MUT/2017/02

COMMITTEE ON MUTAGENICITY
OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT (COM)

Quantitative approaches to the assessment of genotoxicity data II

At the COM meeting in October 2016, Members were given a presentation on quantitative assessment of genetic toxicology data by Dr George Johnson (University of Swansea) who outlined work he is currently involved in, including a project with The Quantitative Analysis Workgroup (QAW) of the Genetic Toxicology Technical Committee (his presentation is appended).

A paper was also considered (MUT/2016/07) which outlined: current risk assessment strategies, such as the margin of exposure (MOE) approach and the threshold of toxicological concern (TTC); the concept of nonlinear/threshold responses in genetic toxicology; comparisons of genotoxic and carcinogenic potencies; and some general publications on the development of quantitative approaches, including conclusions of the IWGT working groups discussions on the topic. Whilst there was some discussion following the presentation, it was agreed that this topic should be further addressed at future meetings. Aspects that could be considered in terms of risk assessment included; the most suitable test systems and endpoints (e.g. gene mutations or micronuclei), what tissues should be analysed, and what critical effect size (CES) or benchmark response (BMR) values were appropriate for each genotoxicity endpoint.

This paper is presented as a general overview which attempts to focus on a number of key areas with view to furthering discussions and gathering Members opinions on how to progress this topic (eg closer examination of datasets, statement). The approaches are complex and many of the key publications are too broad and discursive to summarise. Therefore several of these are appended to facilitate Members consideration of the topic, general discussions and when addressing the questions.

Secretariat

February 2017

COMMITTEE ON MUTAGENICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT (COM)

Quantitative approaches to the assessment of genotoxicity data II

Introduction

1) The quantitative analysis of dose-response data from *in vivo* genotoxicity studies has potential for use in chemical risk assessment, particularly in a margin of exposure (MOE) approach, similar to that utilised in the interpretation of carcinogenicity data. It is also shows promise as a tool in genotoxic potency ranking and further testing prioritisation. The use of benchmark dose (BMD) software to model dose response data for a wide range of toxicological endpoints is widely accepted (EFSA 2009; Davis et al 2011). It has been widely acknowledged that the derived BMD metric has several advantages over the no adverse effect level (NOAEL) for use in chemical risk assessment or for deriving MOE's.

2) At the October 2016 meeting, Members concluded that over the last few years there has been an improvement in the quality of available *in vivo* genotoxicity data (e.g. more endpoints, tissues and dose groups) and that developments in dose response modelling allow genotoxicity data to be analysed quantitatively rather than only qualitatively.

3) Following the initial review at the meeting in October 2016, a number of areas have been identified which are considered important for the COM to address when evaluating the practicality of applying a quantitative approach to risk assessment. These are:

- dose responses and critical effect size;
- the implication of using different genotoxicity endpoints or tissues;
- the importance of study designs;
- the potential usefulness of genotoxicity BMD's in carcinogenicity risk assessment.
- Different computer software (RIVM vs EPA) ;

4) *In vivo* studies are the primary focus of this paper though some *in vitro* analysis are included. Each aspect is presented with questions for consideration.

Dose response modelling and BMD metrics

5) Dose response modelling and selection of an appropriate benchmark response (BMR) metric were briefly introduced in MUT/2016/07 and are examined in a number of the appended publications. What constitutes a critical effect size (CES) is also discussed.

6) The IWGT publications on the quantitative use of genotoxicity data make a number of evaluations and observations about determining dose responses and how these can be best utilised in broad risk assessment strategies (MacGregor et al 2015a - annex). Desirable characteristics include;

> *The method must be well defined and robust, that it is conservative, employing approaches which inherently account for uncertainty in the data underestimating the actual risk (POD statistically lower bound) and that it has an interpretable biological meaning, the undesirable effect can be related to human disease. .*

7) It is indicated that, for use in risk assessment, it should be possible to relate the POD' to acceptable exposure levels extrapolating from data which includes the use of mode of action (MOA) and mechanistic information if available (i.e. so a threshold mechanism if demonstrated can be taken into account). It was also noted that BMD_{10} for quantal and continuous data will be substantially different; for genotoxicity data this represents a percent increase of the spontaneous incidence as opposed to an absolute increase of 10% tumour incidence in a carcinogenicity study.

8) The conclusions on the use of modelling methods and metrics were as follows:

> *Noteworthy conclusions and comments, based on critical comparisons of the methods and metrics, include:*
>
> *(1) BMD modelling almost always yields a good fit to the data,*
>
> *(2) The $BMDL_{10}$ (the lower confidence limit on the benchmark response rate of 10% over background), derived from the best fitting model, generally provided a conservative (lower) value relative to the NOGEL and the BPDL,*
>
> *(3) The $BMDL_{10}$, though recently employed in studies reported in the literature, is an arbitrary choice of minimal response based on the presumption that a BMR of 10% of the spontaneous rate is a minimal increase in response that is close enough to the range of observable responses to be estimated with reasonable accuracy. Since most genotoxicity assays currently employed do not have the statistical sensitivity to detect less than a doubling of the spontaneous rate (i.e., a 100% increase in the spontaneous rate), the choice of a BMR of 10% results in a PoD that is approximately 1/10 (or less for certain assays) of the detectable NOGEL,*
>
> *(4)The BMDL approach is very flexible, can be readily applied to a wide range of datasets, is minimally affected by dose spacing, and requires only three treatment levels. Dose spacing is more critical for determination of the NOGEL and BPD, and BPD analysis requires approximately six treatment groups,*

(5) The $BMDL_{10}$ and $BMDL_{1SD}$ values are essentially always lower than the NOGEL and are therefore more conservative.

9) Gollapudi et al (2013 – annex), a report from the HESI Quantitative analysis working group (QAW) workshop, describes a series of analyses using dose response data to derive POD's using the no observed genotoxicity level (NOGEL), the threshold effect level (Td) and the BMD approach. The focus of the study were EMS and MMS -both *in vitro* and *in vivo* data were analysed. *In vitro* NOGELs were of a similar order of magnitude for gene mutation and MN endpoints for both chemicals. The Td lower bound (TdL) values (lowest estimate of the inflection) were similar to the NOGEL's for all data sets with the exception of *in vivo* micronucleus for MMS which was 3-fold higher. The BMD and Td approaches are recommended, as the estimates have a quantified level of uncertainty. It was also concluded that these analyses provided support for the use of the lower confidence limit for use as a POD.

'the findings support the use of the lower confidence limit of 10% response is adequate POD and BMR for genotoxicity data , when the BMD approach is utilized'

9) Johnson et al (2014 -annex) provide a comprehensive table detailing the advantages, disadvantages and potential limitations of the POD metrics they examined, which includes NOGEL, BPD and also the BMD_{10} (PROAST) and BMD_{1SD} (BMDS-EPA) (but not BMD_{05}). Their conclusions include:

'the BMD approach yields the most conservative POD's (ie $BMDL_{10}$) '

'the BMD10 is comparable to, and recommended alongside, teh BMD1SD as the most suitable metrics for defining POD's for continuous genetic toxicology data'

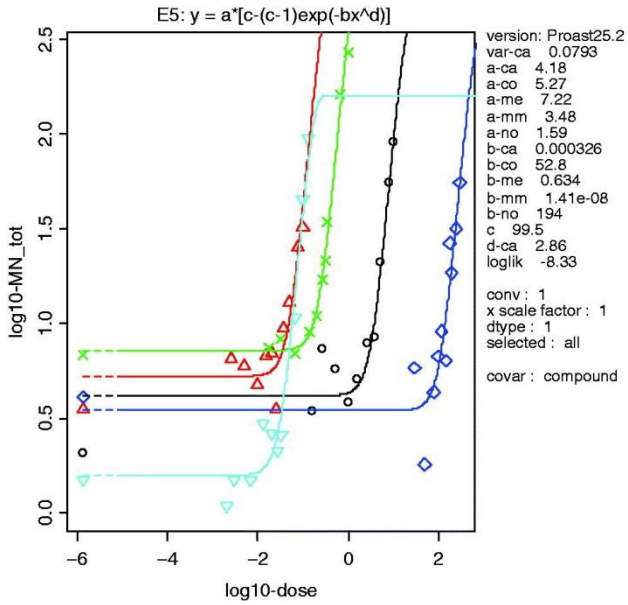
10) A comprehensive evaluation of dose responses for continuous endpoints has been undertaken by Slob and Setzer (2014 -annex) from RIVM and EPA respectively. Their intention was to examine steepness and shape of data from a variety of toxicological endpoints with view to addressing the questions :

" Do the data reveal any pattern or general behaviour in dose–response shapes that we can use to inform the selection of dose–response models in general? Or, do they rather tell us that the shape of a dose–response widely varies from chemical to chemical, so that the selection of a dose–response model indeed needs to be made on a case-by-case basis, as currently done."

11) Their analysis included 139 *in vivo* micronucleus studies of 51 different chemicals (with some variability of study duration, mouse train and sex). These studies were analysed in smaller clusters based on administration route (oral, inhalation or ip) and the oral cluster further subdivided by sex. Five *in vitro* micronucleus studies with 11 concentrations per study (including controls) were also analysed. All responses assumed to be log-normally distributed with homogeneous variance on a log scale. – the models fitted on log scale transformed > corrected for potency and for background response

12) The authors concluded that the shapes of toxicological dose responses relationships were homogeneous, and could be described by a four parameter exponential model and four parameter Hill model, both models giving equally as good fit. The consequences of this for BMD analysis are discussed.

Figure . Dose–responses of MNtot (micronucleus total count) for five genotoxic compounds (carbendazim, colchicine, mebendazole, Methyl methanesulfonate, nocodazole). Here, the exponential model was fitted to the combined dataset assuming equal parameter d. When scaling the curves with respect to background response and potency these dose– responses curves are identical.



13) An aspect which requires consideration is the approach by which parameters are constrained so parameters dose response model means the resulting curve cannot have an infinite slope at zero dose. There is also a commentary on threshold vs linear dose responses and it is argued that a threshold is a visual deception only and dependent on the scales and plots chosen to represent the data. However, from a biological point of view it may be argued that every continuous dose–response has a threshold-like behavior, in the sense that at low enough doses, no biologically significant effects occur, but at the same time dose–responses cannot be excluded to be ‘‘linear’’. The analysis reported takes a very mathematical view, that visual inspection cannot distinguish threshold from non-threshold. Furthermore it has not taken into account examples of a biological plausibility of a ‘threshold’ mechanism and there are many examples of where a biological threshold has dictated risk assessment decision, and indeed such distinction underpins risk analysis policy.

14) Wout Slob (2016) presents a general theory of effect size and its consequences for defining the benchmark response from continuous endpoints.

> *For continuous data, two distinct types of metrics are in use: those expressed relative to the variation (SD) in the controls (in this paper denoted as BMRSD), and those that ignore the within-group variation and focus on the change in mean response (usually a percent change; this BMD metric is also called critical effect size and will be denoted as CES in this paper). The argument behind the use of the BMRSD is that a given change in means does not seem to be comparable between endpoints that largely differ in within-group variation (“natural*

> *variability"). The BMRSD is a way to correct for that, where a larger change in means would be "acceptable" for endpoints with a relatively large "natural variability". The problem of this approach is that the within-group variation in the specific study depends on experimental error. First of all, it is subject to sampling error so that a given fixed value of the BMRSD in two perfectly replicated studies would relate to different changes in mean response. Further, studies examining the same dose– response may differ in within-group variation because the experimental conditions were not kept equally homogenous, or because differed analytical techniques resulted in different measurement errors for the same endpoint. Thus, "natural variability" is not entirely an appropriate term, as the observed within-group variation is study-specific, and includes sources of variation that are irrelevant for the dose-response, and hence for the BMD. As a final difficulty of the BMRSD, two identical dose–response relationships relating to populations that only differ in inter-individual variation will show different BMDs for the same value of the BMRSD. This makes the extrapolation of animal BMDs to human BMDs problematic.*

This is noted as a key differences between the modelling software from EPA (BMDS) and EFSA/RIVM (PROAST see section x)

15) Zeller et al (2016 – annex.) provide some discussion of dose response analysis and propose an alternative model whereby the critical effect size (CES) is tailored to the characteristics of the endpoint. Data on MMS is presented as a test case (described briefly in MUT/16/07). This ratio-based CES is based on criteria commonly used to define a positive response in regulatory tests and aiming to derive the CES value from 'insignificant effect' levels. The authors suggest that it's possible to take into account considerations such as amount of risk and what constitutes a safe dose.

BMD_{10} vs 1SD

16) EFSA 2009 concluded that a default BMR value of 10% be used for quantal data and 5% for continuous data from animal studies in the absence of specific information on what constitutes a biologically relevant change. As stated previously, the default BMR may be modified based on statistical or toxicological considerations.

MacGregor (2015a)

> *In the QWG analyses (Fig. 2), which involved continuous responses, a BMR of 10% was used, corresponding to an increase equal to 10% of the background (negative control) level. It should be noted that for quantal data, such as cancer incidence data, a BMR of 10% in absolute incidence rate is often selected to calculate a PoD. It should be emphasized that while both approaches generate a PoD that is referred to as a "BMD10", they are substantially different because the continuous response analysis, which is generally applied to genotoxicity data when the response of a subject is considered to be a continuous variable, is based on a BMR of a specified percentage increase (often 10%) of the spontaneous incidence, whereas the quantal analysis of cancer data is generally based on an absolute increase of 10% tumor incidence.*

Davis et al (2009)

The U.S. EPA discourages using a percentage change as the basis for a BMR for continuous endpoints without a biological basis to do so; the same percent change can represent very different degrees of response for different endpoints. U.S. EPA's guidance instructs that a BMR of 1 control standard deviation is a more appropriate BMR for continuous endpoints because it takes into consideration the distribution of the data and is more comparable to the 10% extra risk BMR suggested for dichotomous endpoints.

Questions:

- How should the best model be chosen? Hill or Exponential?
- How should model fit be assessed?
- Should models be constrained or not?
- Should transformation of the data be default?
- Which POD statistic is the most appropriate; Benchmark dose (BMD), NOGEL etc?
- What should the BMR/ CES be: SD or per cent change, scaled fractions of maximum effect?
- If the BMD is used is the 95% BMDL the most appropriate?
- Is the best percentage 5% and 10% or some other?
- Can models extrapolate from high dose to low dose?

Optimising study designs

17) At the October 2016 meeting Members commented that it was generally considered preferable from a statistical point of view to have a larger number of dose groups with fewer animals per dose group i.e. as opposed to a lower number of dose groups with more animals per dose. This is contrary to OECD guidelines for *in vivo* genotoxicity assays.

18) Slob (2014) provides some guidance on study design when using BMD compared to NOEL.

"The minimal number of animals needed in a study derives from the required study design performance. Study performance may be defined in two ways, depending on the aim of the study.

1. *In some studies the primary aim is to establish whether any effects (of some type) can at all be observed (hazard identification), in which case study performance is defined by the statistical power of detecting an overall effect.*

2. *In other studies, the aim is to measure the potency of the chemical, or, equivalently, to derive a point of departure (POD) for the purpose of hazard characterization, in which case study performance is defined by the precision of the estimated POD.*

Questions:

- What do Members think are the key elements of study design that need to be considered if study data is to be used in a quantitative manner?
- Can useful BMD analyses be conducted on studies conducted to the current OECD design guidelines?
- Any specific comments on [what constitutes] data quality?

Use of different genotoxicity endpoints and tissues

19) The standard battery of genotoxicity tests includes investigation of three types of genetic damage (gene mutation, clastogenicity) and aneuploidy. It is recommended that a positive response in an *in vitro* test is followed up by an *in vivo* assay examining the same endpoint (ie gene mutation or clastogenicity) thus providing an understanding of the mutagenic mode of action (MOA) of the chemical under test. Furthermore, the comet assay or transgenic mouse mutation assays can be used to examine target organ genotoxicity. The choice of organs for examination will depend on factors such as; known target organs for toxicity or carcinogenicity; organs known to be site of contact or site of metabolic activation.

20) For non-DNA reactive toxicities, a BMD is used to derive a POD from which permitted exposures, acceptable daily intakes or maximum tolerated doses can be established. Generally the endpoint with the lowest NOEL is selected which is considered to generate the most conservative risk assessment. How the most accurate and/or conservative risk estimations should be derived when using genotoxicity data has not yet been broadly addressed. The relationship of each genotoxicity endpoint to a human health effect is not well established. Furthermore, what background levels and induced increases of genotoxicity biomarker represent is not clearly defined. However it is clear that an understanding of the chemical mode of action will be critical in ensuring chemical risk assessments are biologically as well as mathematically reliable. The EMS/Viracept programme of work provides the best comparison of genotoxicity endpoints (despite the use of modelling based on a threshold response, and as outlined in the MUT/2016/07) and there are several studies examining other alkylating agents using different genotoxicity assays, which have been subjected to BMD analyses.

21) Data from the Viracept studies showed that EMS did not induce micronuclei in mouse bone marrow at doses ≤80 mg/kg/day and did not increase lacZ mutation frequencies in Muta™mouse at doses ≤25 mg/kg/day (bone marrow, small intestine) and ≤50 mg/kg/day (liver) (Gocke and Wall 2009). Gollapudi et al (2013 -attach) aimed to critically evaluate the different approaches to quantitative analysis (NOGEL, BMD, Td) and used this EMS data, and MMS data, including *in vitro* and *in vivo* micronucleus assays and *in vivo* gene mutation data from transgenic Muta^{TM}Mouse assays. The results generated from use of the three models from the different endpoints are shown. They did not compare BMDs derived from the two *in vivo* assays. The data from the three MutaMouse tissues following EMS exposure show how the study design and modelling highlights a different pattern of response.

Summary of results for EMS

Assay	Tissue	NOGEL	Tdl	BMDL10
		mg/kg/day		
Muta^{TM}Mouse	Bone marrow	50	21.46	9.29
	Liver	50	25.67	41.00
	GI Tract	25	12.97	12.23
BMMN	Bone marrow	80	56.66	58.68

It also states that '*the findings support the use of the lower confidence limit of 10% response is adequate POD and BMR for genotoxicity data , when the BMD approach is utilized*'

22) Cao et al (2014 –annex) examined the response of EMS in gpt-delta transgenic mice which have a lower spontaneous mutation frequency than MutaMouse. Mice were treated with oral doses of EMS for 28 days (0, 5, 13 ,20, 55 and 100 mg/kg/day) and mutation frequencies derived for lung, kidney, spleen, bone marrow, small intestine and liver. PigA mutation frequency in peripheral blood was also measured. The resulting data were modelled using threshold (Td/LCi) and BMDL (PROAST) analyses and presented as follows:

TABLE I. NOGEL, Td-LCI, and $BMDL_{10}$ Values Derived from *gpt* and *LacZ* Mutant Frequency (MF) and Mutation Frequency (*Mf*), *Pig-a* MF, and Micronucleus (MN) Frequency Data Sets

			Point of departure (PoD) in mg/kg/day			
Mouse	Tissue	Assay	NOGEL	Td-LCI	$BMDL_{10}$: PROAST[a]	$BMDL_{10}$: beta-regression[b]
gpt-delta	Lung	*gpt* MF	5[c,d]	0[c,e]	0.56	0.038[f]
	Kidney	*gpt* MF	5[g]	0[e]	0.59	0.313[h]
	Spleen	*gpt* MF	13[c]	23.24[c]	0.35	3.91[i]
		gpt Mf	13[c]	23.19[c]	3.80	4.04[j]
		gpt G→A *Mf*	20[c]	0[c]	0.52	0.022[k]
	Bone marrow	*gpt* MF	20	0.28[l]	0.37	4.39[m]
	Small intestine	*gpt* MF	20	13.75	0.55	5.08[n]
	Liver	*gpt* MF	55[o]	0	2.30	23.1[p]
	Peripheral blood	RET *Pig-a* MF	5[q]	0	1.18	NT
		RBC *Pig-a* MF	20[q]	0	4.67	
		Day 13 MN	20	23.51	6.79	
		Day 29 MN	55	29.06	8.26	
Muta™Mouse	Small intestine	*LacZ* MF	25	12.97	12.23[r]	NT
	Bone marrow	*LacZ* MF	50	21.46	9.29[r]	NT
	Liver	*LacZ* MF	50	25.67	41.00[r]	NT
CD1	Bone marrow	Day 7 MN	80	56.66	58.68[r]	NT

23) It was reported that the EMS genotoxicity dose responses in gpt-delta mice had lower PoDs than those calculated from the Muta™Mouse and CD1 mouse data. The authors suggested that the magnitude and the shape of mutagenicity dose responses differ between *in vivo* models, with lower PoDs generally detected by gene mutation assays with lower backgrounds mutation frequencies

24) Johnson et al (2014 –annex) provide a comprehensive evaluation of dose responses from MNU and ENU for a variety of endpoints using NOGEL, threshold dose (TdL) and various BMD models. It was demonstrated that BMDL values were generally lower and would provide a more conservative risk estimate. A useful comparison of bioassay results of 4 alkylating agents was provided .

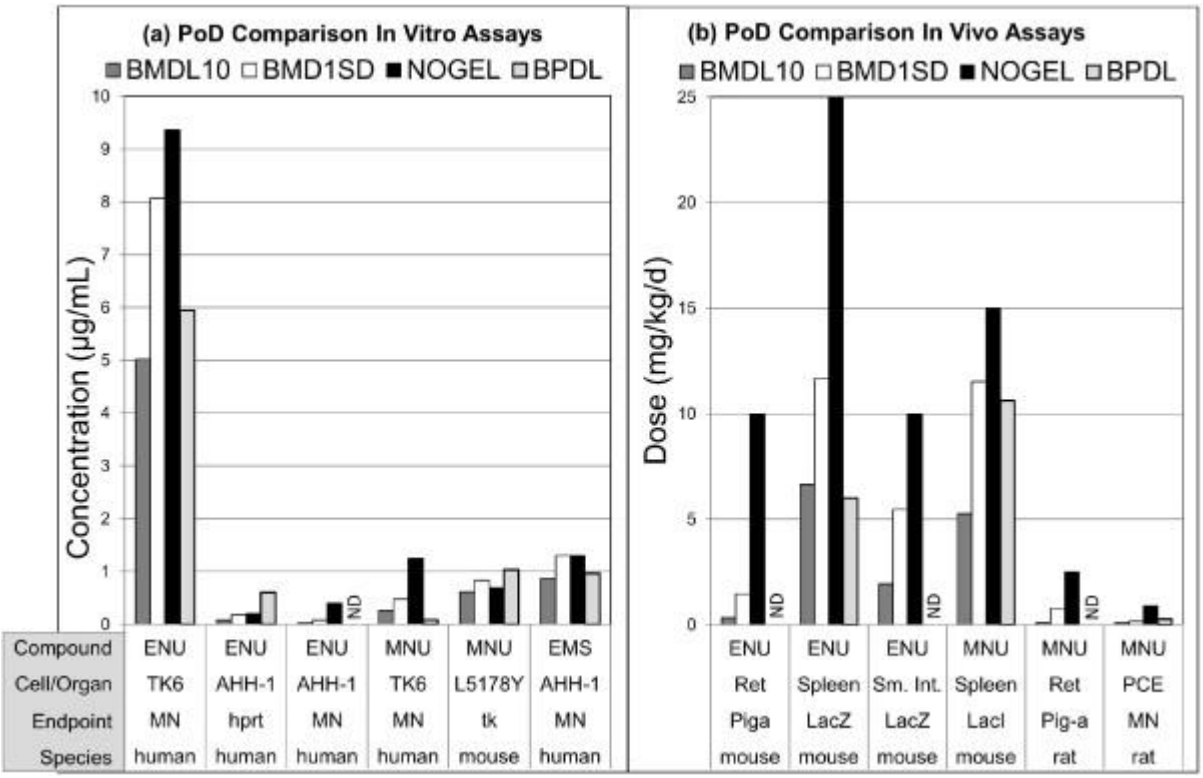
Summary of the lowest derived $BMDL_{10}$value and data from Gollapudi et al 2013: for EMS and MMS

Endpoint		MNU	ENU	MMS	EMS
Gene mutation	In vitro uM	0.006	0.68	4.72	8.70
	In vivo mg/kg	0.0007	0.09	1.34	9.29
Micronucleus	In vitro uM	0.03	0.17	1.00	4.35
	In vivo mg/kg	0.02	1.36	1.74	56.68
Cancer bioassay	mg/kg/day	0.093	0.95	31.8	N/A

See paper for details

25)_ MacGregor et al (2015a) represented these figures as follows:

Fig. 3.
Comparison of PoD metrics for ENU or MNU response in different assays and endpoints. (a) *In vitro* assays for gene mutation and micronucleus induction. (b) *In vivo* assays for gene mutation in different tissues and micronucleus induction in bone marrow. ND, breakpoint was indeterminate. The PoD metrics presented are the $BMDL_{10}$, the BMD_{1SD}, the NOGEL, and the BPDL.



26) Zeller et al 2016 (annex)– was described briefly in the original paper. Considering it more specifically the use of different endpoints, this study provides a comparison of pig-A, comet and BMMN for MMS.

27) Two papers describing empirical analyses of a) in vitro and b) in vivo potency comparisons (Wills et al 2016 a, b) also offer some data to inform the question of the impact of using different genotoxicity endpoints or tissues. In the *in vivo* study (Wills et al 2016b – append?) used a BMD covariate approach to analyse a variety of atasets. Firstly, data from previously published *in vivo* MN and Pig-A assays were used to generate potency rankings of 7 well established clastogens following administration of test compound for 3 or 28 days by oral gavage (Chemicals investigated were: 1,3 propane; azathioprene; chlorambucil; hydroxyurea; melphalan; MMS; thiotepa); 2 or 3 studies for each involving different sampling times. Combinations of datasets were used as covariate. Empirical comparisons were also made with pig-A mutation frequency data using sex as covariate. It is suggested that the combined covariate approach can be used to improve BMD analysis and thus the interpretation of *in vivo* studies.

28) A second set of analyses were derived from a mega Muta^{TM}Mouse study described by Long et al (2016). In this study daily oral doses of 9 different polycyclic aromatic hydrocarbons (PAH's) were administered for 28 days to 7 rats/group (14/controls). The aims were to investigate target tissue mutagenesis and whether some mutagenic carcinogens could be missed by a clastogenic assay (ie in vivo MN). Tissues (liver, lung, BM, stomach, small intestine) were harvested following a 3 day sampling period. In addition to lacZ mutation frequency, BMMN and PAH adduct levels were determined in the same animals. A range of doses for each PAH were based on preliminary range finding studies. All PAH's were positive for lac Z in at least one tissue, 3 were not positive in BMMN but 2/3 of these increased mutation frequency in lacZ in the bone marrow. For example benz(a)anthracene (BaA) was weakly positive in lacZ BM but negative in BMMN . Interestingly dibenz(a,h)anthracene (DBahA) induced a large increase in mutation frequency in the small intestine, wasn't positive in BM lacZ whilst increasing MN in the BMMN assay.

29) Wills et al (2016b) looked specifically at the data following administration of Benzo(a)pyrene BaP (range 0.1 – 50 mg/kg/day) to MutaMouse. BMD analyses were undertaken using PROAST version 50.9 using exponential or both exponential and Hill models (as recommended by EFSA) and the BMR was identified as 100% increase relative to control (i.e. doubling). Covariate analyses included sex, tissue and study as subgroups. Confidence interval data indicated tissue specific differences in BMD values spanned an order of magnitude which would be big enough to impact on human exposure limits if used in a risk assessment or MOE evaluation. However, it was concluded that combined covariate approach across a range of covariates gives a more preecise determination of BMD.

30) There are a number of papers exploring the use of *in vitro* genetic toxicology assays to study potency rankings with goals of reducing the number of animals to predict mutagenic or carcinogenic potency? Bemis et al (2016) aimed to compare clastogenic potency by analysing quantitatively MN data from *in vitro* and *in vivo* studies (from the same animals as Wills et al) for 1,3 propane; azathioprene; chlorambucil; hydroxyurea; melphalan; MMS; thiotepa; = chemicals considered to be model clastogens. For in vitro systems – substantially different shaped dose response curves were noted but largely potencies were similar and with exception of MMS and azthioprene which appear to be more potent in vitro, and potency ranking was comparable to in vivo systems. It is noted that the large datasets compared to a standard study means that the confidence intervals were smaller - The authors suggest that these data contribute to understanding the relationships between different genetic toxicology assays and hence their contribution to risk assessment

31) The importance of genotoxic MOA information in establishing which are the most relevant endpoints to use for POD determination is highlighted in many of the publications (Gollapudi et la 2013; MacGregor et al 2014b, Johnson et al 2014). Furthermore, it is suggested that the selection of choosing the appropriate tissues for a quantitative analysis should be based on the following : site-specific toxicity; mechanisms of toxicity; distribution and metabolism; any chemical accumulation; cell proliferation; the ability for DNA repair capacity to be induced by the chemical . Sensitivity and background mutation or micronucleus frequency of each endpoint will also affect the outcome. The relevance of DNA strand break assays such as the comet assay is not discussed.

Questions:

- What are Members opinions on the use of different genotoxicity endpoints for deriving BMD values?
- How do background levels of damage impact on the analyses – which assays are considered more robust?
- How much can be transposed from the analysis of alkylating agent data to chemicals in general?
- Is the inclusion of covariates in the modelling a suitable approach?
- Is meta-analysis suitable for combining factors in the study design (e.g. sexes) and comparing across different studies?
- What are Members opinions on the use of *in vitro* data for potency ranking and establishing BMD's for use in risk assessment?

Use in carcinogenicity risk assessment.

32) The use of POD's derived from genotoxicity data in place of carcinogenicity data, as part of regulatory strategies in human health risk assessment is the ultimate goal of many developing the quantitative analysis approaches considered in this paper. The QAW discusses the aims in regard to the identification of an exposure level associated with minimal risk of inducing genetic damage (Gollapudi et al 2013; Johnson et al 2014). The precedent for this is the extensive evaluation of genotoxicity data of EMS following its discovery as a impurity in Viracept as discussed previously. This gave rise to ICH Guidance (Human Pharmaceuticals M7 - Assessment and Control of DNA Reactive Mutagenic Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk) which provides a framework for classifying and controlling different mutagenic or carcinogenic substances in pharmaceuticals. However, whilst the evaluation utilised BMD software to derive a POD, it was based on the demonstration of hockey stick curve and a threshold for effect.

> *As a more broadly workable approach it is suggested that it is necessary to establish a level of chemical which does not increase the mutagenic or clastogenic event above an agreed minimal level compared to background levels* (Gollapudi et al 2013).

33) In Johnson et al (2014), use of the $BMDL_{10}$ to support regulatory evaluation is described, using ENU and MNU as the examples. To date it is unclear whether it will be possible to extrapolate assumptions from these small alkylating agents to genotoxic damage induced by other genotoxic chemicals. A more recent paper has utilised $BMDL_{05}$ (Soeteman-Hernandez et al 2016) in their comparison with BMD_{10} from carcinogenicity studies .

34) The use of the MOE approach, outlined in MUT/2016/07, is well established for use in risk communication scenarios where there are unavoidable exposures to carcinogens, or for risk mitigation prioritisation and to set risk management actions. It customarily uses BMD_{10} derived from an animal carcinogenicity study. The derivation of a pragmatic minimal risk level from this POD, an MOE of 10,000 was considered unlikely to be of concern. To date, it has principally been

impurities in drug substances and unavoidable contaminants in food (e.g. acrylamide) which are subject to regulatory assessments.

35) EFSA Scientific Committee (EFSA, 2005, 2009) concluded that, from the options considered, the MOE approach would be the most appropriate one in the risk assessment of substances that are both genotoxic and carcinogenic. They proposed to use the $BMDL_{10}$ from [carcinogenicity study] as the reference point. However, to date, no comments have been made on the use of a POD derived from genotoxicity data in place of a carcinogenicity value.

36) The quantitative use of dose response data in MOE approaches for genotoxic chemicals in food was considered by Benford (2016 –annex). Attention is drawn to the importance of considering factors such as study design and quality, strain and species and chemical MOA when using carcinogenicity data and that these factors would also be critical if genotoxicity data are used. It is noted that a comparison of potency in carcinogenicity and genotoxicity assays is necessary using a broad range of carcinogen classes and MOAs.

From Benford (2016)

> *So far, the MOE approach has been confined to substances for which carcinogenicity data are available. In the absence of carcinogenicity data, evidence of genotoxicity is used only in hazard identification. The challenge to the genetic toxicology community is to develop approaches for characterising risk to human health based on data from genotoxicity studies. In order to achieve wide acceptance, it would be important to further address the issues that have been discussed in the context of dose-response modelling of carcinogenicity data in order to assign levels of concern to particular MOE values, and also whether it is possible to make generic conclusions on how potency in genotoxicity assays relates to carcinogenic potency. So far, the MOE approach has been confined to substances for which carcinogenicity data are available. In the absence of carcinogenicity data, evidence of genotoxicity is used only in hazard identification. The challenge to the genetic toxicology community is to develop approaches for characterising risk to human health based on data from genotoxicity studies. In order to achieve wide acceptance, it would be important to further address the issues that have been discussed in the context of dose-response modelling of carcinogenicity data in order to assign levels of concern to particular MOE values, and also whether it is possible to make generic conclusions on how potency in genotoxicity assays relates to carcinogenic potency.*

37) Publications comparing carcinogenic and mutagenic potency were introduced in MUT/2016/07 and these papers are appended (Hernandez et al 2011, Soeteman Hernandez et al 2016). Some graphical representations are also given in George Johnsons presentation (annex) Consideration of the use of uncertainty factors is required.

> *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 mutagenic 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. Gollapudi et al 2013

Questions:

- What are Members opinions of the publications comparing genotoxic and carcinogenic potency?
- Is it possible to make generic conclusions on how potency in genotoxicity assays relates to carcinogenic potency?
- Can data from *in vivo* studies be used in risk assessment?
- What is the best POD metric from genotoxicity data to compare POD with carcinogenicity data? (or will this vary depending on the assay?)
- Can data from *in vitro* studies (in the absence of *in vivo* studies) be used in risk assessment?
- Can a POD derived from a genotoxicity study be used in a MOE approach to risk management?
    - If so, how should POD's from different genotoxicity endpoints be used
    - If so, how should uncertainly factors be used ?
    - Will this always be at least as protective as the use of carcinogenicity data? If so, is there scope for this approach as a pragmatic alternative when advice is required but there are no carcinogenicity data?

References:

Bemis, J.; Wills, J; Bryce, S.; Torous, D; Dertinger, S.; Slob, W (2016) Comparison of in vitro and in vivo clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data Mutagenesis 31(3) pp. 277– 285

Benford DJ (2016) The use of dose-response data in a margin of exposure approach to carcinogenic risk assessment for genotoxic chemicals in food Mutagenesis 31(3) pp. 329–331.

Cao X, Mittelstaedt RA, Pearce MG, Allen BC, Soeteman-Hernández LG, Johnson GE, Bigger CA, Heflich RH. (2014) Quantitative dose-response analysis of ethyl methanesulfonate genotoxicity in adult gpt-delta transgenic mice. Environ Mol Mutagen. 55(5):385-99.

Davis JA, Gift JS, Zhao QJ. (2011) Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1.Toxicol Appl Pharmacol. 254(2):181-91.

EFSA (European Food Safety Authority). 2005. Opinion of the Scientific Committee on a request from EFSA related to a harmonized approach for risk assessment of substances which are
both genotoxic and carcinogenic. The EFSA Journal 282: 1-31.

EFSA (2016) Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree EFSA Supporting Publication 13(3) EN-1006

EFSA 2009 EFSA (2009) Use of benchmark dose approach in risk assessment EFSA J., 1150 (2009), pp. 1–72

Gocke E, Wall M. (2009) In vivo genotoxicity of EMS: statistical assessment of the dose response curves. Toxicol Lett. 190(3):298-302.

Gollapudi BB, Johnson GE, Hernandez LG, Pottenger LH et al (2013) Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. Environ Mol Mutagen.54(1):8-18.

Hernández 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(7):518-28.

Johnson GE, Soeteman-Hernández LG, Gollapudi BB, et al 2014 Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen. ;55(8):609-23.

Long AS, Lemieux CL, Arlt VM, White PA (2016) Tissue-specific in vivo genetic toxicity of nine polycyclic aromatic hydrocarbons assessed using the Muta™Mouse transgenic rodent assay. Toxicol Appl Pharmacol. 2016 Jan 1;290:31-42.

MacGregor JT, Frötschl R, White PA, et al 2015a IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure-response relationships and points of departure (PoDs). Mutat Res Genet Toxicol Environ Mutagen. 2015 May 1;783:55-65.

MacGregor JT, Frötschl R, White PA, et al 2015b IWGT report on quantitative approaches to genotoxicity risk assessment II. Use of point-of-departure (PoD) metrics in defining acceptable exposure limits and assessing human risk. Mutat Res Genet Toxicol Environ Mutagen. 2015 May 1;783:66-78.

Müller L, Gocke E. (2009) Considerations regarding a permitted daily exposure calculation for ethyl methanesulfonate. Toxicol Lett. 190(3):330-3

Sanner T, Dybing E. (2005) Comparison of carcinogenic and in vivo genotoxic potency estimates. Basic Clin Pharmacol Toxicol. 96(2):131-9.

Soeteman-Hernández LG(1), Fellows MD(2), Johnson GE(3), Slob W(1). (2015) Correlation of In Vivo Versus In Vitro Benchmark Doses (BMDs) Derived From Micronucleus Test Data: A Proof of Concept Study. Toxicol Sci. 148(2):355-67.

Soeteman-Hernández LG, Johnson GE(1), Slob W(2). (2016) Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. Mutagenesis. 31(3):347-58.

Slob, W. (2014a) Benchmark dose and the three Rs. Part I. Getting more information from the same number of animals. *Critical Reviews in Toxicology*, **44**, 557-567.

Slob W (2014b) Benchmark dose and the three Rs. Part II. Consequences for study design and animal use. Critical Reviews in Toxicology. 44: 7. 568-580

Slob, W., and Setzer, R.W. (2014) Shape and steepness of toxicological dose-response relationships of Continuous endpoints. *Crit Rev Toxicol*, **44**, 270-297.

Slob, W. (2016) A general theory of effect size, and its consequences for defining the benchmark response (BMR) for Continuous endpoints. *Crit Rev Toxicol*, 1-10.

Wills JW(1), Long AS(2), Johnson GE(3),et al (2016) Empirical analysis of BMD metrics in genetic toxicology part II: in vivo potency comparisons to promote reductions in the use of experimental animals for genetic toxicity assessment. Mutagenesis. 31(3):265-75

White PA , Johnson GE (2016) Genetic toxicology at the crossroads—from qualitative hazard evaluation to quantitative risk assessment Mutagenesis 31 (3) pp. 233–237

Zeller, A; Tang, L; Dertinger, s Funk, J; Duran-Pacheco, G Guérard, M 2016
A proposal for a novel rationale for critical effect size in dose–response analysis based on a multi-endpoint in vivo study with methyl methanesulfonate Mutagenesis 31(3) pp. 239– 253

ANNEX

Papers to be appended:

Benford DJ (2016) The use of dose-response data in a margin of exposure approach to carcinogenic risk assessment for genotoxic chemicals in food Mutagenesis 31(3) pp. 329–331.

Gollapudi BB et al (2013) Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. Environ Mol Mutagen.54(1):8-18.

Hernández et al (2011) Can carcinogenic potency be predicted from in vivo genotoxicity data?: a meta-analysis of historical data. Environ Mol Mutagen. 52(7):518-28.

Johnson GE, et la 2014 Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen. ;55(8):609-23.

MacGregor et al 2015 IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure-response relationships and points of departure (PoDs). Mutat Res Genet Toxicol Environ Mutagen. 2015 May 1;783:55-65.

MacGregor JT, et al 2015 IWGT report on quantitative approaches to genotoxicity risk assessment II. Use of point-of-departure (PoD) metrics in defining acceptable exposure limits and assessing human risk. Mutat Res Genet Toxicol Environ Mutagen. 2015 May 1;783:66-78.

Soeteman-Hernández et al. (2016) Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. Mutagenesis. 31(3):347-58.

Slob, W., and Setzer, R.W. (2014) Shape and steepness of toxicological dose-response relationships of Continuous endpoints. *Crit Rev Toxicol*, **44**, 270-297.

Slob, W. (2016) A general theory of effect size, and its consequences for defining the benchmark response (BMR) for Continuous endpoints. *Crit Rev Toxicol*, 1-10.

Wills et al (2016) Empirical analysis of BMD metrics in genetic toxicology part II: in vivo potency comparisons to promote reductions in the use of experimental animals for genetic toxicity assessment. Mutagenesis. 31(3):265-75