COM Statement:

Quantitative approaches to the assessment of genotoxicity data

Introduction and background to the current review

1) Genetic toxicology has traditionally been based on the development and implementation of *in vitro* and *in vivo* assays designed to identify substances which cause damage to DNA and/or other cellular components which regulate the fidelity of the genome. The information derived from these testing strategies is used in a qualitative manner, to establish whether or not the chemical is a genotoxic or mutagenic hazard. Accordingly, risk management approaches are based on this dichotomous (yes/no) decision, which helps protect against public exposure to potentially genotoxic [and therefore potentially carcinogenic] agents (COM 2011; EFSA 2011). These assays are also useful during product development to "design out" genotoxic liability. However, this is a conservative approach and in the process, with potentially valuable chemicals being screened out and discarded unnecessarily, or strategies to remove agents from the environment or food have to be undertaken, despite that fact that exposure, and risk, may be very low (Kirkland et al 2007; Pottenger and Gollapudi 2010).

2) Conventional approaches to assessing the risk of chemicals which are toxic /non-genotoxic are [generally] based on establishing a non-toxic level in *in vivo* studies (the reference dose; RfD, or the point of departure; POD) and applying uncertainly factors to estimate an exposure which represents a Health-based Guidance Value (HBGV) such as a maximum acceptable daily intake (ADI) (WHO 2009). In general, for genotoxic carcinogens, the view is that there is no threshold. A margin of exposure (MOE) approach based on a POD derived from a carcinogenicity study, can be utilised for carcinogens that are genotoxic and for which there is unavoidable exposure (EFSA 2005; Barlow et al 2006; Benford et al 2010). Currently, there is considerable interest in the development and evaluation of methodologies which would enable the analysis of genotoxicity dose response data to be carried out in a quantitative manner.

3) Modification of the traditional [yes/no] approach to genotoxicity is a substantial departure from current practises. Development of a strategy based more on quantitative analyses would require extensive evaluations of the dose response methodologies employed and a more detailed understanding of the relationship of the genotoxicity endpoint to a human health effect before it would be possible to establish the appropriateness and/or usefulness of quantitative assessments of genotoxicity data. Reports from the International Workshops on Genotoxicity Testing (IWGT) working group in quantitative approaches to genetic toxicology risk assessment (the QWG) (MacGregor et al 2015 a,b) and publications arising from a workshop organised by the Health and Environmental Sciences Institute (HESI) Genetic Toxicology Technical Committee (GGTC)

(summarised in White and Johnson 2016) provide insight into how international groups are addressing this changing risk assessment environment.

4) It is suggested that refining approaches to assessment of genotoxicity data could contribute to reductions and improvements in the use of animals in toxicity testing (reduction, refinement, replacement; 3R's) (Johnson et al 2014; Soeterman-Hernandez et al 2016).

5) The COM first considered quantitative approaches to assessing genotoxicity data, and how they may be used in chemical risk assessment, at its Horizon Scanning exercise in June 2013. Members were aware of work being conducted by IWGT and HESI on quantifying genotoxic responses and assessing non-linear dose-response relationships, and agreed that the implications of this work should be considered. As many chemicals present in the environment have not been tested for carcinogenicity, the possibility of developing quantitative (or semi-quantitative) methods for the analysis of dose-response data from *in vivo* genotoxicity studies for use in a MOE approach, similar to that utilised in the interpretation of carcinogenicity data, was raised. *These approaches may also facilitate ranking of genotoxins for potency.*

6) The COM were given a presentation by Dr George Johnson (University of Swansea), a member of key working groups, and considered papers summarizing the key research in the field (MUT/2016/07; MUT/2017/02; MUT/2017/03). A number of key themes and questions were addressed including:

- *What dose response modelling methods are available, and which are most appropriate for evaluating genotoxicity data?*
- *Which point of departure (POD) metric is best for assessing genotoxicity data?*
- *How do factors such as endpoint, tissue and study design impact on assessing data quantitatively?*
- *Can quantitative information from genotoxicity data be used in risk assessment, and if so, how?*
- *Is it possible to characterise carcinogenic risk from genotoxicity data alone?*

This statement is a summary of the information considered by the COM and the resultant discussions and opinions. [or what do Members want this statement to be?]

Current hazard and risk assessment approaches

7) The genotoxicity testing strategy currently recommended by COM (*Guidance link*) for the detection of mutagenic hazard is based upon a core set of *in vitro* tests, chosen to

provide information on three types of genomic damage; gene mutation, clastogenicity and aneuploidy. These are followed, if necessary, by appropriate *in vivo* tests designed to investigate whether *in vitro* genotoxic activity, including the specific mutagenic effect identified, also occurs *in vivo* (i.e in the whole animal). The testing strategy may also include assays for specific target organs (e.g. site of contact tissues or site of rodent tumours detected in carcinogenicity bioassays) or germ cells. *If a chemical is considered to be genotoxic it is generally assumed that, other than for a few exceptions (COM guidance, statement on thresholds?), there is no exposure level below which there is no effect. For chemicals, for which potential exposure cannot be eliminated, the ALARA (as-low-as-reasonably-achievable) or ALARP (as-low-as-reasonably- practicable) approach is advised. This suggests that levels of the chemical must be controlled to ensure that intake is minimised to be as low as reasonably, or technically, possible (Barlow et al 2006) and is a widely adopted principle used by regulatory authorities in Europe and many other regions. It is a purely qualitative (hazard-based) risk management approach, and there is no consideration of the genotoxicity or carcinogenicity data in a quantitative manner.*

8) Exceptions to the 'no safe level' assumption have previously been established, based on the demonstration of a non-linear dose response and a mode of action which exhibits a threshold. COM generated a Guidance Statement on thresholds for *in vivo* mutagens in April 2010 (COM 2010). A number of different threshold terms were defined in this document (i.e. true threshold, threshold dose, practical threshold, biologically meaningful threshold, threshold mode of action).

9) An example of a threshold in a mutagenic response that has been extensively investigated is that demonstrated by some low molecular weight alkylating agents, a consequence of the repair of DNA adducts. A detailed evaluation and human risk assessment was undertaken following the discovery that ethyl methanesulfonate (EMS), a known genotoxic carcinogen, was found as an impurity in some tablets of Viracept (nelfinavir mesilate), an HIV protease inhibitor (Walker et al 2009; Muller et al 2009). It was estimated that consumption of contaminated drug batches at the maximal daily dose resulted in patients ingesting EMS at up to 0.045 mg/kg/day (daily Viracept dosage of 2.92 g/day). The responsible pharmaceutical company (Roche) went on to perform a comprehensive quantitative risk assessment of EMS agreed with European regulatory agencies (Muller and Singer 2007) and determined a 'safe level'. *The disparity between the frequency of DNA adducts and of mutations suggested that a DNA repair factor was involved in the conversion of adduct to mutation and that this mechanism exhibits a threshold (Jenkins et al 2005; Doak et al 2007). Therefore it is possible that an organism could be subjected to a low level of DNA damage without deleterious effects because the damage is effectively and efficiently repaired, and it is only when repair mechanisms are exhausted or overwhelmed that a mutation occurs.* The risk assessment was based entirely on

establishing a mode of genotoxic action which had a clear threshold from which a POD was established.

10) The COC have defined approaches to risk characterisation of carcinogens and these are described in a Guidance Statement COC/G-06 (2012 – https://www.gov.uk/government/publications/carcinogenic-dose-response-defining-a-point-of-departure-and-potency-estimates). These are broadly in accordance with those proposed by EFSA (2005). These include the margin of exposure (MOE) approach and the threshold of toxicological concern (TTC). The threshold of toxicological concern (TTC) is a de minimis approach developed to facilitate the risk management of substances, primarily contaminants in food, for which good (or at least conservative) exposure estimates are possible but when chemical-specific toxicity data, including genotoxicity data, are insufficient for normal risk characterisation (Kroes et al 2004; Dewhurst and Renwick 2013). Exposure levels below which safety concerns are not anticipated are given for different classes of chemicals including genotoxic carcinogens.

11) The MOE approach can be applied to chemicals that have been shown to be genotoxic and carcinogenic, and takes into account carcinogenic potency and estimated exposure (EFSA 2005; Barlow et al 2006). The MOE is calculated using a POD derived from suitable rodent bioassay data or human epidemiology information which is divided by the measured or estimated exposure. The resulting value, which is unitless because it is a ratio, has been classified by the COC (based on MOEs calculated using animal carcinogenicity data) as follows:

- may be a concern (MOE<10,000);
- unlikely to be a concern (MOE 10,000-1,000,000) or
- highly unlikely to be a concern (MOE >1,000,000)

12) The recommended POD is based on the benchmark dose (BMD). The approach uses mathematical modelling to calculate the lower 95% confidence interval of a dose (BMDL) causing a defined response, typically a 10% increase in tumours in a cancer bioassay, i.e. the $BMDL_{10}$ (EFSA 2009; 2016). This is replacing the 'traditional' no-observed-adverse-effect level (NOAEL) approach for non-cancer endpoints; and, because the models use all the dose–response data, it provides a quantitative estimate of the uncertainties. This method has gained acceptance by some regulatory bodies (including European Food Safety Authority; EFSA, European Medicine Agency; EMA and World Health Organisation; WHO) for managing genotoxic carcinogens that cannot be avoided (e.g. contaminants), but to date is only useful when good quality carcinogenicity studies are available.

General publications on quantitative risk assessment of genotoxicity data

13) The COM considered a number of publications which examined the application of a range of dose response modelling methods using data from genotoxicity studies and the quantitative analyses resulting from them (Gollapudi et al 2013; Johnson et al 2014; MacGregor et al 2015a,b). Three principle POD metrics were compared; these were:

i) The no-observable genotoxic effect level (NOGEL). This is the highest experimental dose level where there is no statistically significant increase in the genotoxic effect measured in the study.

ii) The threshold effect, lower confidence limit (TdL). This was used in the EMS/Viracept analysis and is based on the assumption of a 'hockey stick' dose-response (Lutz and Lutz 2009; Gocke and Wall 2009). It involves fitting a mathematical model which assumes that the dose response is bi-linear with a region where there is no effect (it is similar to breakpoint dose models). It has been argued that the assumptions made with the use of this model need to be supported by mechanistic data.

iii) The BMD approach. This is determined by mathematical modelling of the dose – response curve and has been widely used in other branches of toxicology. The approach involves, firstly, fitting a mathematical model to experimental dose response data and, secondly, determining the Benchmark Dose (BD) which is estimated to produce a defined increase in the response over the control/background level (termed the benchmark response (BMR) or the critical effect size (CES). For example, using data from a carcinogenicity study a 10% increase in tumours over the control incidence is considered the BMR and the estimated dose is termed the BMD_{10}. The lower 90% confidence limit on the dose, termed the BMDL is then used as the POD in further considerations related to risk assessment such as the derivation of a MOE.

14) Recent interest in the development of quantitative analysis of genotoxicity data has focused on developing a similar BMD approach to that used elsewhere in toxicology. The QWG and HESI groups agreed that BMD modelling is the preferred approach for deriving a POD for genotoxicity data (Gollapudi et al 2013; Johnson et al 2014; MacGregor et al 2015a). It was also noted that the BMDL usually produces a lower and, hence, more conservative value for the POD than the other metrics considered. The BMDL takes account of the amount of variability in the data by considering the width of the confidence interval of the BMD; i.e the ratio of the BMDU (the upper 95% confidence interval of the BMD) to the BMDL has been proposed as a useful metric for the assessment of the uncertainty in the BMD estimate (EFSA 2017).

15) The COM acknowledges that developments in dose response modelling have been made which make it possible for genotoxicity data to be analysed quantitatively rather than only qualitatively. The COM broadly agreed with the conclusion that the BMD approach provides the best representation of the dose response. However, the lack of consensus amongst users of the approach was highlighted. [preliminary comments here?]

Benchmark dose approach

16) A number of areas were identified which were considered important for the COM to address in more detail when evaluating the potential of using genotoxicity data in a quantitative manner. In particular, there appear to be substantial differences in the use of the dose response modelling and in the derivation of BMD metrics. These differences include; choice of software package, the dose response models, the statistical evaluation of model fit, the use of constraints/options, the choice of BMR and methods for selecting or combining multiple BMDs. These areas are highly technical but it is important that the rationale for the choices made can be understood by the toxicologists and risk assessors who will be working with the results or the modelling processes.

Software, dose response modelling and BMD metrics

17) There are two principle software packages for the derivation of BMDs (Davis et al 2011; EFSA 2016). The BenchMarkDose Software (BMDS) package was developed by the US Environmental Protection Agency (EPA) in order to standardize approaches to evaluating dose response assessments. The software has over 30 different mathematical models or model variants which can be used for the analysis of quantal data, continuous data, nested developmental toxicology data, multiple tumour analysis, and concentration-time data. The software is freely available on the EPA website https://www.epa.gov/bmds. There are also extensive documentation, guides and training webinars on its use. New releases of the software are released from time to time.

18) The PROAST software package has been developed by the Dutch National Institute for Public Health and Environment (RIVM), and is freely available from their website http://www.rivm.nl/en/Documents_and_publications/Scientific/Models/PROAST. A comprehensive discussion of the software is available in the EFSA scientific opinion (2009 – appendix p47-72) *Various guides to its use are also provided with the instructions for its installation. The software requires the R computing language to be installed. The current version available at the RIVM website is version 38.9. There are other versions available through the Swansea web site and more advanced versions are used to investigate more advance modelling approaches. (Note that the example in EFSA (2016) uses version 61.6)*

19) Both packages provide methods for fitting similar mathematical models to dose-response data. There are some differences in the methodologies used. Two major differences in the default approaches have been described :

i. In the transformation of response data. PROAST (RIVM) uses the default assumption of a log-normal distribution and transforms the data using logs whereas. BMDS (EPA) recommends choosing the most appropriate transformation of the response data for the analyses (which may or may not be a log transformation) based on an assessment of how well the models describe the data, with the default being no transformation.

ii. Choice of BMR or CES: BMDS uses 1 standard deviation (1SD) above the background as the benchmark response (BMR) for continuous data, as the default whereas PROAST uses a percentage increase e.g. 5%, 10% above the background for the Critical Effect Size (CES). However, recent versions of BMDS can also be used in this way.

20) The COM also discussed the various dose response modelling methods used in BMD analysis. The IWGT consider, for risk assessment, that it should be possible to relate the POD to an acceptable exposure level by extrapolating from data which includes mode of action (MOA) and mechanistic information if available (i.e. so that 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 continuous genotoxicity data this represents a percent increase above a spontaneous incidence as opposed to an absolute increase of a quantal parameter; i.e a 10% increase in tumour incidence relative to the unaffected control population from, say 5% to 15% in the number of animals with tumours[1]in a carcinogenicity study (MacGregor et al 2015a).

21) A direct comparison between PROAST and BMDS, based on $BMDL_{10}$ and BMD_{1SD} values (respectively), from different *in vivo* and *in vitro* genotoxicity studies on methylnitrosourea (MNU) was undertaken by Johnson et al (2014). From this limited analysis, it was concluded that the two approaches produce comparable results and that both can be recommended for defining POD's for continuous data. In a study examining the correlation between genotoxicity and carcinogenicity, the BMD_{05}, calculated from bone marrow micronucleus (BMMN) data, was selected as the POD for comparison with the $BMDL_{10}$ derived from carcinogenicity studies (Soeteman-Hernandez et al 2016). No rationale was given for selecting a 5% increase for the BMMN POD but the authors stated that the choice of BMR was not crucial for their analyses. [*COM thought that the choice of BMD_{05} as a benchmark response in this study may be (arbitrary?) based on the outcome of the comparison with a carcinogenicity POD rather than because the authors believed it to be the most suitable value for use in risk assessment].* EFSA (2009) concluded that a default BMR value of 10% be used for quantal and 5% for continuous toxicological data from animal studies in the absence of specific information on what constitutes a biologically relevant

change. Both EFSA and EPA noted that, where specific information is available, the BMR should be based on statistical or toxicological considerations. However, no specific considerations of genetic toxicity data are given.

22) COM noted that comparison of the models and an understanding of the appropriateness of the different BMR/CES for the various genotoxicity was complex and difficult to conclude on given the limited datasets available. It was considered that more examples and further explanations of the basic assumptions used and the uncertainties that are applied to each model were required before COM would be able to come to any conclusions or make any recommendations on which model or dose response metric was the most suitable to analyse genotoxicity data. Furthermore, COM agreed that it was not obvious at present that the modelling could be transposed directly from its use with other toxicological endpoints to use in genetic toxicology. ***Members comments?***

Endpoints and tissues

23) Members considered how results from the different types of genotoxicity studies (and therefore endpoints) or different tissues will impact on the derivation of POD values for use in potency estimations or risk assessment scenarios. How the most accurate and/or conservative risk estimations should be derived when using genotoxicity data has not yet been broadly addressed. For example, the relative increase in DNA damage measured by the comet assay is likely to differ appreciably from the relative increase in BMMN induced by the same chemical in the same animals, since each response will be influenced by the chemical's MOA. Furthermore, it is not certain what the background levels of damage or the induced increases of each genotoxicity biomarker represent in a risk assessment scenario. A $BMDL_{10}$ (a 10% increase) in bone marrow micronuclei (BMMN) data will be substantially lower than the two fold increase in BMMN currently considered to represent a biologically meaningful positive response. The use of a two fold increase as a CES/ BMR would therefore be equivalent to BMD_{100}. It is clear that an understanding of the chemical's carcinogenic and/or mutagenic mode of action, and the biological relevance of the size of the increase, will be critical in ensuring chemical risk assessments are biologically relevant together with the results of mathematical modelling.

24) The COM examined a number of publications with the aim of addressing the importance of differences in endpoint or tissue when deriving BMDs from genotoxicity data. Many of the published studies investigating the differences in genotoxicity endpoints have focused on the alkylating agents ethyl methanesulphonate (EMS); methyl methanesulphonate (MMS); 1-methyl-1-nitrosourea (MNU); and 1-ethyl-1-nitrosourea (ENU), although some publications also examined polycyclic aromatic hydrocarbons as model genotoxicants.

25) A comprehensive evaluation of the dose-responses generated i*n vivo* following MNU and ENU exposure for a variety of endpoints including BMMN, gene mutations in lac Z transgenic mice, or in pig-A (in mice), was undertaken as part of a programme of work developing POD-based evaluations of genotoxicity data (Gollapudi et al 2013; Johnson et al 2014). The lowest BMDL value for each chemical was derived from the *in vivo* gene mutation studies. This is as expected given the direct acting MOA of the alkylating agents which induce primarily point mutations. These values were also conservative when compared with the values derived from the cancer bioassay. *Is this evidence that a value derived from genotoxicity data will be conservative compared to cancer bioassay for this class of chemicals only***? Members comments...?**

26) Zeller et al (2016) used MMS to examine the relationship of an endpoint with the chemical mode of action and to address the effect this has on the choice of CES/BMR. The results showed that MMS acts primarily as a clastogen and, therefore, its potency as a gene mutagen is lower. Therefore, it would not be appropriate to apply the same CES to both chromosomal damage and gene mutation endpoints for this chemical. Zeller et al concluded that a 'one-size-fits-all' CES for genotoxicity data may be sub-optimal because of the variability in baseline values, scoring systems and the inherent differences in the characteristics of each end-point. However, more data are needed before realistic recommendations for CES with different endpoints can be made. **Members comments... ?**

27) Detailed comparisons of endpoints and dose responses following administration of a number of polycyclic aromatic hydrocarbons (PAH's) including benzo(a)pyrene (BaP) and dibenz(a,h)anthracene (DBahA) to MutaMouse were undertaken with a view to improving the interpretation of genotoxicity dose response data (Wills et al 2016a). BMMN, pig A and lacZ gene mutations were examined from a variety of tissues. Covariate analyses (e.g. pooling data from sexes or different tissues) were used and the BMR was chosen as a 100% increase relative to control (i.e. doubling). *The authors concluded that statistical analysis of the BMD estimates of covariates provided robust potency rankings. Confidence interval data indicated that tissue specific differences in BMD values spanned an order of magnitude. Such large increases could have a significant impact if they were used in a risk assessment or MOE evaluation to establish acceptable human exposure limits***. *Members comments?***

28) An examination of the response of gpt-delta transgenic mice, which have a lower spontaneous mutation frequency than in the earlier Muta^{TM}Mouse studies, to EMS indicated substantially lower POD's in the gpt-delta mice (Cao et al 2014). Members commented that it was not clear whether the results were a consequence of the lower baseline, a different strain of mouse or a particularly sensitive endpoint. Accordingly, it demonstrated the importance of understanding a chemical mode of action, the appropriateness of the endpoint and the sensitivity of the genetic target in interpreting data quantitatively.

29) Establishing genotoxic MOA information was highlighted as vital in deciding on the most relevant endpoints to use for POD determination (Gollapudi et la 2013; MacGregor et al 2015b; Johnson et al 2014). Furthermore, it was suggested that the selection of 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 of each endpoint and background mutation or micronucleus frequency, will also affect the outcome of the analysis.

30) Members broadly agreed with the use of covariate analyses for combining data from different tissues where this was appropriate (a context?] The preliminary data available to them highlighted the importance of the selection of relevant endpoints and tissues if quantitative data were going to be used effectively. ***What do Members want to say about covariates and combining data?***

31) COM agreed that, whilst these types of studies contribute useful information to this area of research and the development of the quantitative analysis approaches, it was not possible to extrapolate findings from specific chemicals or chemical classes (e.g. alkylating agents) to generate broad assumptions. They considered that not enough is known about the quantitative relationships of different genotoxic or mutagenic effects, pre-neoplastic lesions and tumours to be able to interpret dose-response data accurately from a particular endpoint/tissue for each chemical. They suggested that more robust analyses of a larger number of more varied chemicals were required before any conclusions could be reached. They suggested [recommended] that a database which enabled the comparison of BMDs across chemicals, endpoints and tissues would provide useful starting material for a more comprehensive evaluation of the utility of quantitative assessment of genotoxicity data. Members also pointed out that the applicability of the quantitative approaches to germ cell mutagenesis had not been addressed.

Use in carcinogenicity risk assessment

32) *Proponents of these developments have argued that the using quantitative methods for the analysis of genotoxicity data will provide the potential to move away from a 'hazard-only' approach in the chemicals for genotoxicity towards a risk-based approach (Johnson et al 2013; MacGregor et al 2015a,b).* COM examined some publications from groups exploring the possibility of using POD's derived from *in vivo* genotoxicity studies in place of those generated from long term carcinogenicity studies, for example in MOE assessments (Sanner and Dybing 2005; Hernandez et al 2015; Soeterman-Hernandeez et al 2016). It is understood that, with regard to potential exposure to chemicals that are (or could be) genotoxic carcinogens, there are a number of risk management needs. These range from

determination of the potential level of concern for exposure to unavoidable contaminants or constituents of the diet, to market authorisation of new products (such as pesticides and human medicines). Hence, it is unlikely that a single approach would be suitable for all risk management situations. Indeed, as discussed above, the approach currently utilised by the COC varies depending on the risk management context.

33) The quantitative use of dose response data in MOE approaches for genotoxic chemicals in food was considered by Benford (2016). Attention was 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. EFSA (2009) recommend the MOE approach for substances that are both genotoxic and carcinogenic, when risk assessment is necessary. They proposed the use of the $BMDL_{10}$ based upon tumour data from carcinogenicity studies as the POD. To date, no view has been expressed (by anyone?) on the use of a POD derived from genotoxicity data in place of a carcinogenicity value.

34) The COM were provided with a number of publications detail comparisons of mutagenic and carcinogenic potency using BMD dose response modeling. A preliminary evaluation was undertaken by Sanner and Dybing (2005) who concluded that there was a correlation between carcinogenic and mutagenic potencies. A framework, using the lowest effect dose (equivalent to the LOGEL) in a micronucleus study, was proposed as having the potential to be used in regulatory settings when a chemical was considered to be mutagenic but for which carcinogenicity studies are either not available or of poor quality.

35) A comprehensive evaluation of potency estimates was undertaken by Hernandez et al (2011) using 18 chemicals listed as either IARC class 1 or 2A carcinogens. BMD_{10} values for carcinogenicity and genotoxicity were derived using the PROAST dose-response modelling current at the time. Those from genotoxicity data were based on a range of endpoints (BMMN, comet, transgenic (TG) mutations in mice) from various tissues from multiple studies. Some of the carcinogenicity studies, however, used only two treatment dose levels and different exposure routes were used in some cases. Hernandez et al concluded that there was some degree of association and a correlation between the BMDs for mutagenicity and carcinogenicity despite the differences in study designs and routes of exposure.

36) An extension of this study, using similar methodologies, evaluated 48 chemicals, (Soeteman-Hernandez et al 2016) and calculated BMD_{05} from MN studies. The log10 of these BMD_{05} values were plotted against the log10 of the cancer BMD_{10} values. The BMD_{05} were calculated for individual datasets (i.e. if there were several studies available, the data were not pooled) and a version of PROAST (*current at that time?*) was used for dose-

response modelling of both sets of data. The plot of the data showed a wide scatter but the authors concluded that there was a positive correlation between the BMMN and carcinogenic potencies (as measured by the BMDs), although prediction of carcinogenic potency from the genotoxicity data had an uncertainty of two orders of magnitude (i.e. factor of 100.

37) *COM members considered that the causal relation implied by the association of the mutagenicity and carcinogenicity potency was problematic and were unconvinced by the 1:1 associations that seemed to be inferred in the publications. They felt that the complex relationships between adducts, mutations, pre-neoplastic lesions and tumours make it unlikely that evidence for a simple ratio /association is robust. It was considered possible that the dose inducing a biologically relevant genotoxic effect in an appropriate tissue would be lower and therefore be a more conservative RP/POD for protecting health than a $BMDL_{10}$ for cancer. Members, however, felt that analysis of more datasets using a much broader range of chemicals and chemical classes is essential before any assumptions could be made. In particular, they pointed out that much less is known about the pattern of responses for weak genotoxins: for example, styrene, which causes tumours in nasal turbinates, induces relatively weak responses in genotoxicity assays.*

Study designs, data quality and use of uncertainty factors

38) Members considered it important to evaluate the impact of study design, and to consider the quality of the available data before conducting or interpreting quantitative analysis of genotoxicity data in order to generate PODs. It was noted that for an optimal statistical design for BMD modelling it is preferable to distribute a fixed number of animals in a study into more dose groups with fewer animals per group. However, this may not accord with current OECD guidelines for *in vivo* genotoxicity tests. Nevertheless, it was agreed that there is some flexibility within these study designs and that the two designs (i.e. for OECD and for BMD estimation) were not necessarily incompatible and that the current OECD guideline designs of genotoxicity studies were suitable for quantitative analysis

39) Data quality is partly reflected in the width of confidence intervals, which is also dependent upon the number of dose groups and animals per group. Members commented that guidance should be provided on what level of uncertainty in the data and what ratio of BMDU : BMDL would be considered unacceptable.

40) *One important aspect of fitting mathematical models to dose-response data is testing whether the model is a good fit to the data. Models which are not a good fit should not be used. Members noted that the choice of model based solely upon the results of 'goodness of fit' tests is a contentious area when a number of models provide a satisfactory fit. EFSA (2016) recommend that a model averaging approach is used, rather than a single*

*default model. However, it is not clear if there is currently suitable software, readily available to use the method with quantitative data**. Members comments?***

Use of in vitro genotoxicity data for deriving POD's

41) There are a number of recent publications which have examined the use of BMD assessments of *in vitro* genotoxicity studies for potency comparisons, or for comparisons of *in vitro* with *in vivo* BMD's (Soeterman-Hernadez et al 2015; Bemis et al 2016; Wills et al 2016b). *COM commented, that whilst an interesting innovation, these approaches are at an early stage of development and currently cannot be considered for risk assessment scenarios. Useful in potency rankings? COM decided not to consider this use of quantitative models further at this time.*

Overall discussion and conclusions

42) COM considered the current literature on quantitative analyses of dose-response data from genotoxicity studies, including the reports from IWGT and ILSI/HESI, and discussed the recent developments of the approaches in this area.

43) It is noted that a move towards quantitative assessment of data is a significant departure from the current practises which are based principally on establishing only whether a chemical represents a mutagenic hazard. Discussions such as this should enable exploration of concepts which underpin the use of genotoxicity data in risk assessment; for example, would such an approach imply that all genotoxic chemicals have an exposure level below which the risk is considered tolerable and how that might this level be identified? It is evident that such approaches have the potential to improve the interpretation of genotoxicity data, reduce the need for long-term carcinogenicity studies and, hence, reduce the number of animals used in chemical risk assessment (with 3R's benefit). As these approaches are developed and their utility demonstrated, there will be scope for them to be incorporated into regulatory frameworks.

44) COM recognised the importance of the developments in the software and use of BMD methodologies to evaluate genotoxicity quantitatively and were broadly in support of the approaches and their potential utility. However, it was noted that, to date, much of the analyses have been performed by a small number of specialists [without widespread scrutiny?]. Many of the analyses are complex, and will require explanation and clarification before they can be considered by a broader audience. Some aspects of the dose-response modelling continue to evolve whilst other aspects vary between the developers of the

methods. Therefore COM could not conclude on the appropriateness of the different models for use with genotoxicity data.

45) With regards to the usefulness of POD's from genotoxicity data in risk assessment, COM recommend that a detailed evaluation of the different software methodologies utilised is undertaken before any conclusions could be reached. Furthermore, the COM felt that a clarification of the outstanding issues in the use of the methodology was needed (e.g choice of models, use of constraints) so that non-experts in the field were aware of the implications (if any) of the use of the different software and options that have been proposed. COM suggested that precise descriptions of the methodologies and underlying assumptions (explicit and implicit) are developed so that a detailed and informed evaluation can be undertaken by potential users of the methods.

46) Guidance is needed on how to assess data quality and goodness of fit of the models to help decide on the suitability of a dataset for modelling. Clarification is needed on the level of uncertainty in the estimates in terms of the upper to lower confidence limit ratios which are considered acceptable.

47) It is noted that there is a lack of consensus with regards to the selection of an appropriate CES/BMR for specific genotoxicity endpoints and this requires more extensive discussion and evaluation. COM felt that it was unlikely that a similar size response (e.g. 10% increase over the negative control vale) would be suitable for different genotoxicity endpoints such as, for instance, micronucleus induction and gene mutations. Further investigations of what constitutes an appropriate BMR/CES for determining BMDs using a variety of genotoxicity study types is needed, with emphasis placed on the biological relevance of the choice of BMR/CES.

48) COM remain to be convinced of the close associations in comparisons of genotoxicity and carcinogenicity data reported by some investigators and highlighted the need for a more extensive evaluation of suitable datasets including a broader assessment of different chemicals classes, genotoxicity endpoints and tissues. It was considered that BMD's from genotoxicity studies would generally expected to be lower than those from carcinogenicity studies. However, at present, there are insufficient examples and a lack of understanding of the appropriate BMRs for the various endpoints BMR's for COM to draw any definitive conclusions. Consequently, the COM, at present, were unable to make any recommendation for the inclusion of quantitative genotoxicity data in MOE calculations.

References:

Barlow S, Renwick AG, Kleiner J, Bridges JW, et al (2006) Risk assessment of substances that are both genotoxic and carcinogenic report of an International Conference organized by EFSA and WHO with support of ILSI Europe. Food Chem Toxicol.:1636-50.

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 D, Bolger PM, Carthew P, et al (2010) Application of the Margin of Exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. Food Chem Toxicol. 48 Suppl 1:S2-24.

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, et al (2014) Quantitative dose-response analysis of ethyl methanesulfonate genotoxicity in adult gpt-delta transgenic mice. Environ Mol Mutagen. 55(5):385-99.

COM (2010) Guidance: assessment of thresholds for in vivo mutagens https://www.gov.uk/government/publications/assessment-of-thresholds-for-in-vitro-mutagens

COM (2011) A strategy for testing of chemicals for genotoxicity https://www.gov.uk/government/publications/a-strategy-for-testing-of-chemicals-for-genotoxicity

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.

Dewhurst I, Renwick AG. (2013) Evaluation of the Threshold of Toxicological Concern (TTC)--challenges and approaches. Regul Toxicol Pharmacol.;65(1):168-77

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 (2011) Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment EFSA Journal 2011;9(9):2379 [69 pp.].

EFSA (2011b) TECHNICAL REPORT: Use of BMDS and PROAST software packages by EFSA Scientific Panels and Units for applying the Benchmark Dose (BMD) approach in risk assessment. EN-113. [190 pp.]. Available online: www.efsa.europa.eu http://onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2011.EN-113/pdf

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 (2017). Update: use of the benchmark dose approach in risk assessment. EFSA Journal;15(1):465 (http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2017.4658/epdf)

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.

Ji, Zhiying; LeBaron, Matthew J.; et al (2016) Dose–Response for Multiple Biomarkers of Exposure and Genotoxic Effect Following Repeated Treatment of Rats with the Alkylating Agents, MMS and MNU Mutagenesis 31(3) pp. 297–308 H.

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.

Kirkland D, Pfuhler S, Tweats D. Et al (2007) How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. Mutat Res. 2007 Mar 30;628(1):31-55.

Kroes R, Renwick AG, Cheeseman M, et al (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem Toxicol;42(1):65-83.

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(2):118-22.

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

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(8-9):792-9

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, Fellows MD, Johnson GE, Slob W (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.

Walker VE, Casciano DA, Tweats DJ. (2009) The Viracept-EMS case: impact and outlook. Toxicol Lett. 190(3):333-9.

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

WHO (2009) Environmental Health Criteria (EHC) 240: Principles and methods for the risk assessment of chemicals in food, 2009, WHO, Geneva (available at http://www.who.int/foodsafety/publications/chemical-food/en/)

Wills JW; Long AS; Johnson GE et al (2016a) 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

Wills jw; Johnson, GE; Doak, SH et al (2016b) Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations Mutagenesis. 31(3):255-63

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