 0.05$ [Winer, 1977] using SPSS version 16.0.1. This analysis identified a dose where the responses were not statistically distinguishable from the response in the concurrent untreated control samples. The highest dose at which the end point of interest did not differ significantly from the background was identified as the NOGEL. The next highest dose was considered the lowest observed gen-

otoxic effect level (LOGEL). As shown in Table I, data were transformed (log or square root) where necessary, based on Bartlett’s statistic for variance homogeneity [Winer, 1977].

Threshold Effect Level

Definitions of threshold in the literature vary, and this term is often qualified by a descriptor such as “biological thresholds,” “apparent thresholds,” “operational threshold,” etc. [Pottenger and Gollapudi, 2010]. The Workgroup defined a “Threshold Effect Level” (Td) as a statistically identified dose (not limited to tested doses/exposures) below which the effect cannot be distinguished with the available data from the untreated background level, and above which it is possible to observe an increase in the effect above the untreated or negative control level. Td estimates can be derived by mathematical modeling of the dose–response data, and are based upon decision rules applied to the statistical analysis of the experimental response data at different doses. The bilinear hockey stick model [Lutz and Lutz, 2009] and broken stick model [Lynch et al., 2003] are specifically designed for comparing linear models with two parameters to threshold models with three parameters. Statistical rejection of a linear model would indicate that a more complex model, such as one including a Td, may be needed. Threshold values were calculated using the “hockey stick” package described by Lutz and Lutz [2009], with some prerequisites defined in Gocke and Wall [2009] and Johnson et al. [2009]. More specifically, the combined assessment involves four steps: (1) comparison of control groups; (2) rejection of the linear dose–response relationship (for the entire dose range); (3) acceptance of dose–response relationship with zero slope below the NOGEL; (4) application of threshold software developed by Lutz and Lutz [2009] to calculate threshold values including confidence limits [Gocke and Wall, 2009]. All threshold values determined in this study adhered to these steps, and the data analysis followed the four-stage statistical procedure outlined below. To avoid the issues stated previously regarding log–linear and log–log plots, the four stages were carried out on untransformed data, unless otherwise stated. Stage 1 involved a one-way ANOVA for a dose-related effect (SPSS version 16.0.1). If they did not show significant differences, control groups could be cumulated. Stage 2 involved comparison of linear and quadratic models using the coefficient of determination (R^2 , SPSS version 16.0.1). The F distribution was then used to calculate P values in Microsoft Excel 2007. Stage 3 involved determination of NOGEL and LOGEL values using a one-sided Dunnett’s test on either untransformed or log-transformed data (SPSS version 16.0.1). Linear and quadratic models were then compared at the NOGEL and below in the same way as described for Stage 2. Data that had a flat or zero dose–response slope at the NOGEL and below were then suitable for bilinear or hockey stick analysis. Stage 4 involved a comparison of linear versus hockey stick models using the R software package (version 12.2) recommended by Lutz and Lutz [2009]. Parameters, y -intercept, Td, and slope above Td were estimated for best fit of a hockey stick model by minimizing the residual sum of squares. Confidence intervals (CI) were estimated for all parameters using an F distribution [Lutz and Lutz, 2009]. If the 95% CI of the derived Td value does not encompass zero, the model is considered a good fit to the data. This is a key feature of the hockey stick approach, and Lutz and Lutz [2009] note that if the 95% CI of the Td includes zero, it is not statistically possible to distinguish the fit of a hockey-stick response from a linear response.

Benchmark Dose

The BMD approach is based on mathematical modeling of dose–response data, and has been proposed as an improvement on the NOAEL (no observed adverse effect level) approach [Crump, 1984; Slob, 1999, 2007]. The approach has been widely used in other fields of toxicology to define POD values for both cancer and non-cancer endpoints. The BMD approach estimates a dose (i.e., the BMD) that produces some predetermined, and presumably biologically relevant, increase in the

TABLE II. NOGEL, Td-LCI, and BMDL₁₀ Values for Gene Mutation (Gene Mut) and Micronucleus (MN) Endpoints Induced by MMS or EMS (cmpd) In Vitro (vt) and In Vivo (vv)

Cmpd	Effect	VT/VV	System	Units	NOGEL	Td-LCI	BMDL-10	Data source	
EMS	Gene Mut	vt	AHH-1	µg/ml	1	0.95	1.08	Doak et al., 2007; Johnson et al., 2009	
		vv	Bone Marrow	mg/kg bw	50	21.46	9.29	Gocke and Wall, 2009	
	Gene Mut	vv	Liver	mg/kg bw	50	25.67	41.00	Gocke and Wall, 2009	
		vv	GI Tract	mg/kg bw	25	12.97	12.23	Gocke and Wall, 2009	
		vt	TK6	µg/ml	1.17	0.74	0.54	Bryce et al., 2010	
	MN	vt	TK6	µg/ml	6.25	3.92	2.38	Bryce et al., 2010	
		vt	AHH-1	µg/ml	1.3	0.87	1.29	Doak et al., 2007; Johnson et al., 2009	
		vv	Bone Marrow	mg/kg bw	80	56.66	58.68	Gocke and Wall, 2009	
	MMS	Gene Mut	vt	AHH-1	µg/ml	1	0.86	0.56	Doak et al., 2007; Johnson et al., 2009
			vt	L5178Y	µg/ml	1.1	0.52	0.52	Pottenger et al., 2009
MN		vt	AHH-1	µg/ml	0.8	0.14	0.54	Doak et al., 2007; Johnson et al., 2009	
		vt	TK6	µg/ml	0.63	0.48	0.19	Bryce et al., 2010	
		vt	TK6	µg/ml	0.47	0.44	0.11	Bryce et al., 2010	
MN		vv	Blood	mg/kg bw	5	14.07	1.74	LeBaron et al., 2008	

AHH-1 and TK6 are cell lines of human origin; L5178Y is a murine lymphoma cell line.

response over control (i.e., the benchmark response or BMR). The approach employs mathematical dose–response modeling that takes factors such as sample size and shape of the curve into account [Crump, 1984], and a small measurable effect (i.e., BMR) and a critical effect dose (i.e., BMD) are estimated without the need for data transformation [Slob, 1999, 2007; EPA, 2000, 2010]. The BMR refers, for continuous endpoints, to a percentage change compared with background response (i.e., negative control) as estimated by the fitted model (e.g., 5% or 10% change), whereas for quantal endpoints, the BMR is a specific increase in incidence compared with background incidence (e.g., the additional risk or extra risk). The lower limit of the one-sided 95% CI on the BMD is termed the BMDL. Thus, a BMDL₁₀ refers to the estimate of lower 95% CI of a dose that produces a 10% increase over the fitted background level for continuous endpoints, and 10% extra risk for quantal endpoints. The BMDL is often considered an adequate POD for the extrapolation of the dose–response data below the range of available data. Genotoxicity dose–response data can be modeled using BMD methodology since both dichotomous (quantal) and continuous responses can be analyzed, provided that the data set shows a dose-related trend. BMD estimates differ from NOGEL or Td values in that they provide an estimate of the size of the effect associated with the estimated dose.

BMDL₁₀ values were derived using the dose–response modeling software package PROAST, developed at the National Institute for Public Health and the Environment (RIVM) in the Netherlands (www.proast.nl; versions 26.4, 28.1 and 28.3 [Slob, 2002]). This program is analogous to the United States Environmental Protection Agency’s BMDS software [EPA, 2010], and these two agencies are working to achieve consistency between the two software packages and their algorithms. The models used were the exponential models recommended by the European Food Safety Authority [EFSA, 2009]. The exponential family of nested dose–response models used in PROAST is illustrated in Supporting Information Figure 1. The shapes of the curves can vary depending on (1) background levels of the endpoint, (2) the relative efficacy of dose; and (3) the maximum effect (relative to 1) (Slob 2002). Model selection was performed using the log-likelihood ratio test that assesses whether a statistically significant improvement in the fit is achieved by adding additional parameters. The model with additional parameters is only accepted if the difference in log-likelihoods exceeds the critical value at $P = 0.05$. This is automatically performed in PROAST by selecting the “automatic selection of optimal model from nested family” option. The

critical differences in log-likelihood values between two nested models are provided in Supporting Information Table 1. A log-likelihood value is also provided for the “full” model, which is simply the set of the geometric means of the observations at each dose (together with the residual variance). The log-likelihood ratio test can be used to compare the selected model with the full model using a goodness-of-fit test. The model is accepted when the log-likelihood value of the fitted model is significantly better than that of the full model. The BMD₀₅ and BMD₁₀ with their associated lower (BMDL) confidence limits are then derived from the selected model.

RESULTS

The main objective of this exercise was to evaluate critically different approaches for the analysis of genotoxicity dose–response data, and to develop a “tool box” of approaches to be used for data interpretation and risk assessment. As a starting point for quantitative analysis of the genotoxicity data, and to evaluate and refine the currently available quantitative approaches, dose–response data for the induction of gene mutations and micronuclei by the two direct-acting alkylating agents MMS and EMS were examined. The datasets used for this analysis included data-rich in vitro and in vivo studies with multiple dose measurements (see Table I).

Table II summarizes the POD metrics (i.e., NOGEL, Td-LCI (lower confidence interval), and BMDL₁₀) determined for each dataset examined. The BMDL₁₀ was selected because, for several of the datasets, it was not possible to determine a BMDL₀₅ value. Some of these values differ from previously published values because the dose–response models were applied as described in the Methods section. For example, a one-sided Dunnett’s test was used to determine the NOGEL, whereas some of

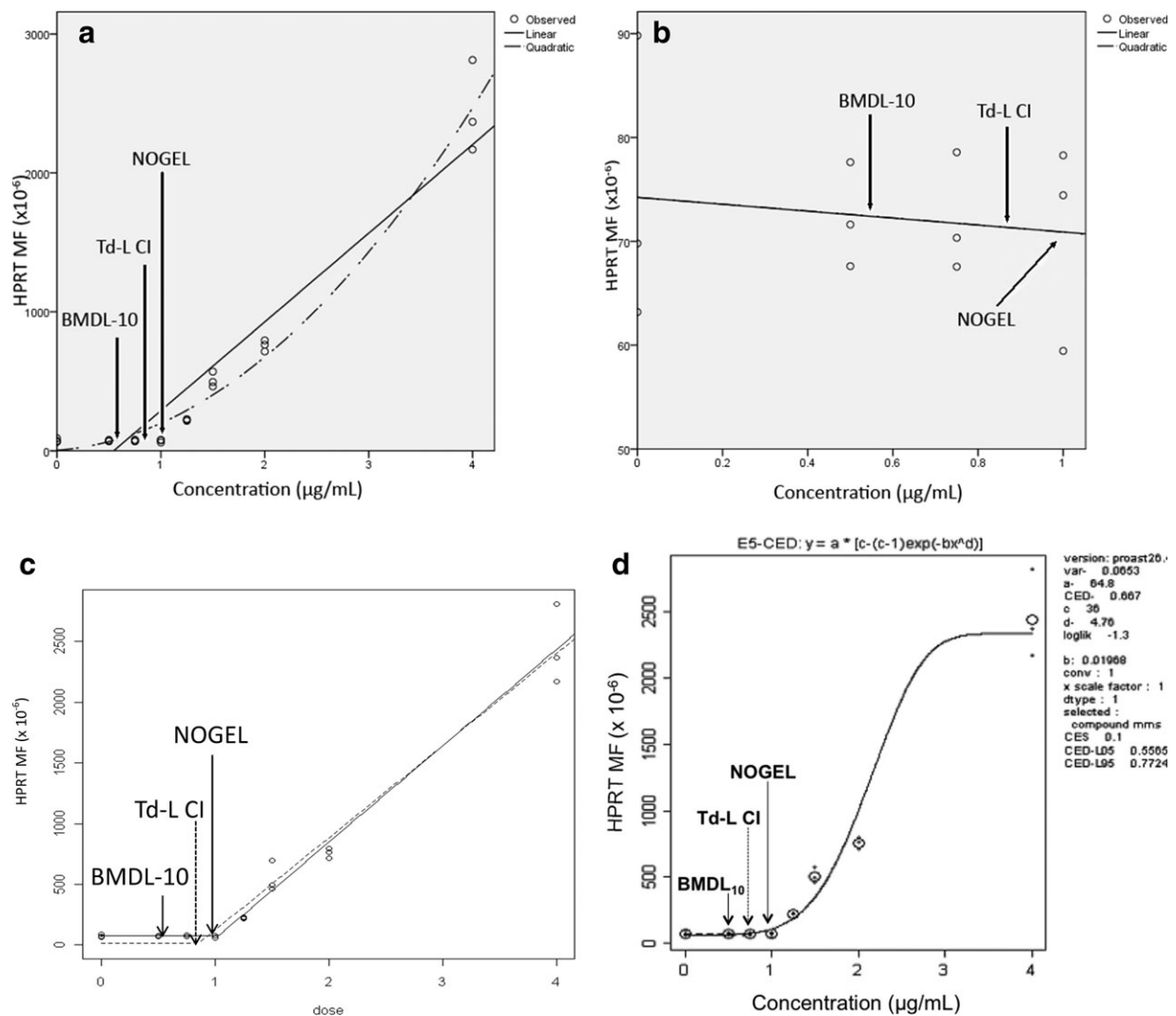


Fig. 2. Dose–response modeling results for HPRT gene mutations induced by MMS *in vitro* in AHH-1 cells. (a) Linear versus quadratic modeling for the entire dose–response. Quadratic model gave the best fit ($P < 0.05$). (b) Linear versus quadratic modeling at the NOGEL and below. Linear model gave the best fit ($P > 0.05$) with negative slope -3.33 ± 12.80 that contains zero ($P > 0.05$). (c) Td-LCI dose–response modeling. The three parameters NOGEL, TD-LCI, and BMDL₁₀ are shown in each graph for comparison. For (c), the solid line shows the best fit of the hockey stick function, and the dotted line shows the hockey stick with the inflection point at the lower 95% confidence limit of the threshold (i.e., Td-LCI) [Lutz and Lutz, 2009]. For (d), CED is the critical effect dose or BMD; CES is the critical effect size, in this case 10% or 0.1; CED-L05 is the lower confidence interval of the BMD or BMDL; CED-L95 is the upper confidence interval of the BMD or the BMDU. For this example, model E5-CED had the highest log-likelihood and provided the best representation of the data.

the previously published values used a two-sided Dunnett’s test. Similarly, some of the Td-LCI values calculated here, as presented in Table II, do not match the published values because top doses were occasionally discounted for hockey stick analysis (i.e., Stage 4) due to supra-linearity (saturation, see Table I), or an increasing exponential dose–response above the Td [Lutz and Lutz, 2009]. This was not the case for BMD analysis where all doses were included in the analyses.

Dose–response functions for induction of HPRT gene mutations by MMS in AHH-1 cells *in vitro* are shown in Figures 2a–2d, as examples of how the data were analyzed. Figures 2a and 2b show the linear versus quadratic

modeling outputs produced using SPSS v16.0.1. These constitute the aforementioned prerequisites (i.e., Stages 2 and 3) that precede analysis with the hockey stick approach. Figures 2c and 2d show the output graphs for the Td-LCI (using the R hockey stick software [Lutz and Lutz, 2009]) and BMD dose–response (using PROAST modeling results), respectively. A comparison of the modeling results across datasets identified several interesting points (Table II). All the *in vitro* NOGEL values were of a similar order of magnitude for gene mutations and micronuclei (~ 0.5 – $1 \mu\text{g/ml}$ for MMS and 1 – $6.25 \mu\text{g/ml}$ for EMS). The calculated Td-LCI values were numerically similar to the calculated NOGEL values for each

dataset, in vivo and in vitro, except for the in vivo MMS micronucleus data where the Td-LCI was about 3-fold higher than the NOGEL (Table II). Although the BMDL₁₀ values tended to be the lowest across all datasets, for both EMS and MMS, all three parameters fell within the same order of magnitude, with a 6-fold maximum difference among the in vitro datasets and an 8-fold maximum difference among the in vivo datasets.

DISCUSSION

Toxicologists evaluate many different types of adverse events, including reproductive, developmental, organ toxicities, and carcinogenic effects. Processes believed to be driven by the interaction of toxicants with non-DNA cellular constituents are often evaluated by assuming a threshold dose below which no effect is expected. Toxicologists generally acknowledge and support the existence of homeostatic mechanisms and employ the supposition that adverse effects occur only when these mechanisms are saturated or overloaded [Piersma et al., 2011]. As a result, a NOEL can be derived from the dose–response information generated for safety assessment studies. Genotoxic mechanisms, on the other hand, have historically been considered as being based upon a stochastic process, and the paradigm employed assumes low dose linearity for induced effects, and the absence of a response threshold.

When selecting the dose levels for toxicology studies that do not include genetic endpoints, investigators emphasize the importance of identifying a NOEL. These NOELs, or LOEL when a NOEL is not available, are used to derive acceptable or tolerable human exposure levels (e.g., reference dose or RfD) by applying one or more uncertainty factors to the calculated POD (i.e., NOEL) that accounts for data uncertainties. For example, the US EPA's Integrated Risk Information System (IRIS) defines an RfD as “an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure for a given duration to the human population (including susceptible subgroups) that is likely to be without an appreciable risk of adverse health effects over a lifetime” [EPA, 2011]. RfD values are derived from BMD, NOEL, LOEL, or other suitable POD. The most commonly used POD values are the NOAELs or specified BMDs (e.g., BMD₁₀).

The three dose–response descriptors reported here were selected because they each represent a different approach. NOGEL values represent the POD metric that is analogous to the statistical NOEL derived from standard toxicology data. As such, they rely entirely on tested doses and do not evaluate the entire dose–response. By definition, NOGELs are strongly affected by the experimental design and the statistical methodology employed. Given

that there is no context for adversity for in vitro genetic toxicity data, and no disease endpoint associated with in vivo genetic toxicology data on surrogate genes or non-heritable endpoints such as micronuclei in bone marrow cells, the term NOGEL provides a more precise descriptor than the standard toxicology terms of NOEL or no-observed-adverse-effect-level (NOAEL) for genotoxicity data. In contrast to the NOGEL, the Td and BMD approaches involve determination of interpolated POD metrics, calculated using the entire dataset, and generated using statistical approaches that also provide measures of uncertainty such as confidence limits [Crump, 1984; Slob, 2002; Lutz and Lutz, 2009]. Confidence limits associated with Td and BMD estimates, such as Td-LCI and BMDL₁₀ values, represent the lower confidence limits of the interpolated POD metrics, and as such, they are inherently more conservative than NOGELs.

Determination of Td-LCI values, which represent the lowest estimate of the inflection point beyond which a response begins to increase significantly, has the most stringent data recommendations (e.g., three doses above and three doses below the estimated Td-LCI, see [Pottenger and Gollapudi, 2010]). The BMD analysis, intended for datasets with three doses plus control, as is often seen with bioassays, is capable of analyzing datasets with more doses, but does not have the stringent requirements associated with Td analysis. Although both analyses can be influenced by high dose data, the Td value is more likely than the BMD value to be influenced by high-dose responses indicative of sub- or supra-linearity. Determination of BMDL₁₀ values for genotoxic effects has the additional advantage that the POD metric can be readily compared to BMDL₁₀ values calculated for other toxicological endpoints including carcinogenicity [Hernandez et al., 2011]. In addition, the ratio of BMDU₁₀ to BMDL₁₀ values provides information on the uncertainty surrounding the BMD estimate.

The three approaches examined in this work (i.e., NOGEL, Td-LCI, and BMD) demonstrate the utility of using quantitative methods to describe the nature of the dose–response curves for the induction of micronuclei and gene mutations after exposure to MMS and EMS in vitro and in vivo, and moreover, derive statistically defensible POD metrics. The analyses suggest that the different POD metrics determined for these chemicals are within the same order of magnitude, with more variability observed for the in vivo assays. Although the magnitudes of the metrics investigated are similar, the different quantitative approaches employed are complementary because they each have advantages and limitations. For example, BMDL can be used with a limited number of doses, although it is preferable that more dose levels are available, whereas Td-LCI determination requires a large number of doses below and above the threshold effect level. The data-rich studies available in the G4 database are

currently being employed to determine the required numbers of doses above and/or below the point of departure for effective determination of Td-LCI. As noted, the BMDL and Td-LCI approaches have the distinct advantage, relative to the experimental design-limited NOGEL approach, of taking into account the variability in the data, and allowing the calculation of confidence intervals.

The assumption that no level of exposure to a genotoxic chemical is risk-free arose from the supposition that even a single molecule could produce a mutagenic lesion in DNA that could lead to a pre-cancerous cell that could proliferate. This assumption has its basis in the “one-hit” model of radiation mutagenesis/carcinogenesis, and the subsequent demonstration of linear dose–responses for DNA adducts induced by alkylating agents [Turteltaub et al., 1993; Swenberg et al., 2008]. Although more recent data demonstrate nonlinear dose–response for exogenously-induced DNA adducts [Swenberg et al., 2011], this paradigm continues to play a dominant role in human health risk assessment, and in regulatory decisions for mutagens and mutagenic carcinogens. The principal exception to this low-dose linear default assumption is for genotoxic chemicals that act via a non-DNA target, e.g., the proteins of the mitotic spindle (Elhajouji et al., 1997). In these cases, limited guidance exists for the quantitative use of dose–response data for human health risk assessment and ensuing regulatory decisions [FDA, 2006; Thybaud et al., 2007; ICH, 2011].

Because genetic damage in a sentinel cell type is not in itself a disease outcome, it is more difficult to calculate potential incremental health risks from incremental increases over the spontaneous mutation frequency in the same way that this is done for an apical endpoint such as cancer. Further, high-dose information has generally been used to generate a binary determination of whether or not a chemical is genotoxic, and regulatory decisions are often made on the basis of this dichotomous information. This was originally considered suitable because mutagenic potential was thought to be a rare property of only a few chemicals to which exposure could simply be prevented. However, with accumulating experience, it has become apparent that many chemicals, both synthetic and natural, can induce genetic damage [Galloway et al., 1987; Moore and Brock, 1988; Seeberg et al., 1988; Kultz and Chakravarty, 2001; Charles et al., 2002; Claxton et al., 2010].

The Quantitative Workgroup recognizes that a quantitative approach is needed to help support rational risk-based decisions regarding agents that induce genetic alterations. In common with other toxicological endpoints, there are different mathematical methods to characterize dose–response data and derive POD metrics. In order to identify an exposure level associated with a minimal risk of inducing genetic damage, it is necessary to define an exposure that either fails to increase the existing level of the toxic event of interest (e.g., mutant frequency for muta-

genic chemicals) by an agreed-upon minimal level over control or background values, or fails, through the application of an appropriate experimental and mathematical method, to induce a specified absolute frequency of the toxic event. The acceptable/tolerable increase can be defined relative to the existing spontaneous frequency or specified as an absolute frequency or rate.

The Workgroup also recognizes that zero exposure to hazardous substances is not always possible, and that it is timely to explore ways to maximize information that can be obtained from detailed dose–response data from genotoxicity studies. It is clear that organisms experience a substantial endogenous level of DNA damage (=30,000 DNA lesions/cell; [Swenberg et al., 2011]), and a spontaneous mutation rate that is not zero. Given the ubiquitous presence of background DNA damage, human health risk assessment for genotoxins and mutagens should consider this natural background of genetic damage, and subsequently, using quantitative methods such as those described in this study, attempt to minimize/prevent exposures that add to the mutational burden of a given population.

In conclusion, this article explored the applicability of three different approaches to assess genotoxicity dose–response data and derived POD metrics such as BMD, Td, and NOGEL. The BMD and Td estimates have a quantifiable measure of uncertainty associated with them and hence are recommended as initial acceptable approaches if adequate data are available, whereas the NOGEL approach parallels the commonly-used statistical NO(A)EL. Given that the BMD approach is already widely used in the risk assessment community for other toxicology endpoints, it may become the preferred approach when analyzing genotoxicity data, with the Td estimation serving as a useful adjunct for more refined analysis in certain circumstances. Moreover, the findings support the use of the lower confidence limit of a 10% response as an adequate POD and BMR for genotoxicity data, when the BMD approach is utilized.

Ongoing work is expanding the analyses to include additional chemicals in the G4 database (e.g., ENU and MNU) and additional endpoints (e.g., DNA strand breaks as measured in a Comet assay, DNA adducts), and moreover, evaluating the ability to use quantitative data from experimental models to calculate POD metrics and use the metrics for risk assessment and determination of acceptable human exposure levels. In addition, ongoing analyses are developing approaches for quantitative comparisons of *in vitro* and *in vivo* responses. Finally, ongoing work is extending the comparative analysis of different statistical methods and different PODs, with the ultimate goal of producing a standard operating protocol to guide effective quantitative analysis of genetic toxicity dose–response data. It is hoped that application of quantitative approaches, such as those used here, will help to

bring the field of genetic toxicology closer to other fields of toxicology. It is conceivable that the POD values along with uncertainty factors or a margin of exposure approach, and a mode-of-action analysis, can be used to help set acceptable/tolerable exposure levels or reference doses for genotoxic materials.

AUTHOR CONTRIBUTION

Drs. Bhaskar Gollapudi and Véronique Thybaud provided leadership for the IVGT QAW during database construction, data collection, and data analysis, and led discussions of the group on results interpretation. Dr. Gollapudi is also the lead author of the paper. Drs. Lya Hernandez, George Johnson, and Lynn Pottenger performed and contributed the quantitative approaches for analyzing dose–response curves. Dr. Elizabeth Julien was the technical lead for construction of the database and data entry. Drs. Kerry Dearfield, Alan Jeffrey, David Lovell, Jim MacGregor, Martha Moore, Jan van Benthem, Paul White, and Errol Zeiger contributed their perspectives and expertise on the interpretation of the analyses and the results obtained. Dr. James Kim provided logistical, organizational and editorial support for the project. All authors contributed to addressing the comments from the reviewers.

REFERENCES

- Bryce SM, Avlasevich SL, Bemis JC, Phonethpswath S, Dertinger SD. 2010. Miniaturized flow cytometric in vitro micronucleus assay represents an efficient tool for comprehensively characterizing genotoxicity dose–response relationships. *Mutat Res* 703:191–199.
- Charles GD, Linscombe VA, Tornesi B, Mattsson JL, Gollapudi BB. 2002. An in vitro screening paradigm for extracts of whole foods for detection of potential toxicants. *Food Chem Toxicol* 40:1391–1402.
- Claxton LD, Umbuzeiro Gde A, DeMarini DM. 2010. The Salmonella mutagenicity assay: The stethoscope of genetic toxicology for the 21st century. *Environ Health Perspect* 118:1515–1522.
- Crump KS. 1984. A new method for determining allowable daily intakes. *Fundam Appl Toxicol* 4:854–871.
- Dearfield KL, Auletta AE, Cimino MC, Moore MM. 1991. Considerations in the U.S. Environmental Protection Agency's testing approach for mutagenicity. *Mutat Res* 258:259–283.
- Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS, Hester S, Kim JH, Kirkland D, Levy DD, Lorge E, Moore MM, Ouedraogo-Arras G, Schuler M, Suter W, Sweder K, Tarlo K, van Benthem J, van Goethem F, Witt KL. 2011. Follow-up actions from positive results of in vitro genetic toxicity testing. *Environ Mol Mutagen* 52:177–204.
- Doak SH, Jenkins GJ, Johnson GE, Quick E, Parry EM, Parry JM. 2007. Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens. *Cancer Res* 67:3904–3911.
- EFSA (European Food Safety Authority). 2009. Use of the benchmark dose approach in risk assessment: Guidance of the Scientific Committee. *EFSA J* 1150:1–72. Also available at: <http://www.efsa.europa.eu/en/efsajournal/doc/1150.pdf>. Accessed July 19, 2012.
- Elhajouji A, Tibaldi F, Kirsch-Volders M. 1997. Indication for thresholds of chromosome non-disjunction versus chromosome lagging induced by spindle inhibitors in vitro in human lymphocytes. *Mutagenesis* 12:133–140.
- EPA (U.S. Environmental Protection Agency). 2000. Benchmark dose technical guidance document. Available at: http://www.epa.gov/ncea/www1/pdfs/bmds/BMD-External_10_13_2000.pdf. Accessed July 19, 2012.
- EPA (U.S. Environmental Protection Agency). 2010. Benchmark dose software (BMDS). Available at: <http://www.epa.gov/ncea/bmds/>. Accessed July 19, 2012.
- EPA (U.S. Environmental Protection Agency). 2011. IRIS Glossary/ Acronyms and Abbreviations. Available at: http://www.epa.gov/iris/help_glosshtm#r. Accessed July 19, 2012.
- FDA (U.S. Food and Drug Administration). 2006. Guidance for Industry and Review Staff Recommended Approaches to Integration of Genetic Toxicology Study Results, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), January 2006. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm079257.pdf>. Accessed July 19, 2012.
- Flamm WG, Valcovic LR, deSerres FJ, D'Aguanno W, Fishbein L, Green S, Mallig HV, Mayer V, Prival M, Wolff G, Zeiger E. 1977. Approaches to determining the mutagenic properties of chemicals: risk to future generations. *J Environ Pathol Toxicol* 1:301–352.
- Galloway SM, Armstrong MJ, Reuben C, Colman S, Brown B, Cannon C, Bloom AD, Nakamura F, Ahmed M, Duk S, Rimpo J, Margolin BH, Resnick MA, Anderson B, Zeiger E. 1987. Chromosome aberrations and sister chromatid exchanges in Chinese hamster ovary cells: Evaluations of 108 chemicals. *Environ Mol Mutagen* 10 (Suppl 10):1–175.
- Gocke E, Wall M. 2009. In vivo genotoxicity of EMS: Statistical assessment of the dose response curves. *Toxicol Lett* 190:298–302.
- Gollapudi BB, Thybaud V, Kim JH, Holsapple M. 2011. Strategies for the follow-up of positive results in the in vitro genotoxicity assays—An international collaborative initiative. *Environ Mol Mutagen* 52:174–176.
- Hernandez LG, Slob W, van Steeg H, van Benthem J. 2011. Can carcinogenic potency be predicted from in vivo genotoxicity data? A meta-analysis of historical data. *Environ Mol Mutagen* 52:518–528.
- ICH (International Conference on Harmonisation). 2011. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use, ICH Harmonised Tripartite Guideline ICHS2(R1): Guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use, November 9, 2011. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S2_R1/Step4/S2R1_Step4.pdf.
- Jeffrey AM. 2009. Thresholds: absolute, practical, pragmatic, apparent and virtual. *Toxicol Lett* 191:365; author reply 366–367.
- Johnson GE, Doak SH, Griffiths SM, Quick EL, Skibinski DOF, Zair ZM, Jenkins GJ. 2009. Non-linear dose-response of DNA-reactive genotoxins: Recommendations for data analysis. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 678:95–100.
- Kultz D, Chakravarty D. 2001. Hyperosmolality in the form of elevated NaCl but not urea causes DNA damage in murine kidney cells. *Proc Natl Acad Sci USA* 98:1999–2004.
- LeBaron MJ, Kan HL, Geter DR, Pottenger LH, Zhang F, Schisler MR, Bartels MJ, Gollapudi BB. 2008. Non-linear low dose response in methyl methanesulfonate (MMS)-treated rats as measured by multiple endpoints. 47th Annual Meeting of the Society of Toxicology, Abstract no. 1429. *Toxicologist* 62: 293.

- Lutz WK, Lutz RW. 2009. Statistical model to estimate a threshold dose and its confidence limits for the analysis of sublinear dose-response relationships, exemplified for mutagenicity data. *Mutat Res* 678:118–122.
- Lynch A, Harvey J, Aylott M, Nicholas E, Burman M, Siddiqui A, Walker S, Rees R. 2003. Investigations into the concept of a threshold for topoisomerase inhibitor-induced clastogenicity. *Mutagenesis* 18:345–353.
- Lynch AM, Sasaki JC, Elespuru R, Jacobson-Kram D, Thybaud V, De Boeck M, Aardema MJ, Aubrecht J, Benz RD, Dertinger SD, Douglas GR, White PA, Escobar PA, Fornace A Jr, Honma M, Naven RT, Rusling JF, Schiestl RH, Walmsley RM, Yamamura E, van Benthem J, Kim JH. 2011. New and emerging technologies for genetic toxicity testing. *Environ Mol Mutagen* 52:205–223.
- McCann J, Ames BN. 1976. Detection of carcinogens as mutagens in the Salmonella/microsome test: Assay of 300 chemicals: Discussion. *Proc Natl Acad Sci USA* 73:950–954.
- Moore MM, Brock KH. 1988. High concentrations of sodium chloride induce a “positive” response at the TK locus of L5178Y/TK+/- mouse lymphoma cells. *Environ Mol Mutagen* 12:265–268.
- Piersma AH, Hernandez LG, van Benthem J, Muller JJ, van Leeuwen FX, Vermeire TG, van Raaij MT. 2011. Reproductive toxicants have a threshold of adversity. *Crit Rev Toxicol* 41:545–554.
- Pottenger LH, Gollapudi BB. 2010. Genotoxicity testing: Moving beyond qualitative “screen and bin” approach towards characterization of dose-response and thresholds. *Environ Mol Mutagen* 51:792–799.
- Pottenger LH, Schisler MR, Zhang F, Bartels MJ, Fontaine DD, McFadden LG, Bhaskar Gollapudi B. 2009. Dose-response and operational thresholds/NOAELs for in vitro mutagenic effects from DNA-reactive mutagens, MMS and MNU. *Mutat Res* 678:138–147.
- Seeberg AH, Mosesso P, Forster R. 1988. High-dose-level effects in mutagenicity assays utilizing mammalian cells in culture. *Mutagenesis* 3:213–218.
- Slob W. 1999. Thresholds in toxicology and risk assessment. *Int J Toxicol* 18:259–268.
- Slob W. 2002. Dose-response modeling of continuous endpoints. *Toxicol Sci* 66:298–312.
- Slob W. 2007. What is a practical threshold? *Toxicol Pathol* 35:848–849.
- Swenberg JA, Fryar-Tita E, Jeong YC, Boysen G, Starr T, Walker VE, Albertini RJ. 2008. Biomarkers in toxicology and risk assessment: Informing critical dose-response relationships. *Chem Res Toxicol* 21:253–265.
- Swenberg JA, Lu K, Moeller BC, Gao L, Upton PB, Nakamura J, Starr TB. 2011. Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci* 120 (Suppl 1):S130–S145.
- Thybaud V, Aardema M, Clements J, Dearfield K, Galloway S, Hayashi M, Jacobson-Kram D, Kirkland D, MacGregor JT, Marzin D, Ohyama W, Schuler M, Suzuki H, Zeiger E. 2007. Strategy for genotoxicity testing: hazard identification and risk assessment in relation to in vitro testing. *Mutat Res* 627:41–58.
- Turteltaub KW, Frantz CE, Creek MR, Vogel JS, Shen N, Fultz E. 1993. DNA adducts in model systems and humans. *J Cell Biochem Suppl* 17F:138–148.
- Winer BJ. 1977. *Statistical Principles in Experimental Design*, 2nd ed. New York:McGraw-Hill.

Accepted by—
M. Shelby