Environmental and Molecular Mutagenesis

WILEY

Environmental and Molecular Mutagenesis | Environmental Mutagenesis and Genomics Society

RESEARCH ARTICLE OPEN ACCESS

Benchmark Response (BMR) Values for In Vivo Mutagenicity Endpoints

Paul A. White[1] | Guangchao Chen[2] | Nikolai Chepelev[1] | Madison A. Bell[3] | Lauren R. Gallant[3] | George E. Johnson[4] | Andreas Zeller[5] | Marc A. Beal[6] | Alexandra S. Long[7]

[1]Environmental Health Science and Research Bureau, Environmental and Radiation Health Sciences Directorate, Health Canada, Ottawa, Canada | [2]Centre for Prevention, Lifestyle and Health, RIVM (National Institute for Public Health and the Environment), Bilthoven, the Netherlands | [3]Department of Biology, University of Ottawa, Ottawa, Canada | [4]Swansea University Medical School, Swansea University, Swansea, UK | [5]Pharmaceutical Sciences, pRED Innovation Center Basel, F. Hoffmann-La Roche Ltd., Basel, Switzerland | [6]Bureau of Chemical Safety, Food and Nutrition Directorate, Health Canada, Ottawa, Canada | [7]Existing Substances Risk Assessment Bureau, Safe Environments Directorate, Health Canada, Ottawa, Canada

Correspondence: Paul A. White (paul.white@hc-sc.gc.ca)

Received: 20 January 2025 | **Revised:** 19 February 2025 | **Accepted:** 24 February 2025

Accepted by: B. Parsons

Funding: This work was supported by Rijksinstituut voor Volksgezondheid en Milieu, I/133002/02/DR and Health Canada.

ABSTRACT

The benchmark dose (BMD) approach constitutes the most effective and pragmatic strategy for the derivation of a point of departure (PoD) for comparative potency analysis, risk assessment, and regulatory decision-making. There is considerable controversy regarding the most appropriate benchmark response (BMR) for genotoxicity endpoints. This work employed the Slob (2017) Effect Size (ES) theory to define robust BMR values for the in vivo transgenic rodent (TGR) and *Pig-a* mutagenicity endpoints. An extensive database of dose–response data was prepared and curated; BMD analyses were used to determine endpoint-specific maxima (i.e., parameter *c*) and within-group variance (i.e., *var*). Detailed analyses investigated the dependence of *var* on experimental factors such as tissue, administration route, treatment duration, and post-exposure tissue sampling time. The overall lack of influence of these experimental factors on *var* permitted the determination of typical values for the endpoints investigated. Typical *var* for the TGR endpoint is 0.19; the value for the *Pig-a* endpoint is 0.29. Endpoint-specific *var* values were used to calculate endpoint-specific BMR values; the values are 47% for TGR and 60% for *Pig-a*. Endpoint-specific BMR values were also calculated using the trimmed distribution of study-specific standard deviation (SD) values for concurrent controls. Those analyses yielded endpoint-specific BMR values for the TGR and *Pig-a* endpoints of 33% and 58%, respectively. Considering the results obtained, and the in vivo genetic toxicity BMR values noted in the literature, we recommend a BMR of 50% for in vivo mutagenicity endpoints. The value can be employed to interpret mutagenicity dose–response data in a risk assessment context.

1 | Introduction

There is increasing agreement that the benchmark dose (BMD) approach constitutes the most effective and flexible strategy for deriving dose–response point-of-departure (PoD) values. BMD values can be used for margin of exposure (MOE) determination and, via extrapolation, setting human exposure limits such as the tolerable daily intake (TDI) or Permitted Daily Exposure

Paul A. White and Guangchao Chen contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 His Majesty the King in Right of Canada, F. Hoffmann-La Roche, Ltd and The Author(s). *Environmental and Molecular Mutagenesis* published by Wiley Periodicals LLC on behalf of Environmental Mutagenesis and Genomics Society. Reproduced with the permission of the Minister of Health Canada.

(PDE) (Crump 1984; Slob 2002; Haber et al. 2018; White et al. 2020, 2019; Chepelev et al. 2023; Johnson et al. 2014, 2021). Additionally, BMD values can be used for comparative analyses across covariates such as compound, tissue, and endpoint (Wills et al. 2016a, 2016b, 2017; Long et al. 2018). The approach, which was established in 1984 (Crump 1984), and is broadly used for the interpretation of toxicological dose–response data (e.g., Hardy et al. 2017; USEPA 1988; Haber et al. 2018; EFSA Scientific Committee 2022; USEPA 2022), determines the dose required to elicit a predefined response change relative to the concurrent control (i.e., background). The BMD confidence limits (i.e., BMDL and BMDU) can be used to evaluate BMD precision. The response change employed for BMD determination, which is often expressed as a fractional change relative to control (e.g., 10% increase), is referred to as the Critical Effect Size (CES) or Benchmark Response (BMR).

Several works have demonstrated the utility of the BMD covariate approach for simultaneous analysis of genetic toxicity dose–response datasets and subsequent comparisons of BMD values across experimental factors such as test article, sex, assay type, treatment duration, route of administration, and cell type (Wills et al. 2016a and 2016b; Allemang et al. 2018; Guo et al. 2018, 2016; Mittelstaedt et al. 2021). Additionally, several works have promoted quantitative interpretation of genetic toxicity dose–response data in a regulatory context; specifically, the use of genetic toxicity BMD values to derive human exposure limits and/or MOE values that can be used for risk assessment and regulatory decision-making (White et al. 2020). Indeed, genetic toxicity dose–response data have been used to derive human exposure limits for noteworthy mutagens such as ethyl methanesulfonate (EMS), *N*-methyl-*N*-nitrosourea (MNU), *N*-ethyl-*N*-nitrosourea (ENU), ethylene oxide, and selected *N*-nitrosamines (Johnson et al. 2021, 2014; Müller and Gocke 2009; Lutz 2009; Gollapudi et al. 2020). With respect to the latter, Johnson et al. (2021) employed the BMD approach, which was applied to carcinogenicity and mutagenicity dose–response data, to derive PDE values for *N*-nitrosodimethylamine (NDMA) and *N*-nitrosodiethylamine (NDEA). Chepelev et al. (2023) recently employed BMD values to calculate carcinogenicity- and mutagenicity-derived MOE values.

Despite the aforementioned works by Johnson et al. (2014), Johnson et al. (2021), Gollapudi et al. (2020), and Chepelev et al. (2023), some researchers have expressed reservations and concerns about the use of genetic toxicity BMD values for risk assessment and/or regulatory decision-making (White et al. 2020; Menz et al. 2023). More specifically, White et al. (2020) noted that prior to routine regulatory applications of BMD-based approaches, it will be necessary to determine robust BMR values for genetic toxicity endpoints (e.g., in vivo mutation and chromosome damage). Although some regulatory organizations that support the use of the BMD approach already provide BMR recommendations, there are currently no formal recommendations for genetic toxicity endpoints. In 2017, the European Food Safety Authority (EFSA) recommended a BMR of 5% for continuous toxicological endpoints such as organ weight, body weight, and hematologic parameters (Hardy et al. 2017). This recommendation was later supplanted by a tiered approach that considers: (i) the existence of a value already deemed to be biologically relevant, or (ii) quantitative determination of a biologically relevant value using methods specified in the literature, e.g., the Slob (2017) effect size (ES) theory or the United States Environmental Protection Agency's (US EPA's) 1SD approach (USEPA 2012; USEPA 2022). With respect to the latter, the US EPA has recommended a BMR of one standard deviation (SD) from the mean of the concurrent control (i.e., BMD_{1SD}) for continuous endpoints (USEPA 2012; Wignall et al. 2014; Haber et al. 2018). This 1SD approach has been criticized, particularly for endpoints with low response variability whereby it is unlikely that a 1SD change from control could be deemed adverse (Haber et al. 2018). Conversely, for endpoints with high background variability, the 1SD approach will yield higher BMR values, that is, the percentage increase corresponding to a 1SD increase above control will be relatively large (White et al. 2019). Larger BMR values will yield larger BMD values that may be less desirable from a health protection point of view. Nevertheless, the 1SD approach might be used when there is no consensus on a biologically defensible BMR for the endpoint under consideration (USEPA 2012). More recently, the US EPA's BMD determination software manual provides a list of options for denoting BMR values for continuous endpoints, e.g., fractional deviation relative to control, 1SD relative to control, and absolute deviation (USEPA 2022).

Although some researchers have employed a BMR of 10% for the analysis of in vivo mutagenicity dose–response data (e.g., Johnson et al. 2014), others have noted that 10% is generally too low for genotoxicity endpoints (Zeller et al. 2017). Specifically, using extensive analyses of endpoint-specific historical control data, Zeller et al. (2017) noted that BMR values for in vivo genotoxicity endpoints such as chromosome damage, mutation, and DNA strand breaks should be in the range of 34%–76%. White et al. (2020), who summarized the state of the science regarding BMR values for in vivo genetic toxicity endpoints, indicated that values should be in the range of 18%–79%. Indeed, Johnson et al. (2021), Douglas et al. (2022), Marchetti et al. (2021), Chepelev et al. (2023), Gollapudi et al. (2020), Bercu et al. (2023), Powley et al. (2024), Lynch et al. (2024), and Zhang et al. (2024) used a BMR of 50% in their analyses of in vivo transgenic rodent (TGR) mutagenicity dose–response data. Additionally, Bercu et al. (2023) and Zhang et al. (2024) used a BMR of 50% in their analysis of error-corrected next-generation sequencing (ecNGS) mutagenicity dose–response data. Finally, Gollapudi et al. (2020) and Chepelev et al. (2023) used a BMR of 50% in their analyses of in vivo chromosomal damage and in vivo micronucleus (MN) dose–response data, respectively.

White et al. (2020) noted the considerable inconsistency regarding the approach used to select an appropriate BMR for interpretation of genetic toxicity dose–response data. Whereas EFSA used an extensive historical overview of No-Observed-Adverse-Effect-Level (NOAEL) values to uniformly justify 5% (EFSA 2009; Hardy et al. 2017), the US EPA noted that low BMR values may necessitate undesirable extrapolation outside the range of observations (USEPA 2012). The US EPA noted that acceptance of BMR values lower than 10% should be based on biological considerations and the range of observed responses (USEPA 2012). EFSA similarly noted that "the default BMR (CES) may be modified based on statistical or biological considerations" (Hardy et al. 2017). This was later reiterated in EFSA's updated guidance on the use of the BMD approach (EFSA Scientific Committee 2022). Importantly, the work of Slob (2017)

demonstrated that robust determination of BMR values requires careful consideration of dose–response shape, and concomitant variations in selected parameters across endpoints and compounds. With respect to in vivo chromosome damage, the ES theory presented by Slob (2017) indicates that the endpoint-specific BMR value for the micronucleus (MN) frequency endpoint should be in the range of 71%–79%. Although based on more limited data, additional analyses suggested that the BMR for *Pig-a* gene mutant frequency in peripheral blood should be in the range of 58%–70%; the value for transgene (*lacI*) mutant frequency in BigBlue rat liver should be in the range of 46%–74%. In a separate analysis of TGR mutagenicity data, Wills et al. (2017) used the ES theory approach to determine that BMR values for different variants of the TGR mutagenicity assays should be in the range of 18%–66%.

According to the Slob (2017) ES theory, the maximum fold-change in response, which is designated as M, can be used to specify an endpoint-specific BMR. Specifically, an endpoint-specific BMR can be scaled to M, according to, for example, $\text{BMR} = \sqrt[4]{M}$ or $\text{BMR} = \sqrt[8]{M}$. Moreover, the theory predicts that the maximum fold-change in response will be positively related to the within-group SD according to:

$$\ln(M) = zs \tag{1}$$

where M is the maximum fold-change in response, s is the within-group SD, and z is a constant estimated to be 7 with a 90%-confidence interval of 5.9–8.9. This empirical relationship between M and s was based on the analysis of dose–response data for 27 different toxicological endpoints, including three genetic toxicity endpoints (i.e., micronucleus frequency, *lacI* mutant frequency, *Pig-a* mutant frequency). More specifically, Slob (2017) showed a positive correlation between M and s, documenting the aforementioned proportionality constant and its associated confidence limits. Importantly, since the estimated value of M is often imprecise, the value of s can be used as a surrogate for M when establishing an endpoint-specific BMR value. The value of s can generally be precisely estimated, in particular when a wide range of datasets are available for the endpoint(s) being considered. Thus, for effective BMR determination, it is essential to collect, collate, curate, and analyze large databases containing dose–response data across a wide range of studies, test articles, and experimental factors such as tissue, species, administration route, treatment duration, and post-exposure tissue sampling time.

In this work, we used an extensive collection of publicly available dose–response data to calculate robust endpoint-specific BMR values for the TGR somatic and germ cell gene mutation assays and the mammalian erythrocyte *Pig-a* gene mutation assay (OECD 2022a, OECD 2022b). The PROAST software (www.rivm/nl/proast) was used for the majority of the analyses; for continuous data, PROAST denotes the maximum fold change of the response as parameter *c* and the within-group variance as *var*. Initial analyses examined the dependence of these parameters on experimental factors such as tissue, treatment duration, administration route, and post-exposure tissue sampling time. Subsequent analyses used robust *var* estimates and their confidence limits to determine endpoint-specific BMR values. Additional analyses, which were conducted using SAS software v9.4 (SAS Institute Inc. Cary, NC), determined BMR values in accordance with the aforementioned 1SD approach, that is, values based on the mean and SD of study-specific control values. These values were compared with those based on the Slob (2017) ES theory. The work constitutes an important step forward toward formal adoption of quantitative methods for interpretation of genetic toxicity dose–response data in a risk assessment context.

2 | Materials and Methods

Detailed dose–response data were collected from the peer-reviewed literature and/or extracted from existing databases. For the TGR somatic and germ cell mutations assays, the majority of the data were extracted from the Transgenic Rodent Assay Information Database (TRAiD) housed at Health Canada (Lambert et al. 2005; Kirkland et al. 2019). For the mammalian erythrocyte *Pig-a* gene mutation assay, the majority of the data were obtained from the database housed on a University of Maryland server (www.pharmacy.umaryland.edu/centers/cersi-files) (Shemansky et al. 2019). These databases contain data from the peer-reviewed literature; they were supplemented with data published in theses (e.g., Singer 2006) and/or unpublished data provided by colleagues. In both cases, the scientific literature was searched to identify more recent publications; data were extracted, collated, and added to the existing databases. Where data were only presented in figures, required values were extracted using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer). For the dose–response analysis, an individual dataset was defined as the lowest stratum of the dose–response data, that is, apart from test article dose, the levels of the recorded experimental factors are the same. For example, for a given test article dose, variables such as sex, administration route, treatment duration, tissue, and post-exposure tissue sampling time are the same. The TGR dose–response database employed for the analyses included a total of 9472 records from 527 dose–response datasets associated with 148 studies of 114 test articles. The *Pig-a* dose–response database employed for the analyses included a total of 7813 records from 391 dose–response datasets associated with 147 studies of 43 test articles. Both databases are freely available from the corresponding author.

All selected datasets were required to have a minimum of 2 doses plus control; the data collected included individual animal values (i.e., replicate-type) or dose-group summaries (i.e., mean, group size, and variance). For the purposes of the analyses conducted using PROAST v70.2 (www.rivm.nl/en/proast) in the R computing environment (v3.6.3, R Core Team 2022), all replicate-type data were converted to summary data prior to the analyses described below. The selected datasets all included information about test article, species, sex, tissue, study design, treatment duration, post-exposure sampling time, and administration route. The TGR database also included information about TGR assay type (e.g., MutaMouse, BigBlue, etc) and transgene (e.g., *lacZ*, *lacI*, *cII*, *gpt*). The *Pig-a* database also included information about animal strain. Rat red blood cell (RBC) *Pig-a* gene mutant frequency values are expressed per 10^6 scored cells (i.e., total RBCs). Tissue-specific transgene mutant frequency values are expressed per 10^5 plaque-forming units.

All dose values are expressed as mg/kg BW/day; where necessary, age-specific body weight values were obtained from the

scientific literature. In addition, published food or drinking water daily consumption rates were used to convert concentrations to dose rate values in mg/kg BW/day. For inhalation studies, concentration values were converted to dose rates in mg/kg BW/day using published tidal volume and respiration rate values. Concentration values expressed as ppm or ppb were converted to mg per cubic meter using molecular weight and molar volume at standard temperature and pressure (https://www.cdc.gov/niosh/docs/2004-101/calc.html). The required rodent physiological data were obtained from Rocha et al. (2004), Bachmanov et al. (2002), USEPA (1988), Sucko et al. (2001), Strohl et al. (1997), and Mauderly (1986). Before investigating the effect of tissue on model parameters *c* and *var* (see below), tissue labels were simplified by combining some tissue subtypes. For example, where tissue was listed as a specific section of the small intestine (e.g., duodenum), the term *small intestine* was used as the tissue label.

Since datasets without a significant trend do not contribute to information on the maximum fold-change, a preselection of datasets was performed to improve the efficiency of the analyses, that is, only datasets with a significantly increasing response trend were selected for further analysis using PROAST. Model 15 in PROAST (i.e., a 4-parameter exponential model) was used for the preselection of datasets:

$$y = a \cdot c^{1-e^{-\left(\frac{x}{b}\right)^d}} \quad (2)$$

where x is dose, y is the response, parameter a is the mean response at dose zero (i.e., the background), b is the potency parameter, c is the maximum fold-change relative to the background, and d is the steepness of the curve. This model was previously shown to be appropriate for describing a wide range of toxicological dose–response relationships, including those pertaining to genotoxicity endpoints (i.e., in vivo and in vitro micronucleus frequency) (Slob and Setzer 2014). When model 15 is applied, PROAST directly estimates the BMD instead of parameter b, that is, BMD is an output parameter and parameter b is an intermediate parameter. The relationship between BMR and the model parameters can be expressed as:

$$1 + BMR = c^{1-e^{-\left(\frac{BMD}{b}\right)^d}} \quad (3)$$

For the dataset preselection, the following criteria were applied:

i. The BMD is less than 10,000 mg/kg bw/day. This value was chosen by scrutinizing the ranges of administered doses in the collected data and taking a rounded-up upper limit.

ii. The lower limit for the maximum fold-change of response is 11, that is, the value of parameter *c* is larger than 11. This choice was based on preliminary analyses where one single *c* was estimated on the basis of more than 200 individual datasets;

iii. The limit for parameter *d* (i.e., steepness of the dose–response curve) is 5. This choice follows the empirical analysis presented in Slob and Setzer (2014);

iv. The response increases with increasing doses.

With respect to (iii) above, Slob and Setzer (2014) noted that parameter *d* (i.e., the steepness of the of the curve) can be assumed to be constant for a given continuous endpoint. After preselection, the number of retained datasets for the TGR and *Pig-a* endpoints was 171 and 172, respectively (Table 1, Table 2).

As noted earlier, the Slob (2017) ES theory uses information on *var* and *c* to estimate an endpoint-specific BMR. However, before using all the available data for each endpoint to determine single, endpoint-specific *var* and *c* values, it is necessary to investigate whether *var* and *c* are influenced by experimental factors. More specifically, it is necessary to conduct analyses that determine whether it is reasonable to determine a universal maximum fold-change (i.e., parameter *c*) and within-group variance (i.e., *var*) for a given toxicological endpoint, that is, regardless of experimental factors such as tissue, administration route, treatment duration, etc. Consequently, prior to determining single *var* and *c* values for each endpoint, the effects of each experimental factor were incrementally investigated. More specifically, we comparatively examined the 90%-confidence intervals (CIs) of *var* and *c* across each level of each experimental factor investigated. Taking tissue as an example, tissue was used as a covariate, and the CIs for *var* and *c* were subsequently compared across each of the tissue types examined (e.g., liver, kidney, bone marrow, etc.). Analogously, all other experimental factors were investigated to critically examine their influence on parameters *var* and *c*. In all cases, the dataset was set as a covariate for parameters *a* and *b*.

Final determination of endpoint-specific values for *var* and *c* required simultaneous dose–response analysis of all preselected datasets, that is, single *var* and *c* values that can be used for BMR determination. With respect to the available data, these single *var* and *c* values can be viewed as typical for the endpoint.

TABLE 1 | Overview of the preselected TGR mutagenicity dose–response data used for BMD analyses (171 datasets).

Factors	Levels of factor
Species	Mouse, rat
Transgene	*cII*, *gpt*, *lacI*, *lacZ*, spi selection
TGR	BigBlue mouse, BigBlue rat, *gpt* delta mouse, *gpt* delta rat, MutaMouse
Tissues	Bone marrow, brain, colon, kidney, liver, lung, mammary gland, omentum, ovary, skin, small intestine, sperm, spleen, stomach, testes
Administration route	Diet, drinking water, gavage, inhalation, intratracheal instillation (*it*), intraperitoneal injection (*ip*), subcutaneous administration, topical administration
Treatment duration (days)	1, 3, 4, 5, 8, 10, 11, 14, 21, 28, 30, 56, 60, 84, 90, 91, 112, 336
Post-exposure sampling time (days)	0, 1, 2, 3, 7, 10, 14, 21, 28, 30, 33, 35, 42, 49, 56, 70, 73, 84, 98, 100

TABLE 2 | Overview of the preselected *Pig-a* RBC mutagenicity dose–response data used for BMD analysis (172 datasets).

Factors	Levels of factor
Species	Rat
Strain	Sprague–Dawley, Fischer 344, Wistar Han
Administration route	Gavage, intraperitoneal injection (*ip*), intravenous injection (*iv*)
Treatment duration (days)	1, 3, 9, 15, 16, 28
Post-exposure sampling time (days)	0, 4, 7, 13, 15, 18, 21, 27, 29, 39, 42, 49, 57, 87

PROAST model 15 was used for the final analysis, setting the dataset as a covariate for parameters *a* and *b*. In accordance with the results obtained (see below), parameters *var* and *c* were assumed to be constant across all datasets.

Additional analyses employed the aforementioned 1SD approach (USEPA 2012) to determine endpoint-specific BMRs that could be compared with those calculated using the ES theory outlined above. Accordingly, the mean and SD of replicate background mutant frequency values were calculated for each dataset. Prior to analyses and interpretation, the distribution of study-specific SD values was examined using a normal probability plot, and values were log transformed to generate a normal distribution. Similar to Zeller et al. (2017), the distribution of study-specific SD values was truncated before the determination of endpoint-specific BMR values; truncation involved the removal of the upper- and lower-most 5% of SD values in the distribution. Like the Zeller et al. analyses, such values were assumed to be technical or biological outliers. Geometric mean BMR values were calculated across all the SD values included in the truncated dataset. In total, 521 datasets containing replicate animal values were examined for the TGR assay; 382 datasets for the *Pig-a* assay. Analyses were conducted using SAS software v9.4 for Windows (SAS Institute Inc., Cary, NC). Figures illustrating the distribution of study-specific BMR values were prepared using *ggplot2* v3.5.1 in RStudio (RStudio Team 2023).

3 | Results

Tables 1 and 2 provide descriptive overviews of the databases containing the collected dose–response data. Table entries are based on the preselected data employed for the described BMD analyses.

The data preselection resulted in 171 TGR datasets showing a significant dose–response relationship (Table 1). The experimental factors of primary importance for the TGR data include species, transgene, TGR, tissue, administration route, treatment duration, and post-exposure sampling time (OECD 2022a). The levels of many of these factors are very heterogeneous (i.e., tissue, administration route, treatment duration, and post-exposure sampling time); the influence of these experimental factors on *var* was appropriately investigated (see below).

The collected TGR dose–response data includes a wide range of tissues, including bone marrow, brain, colon, kidney, liver, lung, mammary gland, omentum, ovary, skin, small intestine, sperm, spleen, stomach, and testes (Table 1 and Figure 1A). Accordingly, the data include diverse somatic tissues with different proliferation rates, as well as sperm. Figure 1A illustrates the influence of tissue on *var*. The figure shows that the most examined tissues are liver (i.e., 54 datasets) and bone marrow (i.e., 33 datasets), which account for approximately half of the analyzed datasets. This is aligned with the eventual recommendation to choose liver in the absence of background knowledge for tissue collection, plus at least one rapidly dividing tissue, such as bone marrow or glandular stomach (OECD 2022a). Besides liver and bone marrow, small intestine, lung, stomach, and spleen were commonly selected tissues. Consistent with data availability, the *var* values for liver and bone marrow have CIs compared to those for tissues with fewer data. The estimated *var* for all tissues lies mainly in the 0.1–0.3 range; no appreciable cross-tissue differences were observed. Considering these results, and the availability of the dose–response data, the assumption that tissue does not influence the *var* of transgene mutant frequency cannot be rejected (Figure 1A).

The collected dose–response data also include a variety of administration routes (Table 1 and Figure 1B). These include diet, drinking water, gavage, inhalation, intratracheal instillation (*it*), intraperitoneal injection (*ip*), subcutaneous injection, and topical. Figure 1B illustrates the influence of the administration route on *var*. Among the 171 preselected TGR datasets, gavage accounts for most of the datasets (i.e., 114). The next most common routes of administration are *ip*, diet, and drinking water. It is normally assumed that studies that are more prone to dosing variability, such as diet, drinking water, and inhalation, will be associated with higher dose-specific variance in comparison with gavage studies (Jones et al. 2016; Turner et al. 2011). With respect to diet and drinking water studies, the calculated doses are dependent on average cage-specific daily consumption rates and medium test article concentrations; the exact amount consumed by each animal is generally not known (e.g., Manjanatha et al. 2015, 2006; Dragsted et al. 2002). Similarly, inhalation studies assume a constant airborne concentration of the test article; average inhalation rate and tidal volume values are used to calculate dose rate values. This can also lead to deviations from the actual per-animal dose. Nevertheless, the *var* value for inhalation was in the same range as that observed for other routes of administration.

The estimated *var* for gavage exposures is 0.159. Given the relatively large amount of data for gavage studies and the implicit dosing precision, the value has a very narrow CI (i.e., 0.152–0.166, Figure 1B). Apart from drinking water and diet, the estimated *var* for different administration routes is mostly within the 0.1–0.3 range. To check if the deviation of *var* for diet and drinking water could be caused by other factors instead of the administration route alone, the data were examined more carefully. With respect to diet, the data included both rat and mouse, three types of transgene scoring (i.e., *lacI*, *gpt*, and spi selection), two tissues (i.e., liver and colon), seven different treatment durations (i.e., 28, 30, 60, 84, 90, 91, and 112 days), and three different post-exposure sampling times (i.e., 0, 7, and 14 days). Similarly, for drinking water, the data included both rat and mouse, four types of transgene scoring (i.e., *cII*, *gpt*, *lacI*, and spi), four tissues (i.e.,

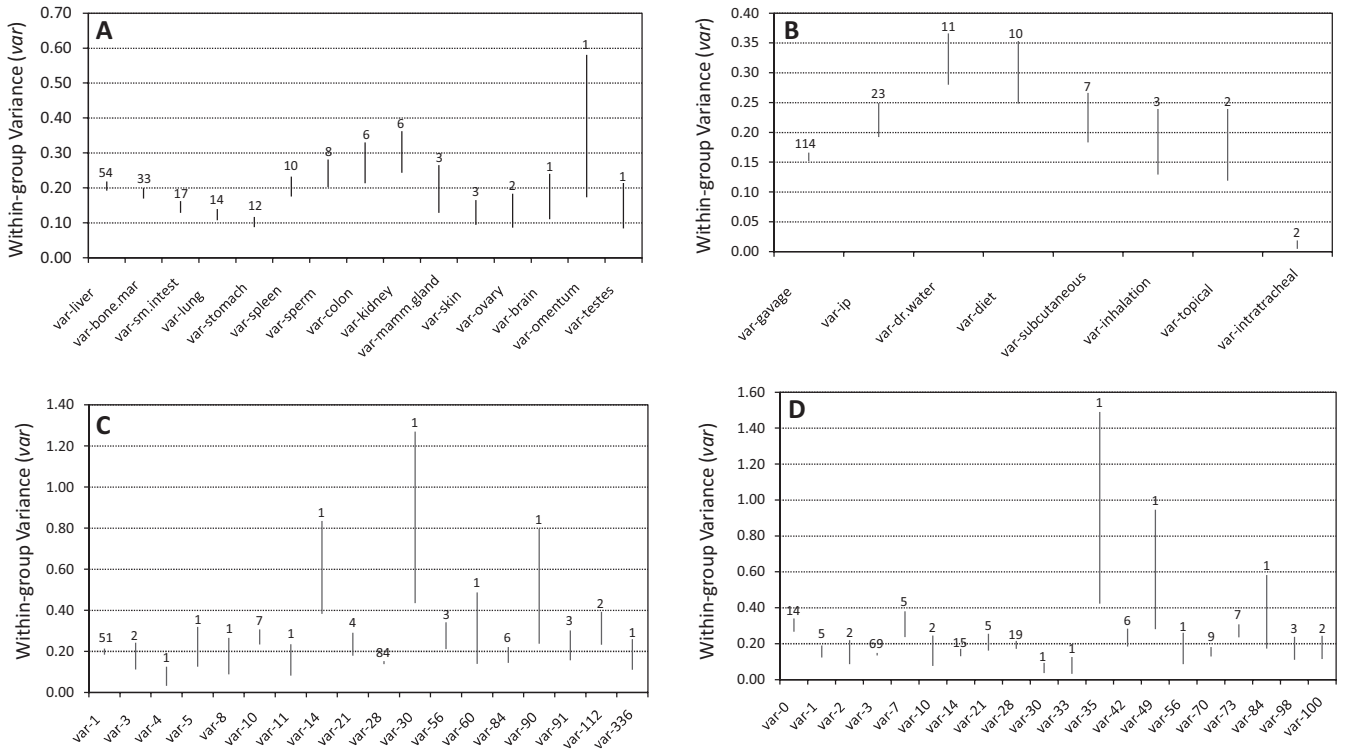

FIGURE 1 | The influence of tissue (A), administration route (B), treatment duration (days) (C), and post-exposure sampling time (days) (D) on the within-group variance (*var*) for the Transgenic Rodent (TGR) dose-response data. Bar extremes indicate upper and lower 90% confidence limits. The values above the bars indicate the number of datasets within each level of the factor.

kidney, liver, lung, and sperm), four treatment durations (i.e., 14, 28, 56, and 112 days), and two different post-exposure sampling times (i.e., 0 and 21 days). Therefore, it does not seem likely that factors other than the administration route (i.e., species, transgene, tissue, treatment duration, post-exposure sampling time) could explain the decreased relative precision of *var* values for diet and drinking water. This is consistent with the assertion that the assumed dosing uncertainty for diet and drinking water studies (Bachmanov et al. 2002; Turner et al. 2011; Jones et al. 2016; Manjanatha et al. 2015, 2006; Dragsted et al. 2002) can translate into larger variability in the observed individual responses (i.e., lower *var* precision). However, considering the limited data availability, the assumption needs to be further verified when more suitable data become available. Overall, the assumption that the administration route does not have an appreciable effect on *var* cannot be rejected (Figure 1B).

Another important factor that could influence *var* is the treatment duration. For the TGR data (Table 1, Figure 1C), the treatment durations employed range from 1 day to 336 days. As illustrated in Figure 1C, the 1- and 28-day studies dominate the database (i.e., total of 135 datasets). This is not surprising since the latter is aligned with the repeated-dose regimen recommended in OECD Test Guideline 488 (OECD 2022a). The recommendation is consistent with the need to permit the accumulation of detectable levels of mutants in slowly proliferating tissues. Besides 1 and 28 days, less common treatment duration times include 10, 21, 56, and 84 days.

The values and CIs for *var* for different treatment durations were determined, and the results are displayed in Figure 1C. To determine if *var* is affected by treatment duration, the values are displayed from short durations on the left to long durations on the right. As indicated in Figure 1C, *var* remains mainly in the 0.1–0.3 range regardless of the treatment duration. This applies to the 1- and 28-day studies, which have the most individual datasets, as well as other studies with lower data availability (i.e., 3, 8, 11, 84, and 336 days). It does appear that *var* values associated with 14, 30, 60, and 90 days have relatively wider CIs. Even though data availability could be a possible explanation (i.e., only one individual dataset for each of these four treatment durations), it should be noted that the administration routes for these four datasets were diet or drinking water. Thus, the source of the variability in *var* for these datasets may indeed be associated with the administration route. Overall, the assumption that *var* is not appreciably influenced by increases in treatment duration cannot be rejected (Figure 1C).

The last factor of interest is the post-exposure tissue sampling time. In the collected datasets, sampling times vary widely from 0 to 100 days (Table 1, Figure 1D). As illustrated in Figure 1D, the most commonly employed sampling time is three days (i.e., 69 datasets), followed by 28 days (i.e., 19 datasets), 14 days (i.e., 15 datasets), and 0 days (i.e., 14 datasets). The sampling time employed for TGR assays is critical for the fixation of mutations; the required fixation time relates to the tissue-specific mitotic index (White et al. 2017). For instance, bone marrow and intestine can respond more rapidly relative to a tissue such as the liver. For slowly proliferating tissues, a 28-day sampling time is recommended; 3 days is recommended for rapidly proliferating tissues (OECD 2022a). This explains the preponderance of 3- and 28-day sampling times in the database. In Figure 1D, the *var* estimate

for the different sampling times is plotted from short durations on the left to long durations on the right. Similar to the results obtained for treatment duration, the data do not show any significant trend, that is, *var* is not dependent on post-exposure tissue sampling time. That said, the limited data availability likely contributed to the wide CIs for 35 and 49 days. Overall, *var* lies in the 0.1–0.3 range, and the assumption that sampling time does not impact *var* cannot be rejected (Figure 1D).

Additional analyses investigated the influence of sex, TGR rodent model, and transgene on *var*; the results obtained also indicated that these experimental factors do not influence the *var* value (results not shown).

Following the aforementioned preselection process, 172 *Pig-a* dose–response datasets were available for detailed analysis. These datasets include relatively fewer experimental factors compared to those in the TGR database. The primary experimental factors pertaining to the *Pig-a* datasets are the administration route, treatment duration, and post-exposure tissue sampling time. The selected *Pig-a* datasets include only rat RBC dose–response data.

For the 172 selected *Pig-a* dose–response datasets, only three administration routes are included, that is, gavage, *ip*, and *iv* (Table 2, Figure 2A). Similar to that observed for the TGR data, and as illustrated in Figure 2A, gavage is the most commonly employed route of administration (i.e., 152 datasets). The data for *ip* and *iv* are comparatively marginal, i.e., 18 and 2 datasets, respectively. The estimated *var* for gavage is 0.303; considering the amount of data and the precision generally associated with the administration strategy, the confidence limit was relatively narrow (i.e., 0.291–0.317). The estimated *var* for *ip* is 0.424 (0.380–0.476); for *iv*, it is 0.220 (0.160–0.314). The CI for *ip* does not overlap with that of gavage and *iv* (Figure 2A); however, considering the amount of available data, it is not possible to conclude whether this is due to the influence of the administration route or the scarcity of dose–response data. Thus, the assumption that the administration route does not appreciably influence *var* cannot be rejected (Figure 2A).

The treatment durations included in the preselected *Pig-a* dose–response datasets range from 1 day to 28 days (Table 2, Figure 2B); the range is narrow compared to that observed for the TGR datasets (Table 1). As illustrated in Figure 2B, the 3- and 28-day studies are the most common in the selected *Pig-a* data; together they constitute 125 out of the 172 datasets analyzed. That said, 1 administration day is not uncommon in the analyzed data; less commonly employed treatment durations include 9, 15, and 16 days. The *var* values across different treatment durations are displayed in Figure 2B; the results are plotted with short durations on the left and long durations on the right. The results indicate that, regardless of data availability, *var* is in the 0.15–0.35 range. Overall, the data analyses revealed that the assumption that treatment duration does not appreciably influence *var* cannot be rejected (Figure 2B).

The post-exposure sampling time in the *Pig-a* database ranges from 0 to 87 days (Table 2, Figure 2C). As illustrated in Figure 2C, the most common sampling times are one day (i.e., 23 datasets), followed by 12 days (i.e., 22 datasets), 27 days (i.e., 18 datasets), and 28 days (i.e., 17 datasets). The *var* values for the different sampling times are displayed in Figure 2C; the lowest values are shown on the left and higher values on the right. Similar to that

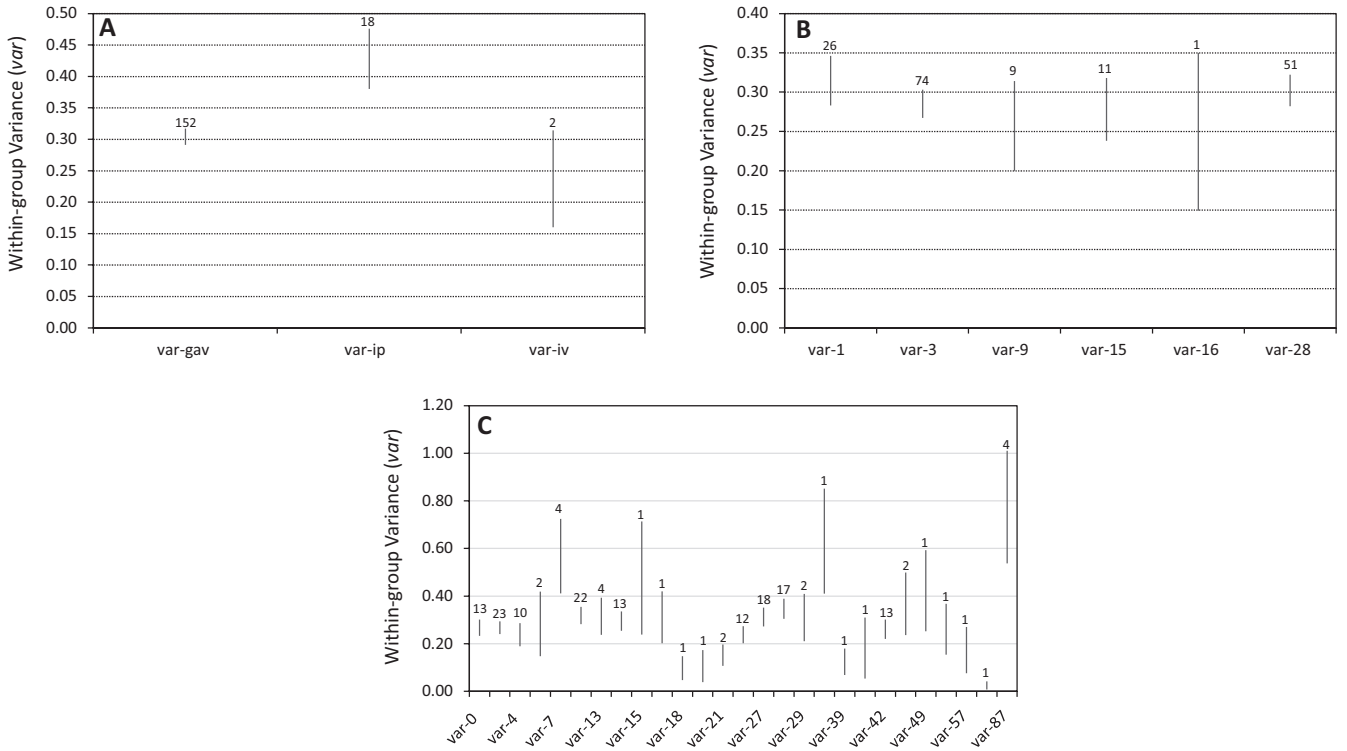

FIGURE 2 | The influence of administration route (A), treatment duration (days) (B), and post-exposure sampling time (days) (C) on the within-group variance (*var*) for the *Pig-a* gene mutation dose-response data. Bar extremes indicate upper and lower 90% confidence limits. The values above the bars indicate the number of datasets within each level of the factor.

observed for the effect of post-exposure sampling time for the TGR data, we did not observe any trend in *var* values with increasing sampling time; the overall *var* is in the 0.2–0.4 range. Thus, the assumption that the post-exposure sampling time does not appreciably influence *var* cannot be rejected (Figure 2C).

In summary, based on detailed analyses of the preselected TGR and *Pig-a* dose–response datasets, we can conclude that experimental factors such as tissue, administration route, treatment duration, and post-exposure tissue sampling time do not significantly influence *var*. Nevertheless, it does appear that TGR studies that employed dietary or drinking water administration routes may be associated with higher *var* values. However, considering the limited data availability, this assertion would need to be verified when more data become available. In addition to analyses that investigated the effects of experimental factors on *var*, the influence of experimental factors on maximum fold-change (i.e., parameter *c*) also needs to be investigated. The results obtained from that analysis are described below.

In principle, the influence of experimental factors on parameter *c* could be examined in the same manner as that conducted for *var* (Figures 1 and 2), i.e., set *c* dependent on factors such as tissue, administration route, treatment duration, or post-exposure sampling time. These analyses were conducted, and the results revealed that the precision of estimated *c* values is low, with CI values across different levels of certain factors often spanning several orders of magnitude (results not shown). Taking tissue for the TGR data as an example, for organs with relatively more data, *c* was estimated to be 24.6 for lung (i.e., 14 datasets), 90.4 for liver (i.e., 54 datasets), and 154.1 for bone marrow (i.e., 33 datasets). When considering the tissues with very limited data, the range of *c* is much larger across the different tissues examined, for example, 242,600 for omentum and 10,320 for testes. Such extreme values are likely due to the fact that the data used for the analysis are not well defined at the upper end of the dose–response relationship. To verify this assertion, we examined normalized dose–response relationships, whereby response is normalized according to the background (i.e., parameter *a*) for each dataset, and dose is normalized using the BMD10 for each dataset. The relationship, which is illustrated in Figure S1, supports the contention that the datasets contain limited information regarding maximum response (i.e., parameter *c*). Consequently, as noted earlier, the estimate of *c* is often imprecise, while *var*, and *s* by extension, can often be estimated more precisely. This is particularly true when a wide range of datasets is available. Therefore, in congruence with the aforementioned ES theory, it was deemed prudent to use *s* as a surrogate of *c*; subsequently employing *s* to determine endpoint-specific BMR values.

Since the experimental factors investigated do not appear to have a significant influence on *var*, the next step was the determination of a typical *var* for each endpoint. This required the simultaneous analysis of all preselected dose–response data with parameters *a* and *b* set as dataset-dependent; the single *var* value across all datasets can be used for calculating the endpoint-specific BMR. Model 15 in the PROAST software (i.e., the aforementioned 4-parameter exponential model) was used to analyze the data; the results obtained are displayed in Figure 3A,B.

Based on all 171 individual TGR datasets, the results revealed a single *var* of 0.191, with a narrow CI of 0.184–0.198. Based on all 172 *Pig-a* datasets, the results revealed a single *var* of 0.29, with a narrow CI of 0.28–0.30. Due to the aforementioned extreme *c* values, and the related fact that the datasets contain comparatively little information about maximum response (Figure S1), endpoint-specific *var* values were preferentially used to determine endpoint-specific BMR values. Applying the ES theory, the

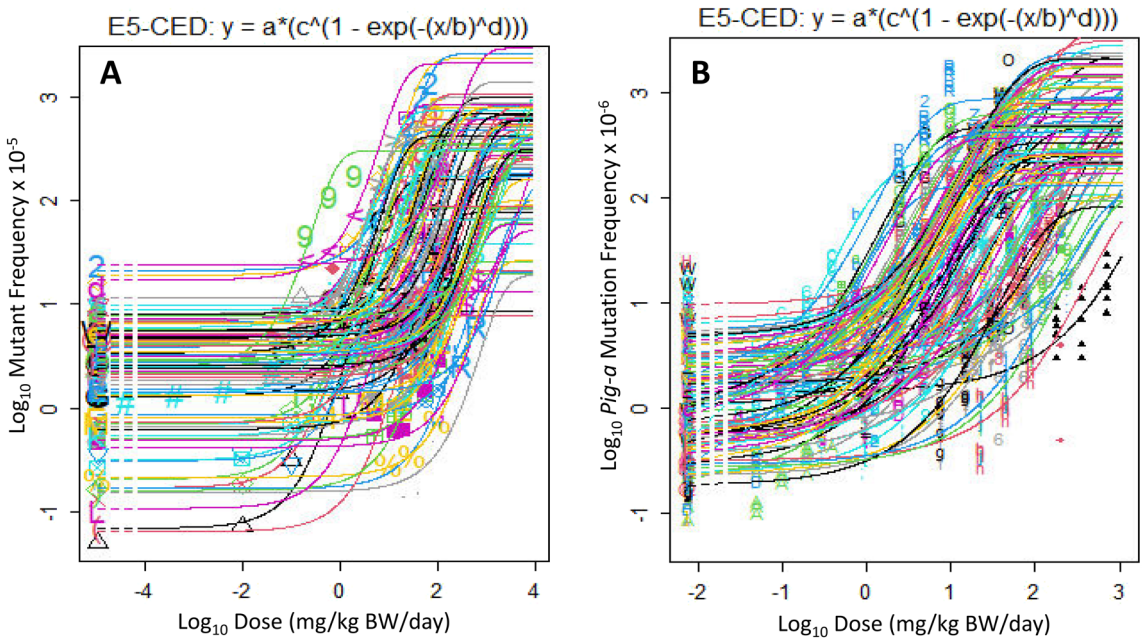

FIGURE 3 | PROAST output of BMD analysis using the selected TGR (A) and *Pig-a* (B) data. In each analysis, one single *var* and one single *c* were estimated together with their respective 90% confidence intervals. Parameters *a* and *b* were dataset dependant. The result provide the typical *var* values required for calculation of endpoint-specific BMR; *var* values are 0.191 for the TGR assays and 0.292 for the *Pig-a* assay (Table 3).

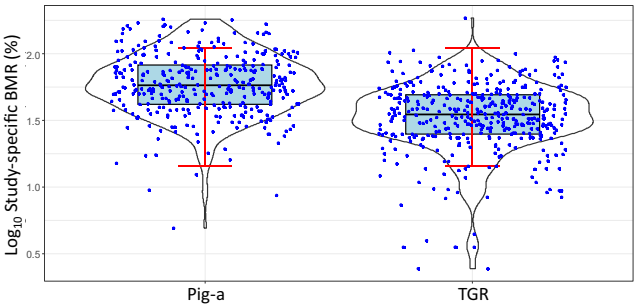

FIGURE 4 | The distribution of study-specific Benchmark Response (BMR) values determined using the 1SD approach. Values are expressed as percentage increases above background, and BMR determination employed the trimmed distribution of study-specific SD values. The horizontal lines in the boxes are median values; the box limits display the interquartile range, that is, 25th and 75th percentiles. The lower and upper whiskers indicate the 5th and 95th percentiles, respectively. Dots indicate each observation; the overlay indicates the density of observations across the range of BMR values. The results are based on 521 TGR datasets and 382 Pig-a datasets. The horizontal distribution of observations was adjusted to improve visibility.

endpoint-specific BMR values based on *s* (i.e., the square route of *var*), which are expressed as a percentage, were calculated according to Equation (4):

$$\mathrm{BMR} = e^{\left(\frac{1}{8}*7*s\right)} - 1 \tag{4}$$

The calculated BMR values for the TGR and *Pig-a* mutagenicity endpoints are 47% and 60%, respectively (Table 3). Following Slob (2017), the analyses assumed that the 8th root of *M* can be considered a small effect.

To compare these BMR values with those that would be associated with the aforementioned 1SD approach, the BMR was also calculated using study-specific mean and SD values for the background response (i.e., concurrent controls). Study-specific BMR values were calculated according to Equation (5):

$$\text{Study } BMR\,(\%) = 100 * \left(\frac{\text{SD backround}}{\text{mean background}}\right) \tag{5}$$

Prior to BMR calculation, the distribution of study-specific SD values was screened for normality, and the upper- and lower-most 5% were removed. The effects of experimental factors on the SD of concurrent replicate controls were also investigated, for example effects of tissue, administration route, treatment duration, and post-exposure sampling time. The results obtained revealed that study-specific SD values are stable across all factors examined (results not shown). That said, with respect to the *Pig-a* mutagenicity assay, the analyses did reveal significantly lower SD values for studies that employed inhalation exposure. However, the data for this route of administration are very limited, that is, $N = 10$ datasets. Figure 4 displays the distribution of study-specific BMR values calculated using Equation (5); the calculation employed the aforementioned trimmed distribution of SD values. The displayed results were generated by analyses of 521 TGR datasets and 382 *Pig-a* datasets. The geometric mean BMR based on the trimmed SD distributions are 33% and 58% for the TGR and *Pig-a* mutagenicity datasets, respectively (Table 3). Interestingly, the confidence limits of the study-specific background SD values are much greater than *s*, that is, the value derived from within-group variance *var*. Consequently, the range of BMR values calculated according to the 1SD approach (Equation 5) is much wider than those calculated using *s* (Equation 4).

TABLE 3 | Typical *var* and control SD values for the TGR and *Pig-a* mutation endpoints. Endpoint-specific BMR values were calculated according to Equations (4) and (5); *s* was used as a surrogate for *c*. Values in parentheses are the 90% confidence limits. Values based on study-specific background responses are geometric means.

	TGR	*Pig-a*
var	0.19 (0.18–0.20)	0.29 (0.28–0.30)
BMR based on *s*	47% (46%–48%)	60% (59%–62%)
Mean background SD	1.08 (0.13–3.99)	0.58 (0.082–3.25)
Mean BMR based on background SD	33% (11%–77%)	58% (26%–142%)

4 | Discussion

White et al. (2020) outlined knowledge gaps impeding the use of genetic toxicity dose–response data for risk assessment and regulatory decision-making, e.g., derivation of human exposure limits and/or MOE values. Among the primary knowledge gaps is the lack of suitably determined endpoint-specific BMR values. Several works, including some by regulatory authorities, do provide guidance regarding BMR selection (e.g., Hardy et al. 2017; Haber et al. 2018; USEPA 2012; White et al. 2020; EFSA Scientific Committee 2022; USEPA 2022); however, there is still no uniform agreement on the most appropriate BMR values for genetic toxicity endpoints.

The aforementioned work of Slob (2017) indicates that robust determination of suitable BMR values requires careful consideration of dose–response shape and related variations in dose–response function parameters across endpoints and compounds. The approach requires collection, collation, curation, and analysis of large data sets containing dose–response data across a wide range of studies and test articles; moreover, across studies that have examined a wide range of experimental factors. Aside from the limited analyses mentioned earlier (e.g., Slob 2017; Wills et al. 2017; Zeller et al. 2017), investigations have not previously been conducted for genotoxicity endpoints. Consequently, there is no consensus regarding appropriate BMR values for genotoxicity endpoints. Therefore, this study used the Slob (2017) ES theory, applied to an extensive collection of dose–response data, to determine robust BMR values for the in vivo TGR and *Pig-a* gene mutation endpoints. More specifically, the Slob (2017) ES approach was employed to determine endpoint-specific maximum fold-change (i.e., *c*) and within-group variance (i.e., *var*) for the in vivo TGR and *Pig-a* data; thereby permitting the use of *var* to calculate endpoint-specific BMR values according to Equation (4).

The results based on endpoint-specific *var* values (Table 3) indicate that endpoint-specific BMR values for the in vivo TGR and *Pig-a* mutagenicity assays should be 47% and 60%, respectively. Results

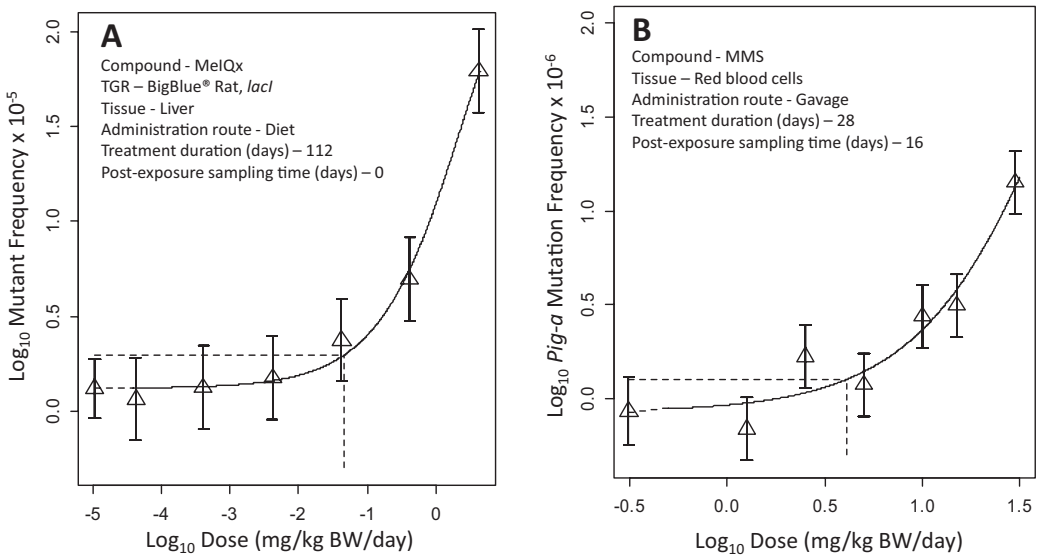

FIGURE 5 | Use of an endpoint-specific BMR of 50% for analyses of TGR assay (A) and *Pig-a* mutagenicity assay (B) dose-response data. Dose-response analyses yielded BMD values of 0.047 (left) and 4.12 (right) mg/kg/day, respectively. The test article, tissue, administration route, treatment duration, and post-exposure sampling time are indicated. The triangles indicate dose-specific mean responses; upper and lower confidence limits are also shown. The vertical dotted lines indicate the BMDs; the horizontal dotted lines indicate the BMRs. MeIQx: 2-Amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline. MMS: methyl methanesulfonate.

based on the distribution of study-specific SD values for background responses indicate that endpoint-specific BMR values for the in vivo TGR and *Pig-a* mutagenicity assays should be 33% and 58%, respectively. Importantly, scrutiny of a wide range of experimental factors was employed to investigate the covariate dependence of *var*; the results obtained justify the ability to specify single BMR values for each of the investigated endpoints. Similarly, the results obtained did not reveal significant variation in the SD of control values across experimental factors such as tissue, administration route, treatment duration, and post-exposure sampling time. Evaluation alongside the genetic toxicity BMR recommendations published by Slob (2017), Zeller et al. (2017), and Wills et al. (2017) (i.e., 18%–79%) triggers specification of a pragmatic uniform BMR of 50% for in vivo genetic toxicity endpoints. The value is consistent with those specified by Slob (2017) and Wills et al. (2017), that is, 18%–66% and 46%–79%, respectively. Moreover, it is consistent with the values specified by Zeller et al. (2017) for the mutagenicity endpoints investigated therein (i.e., 42%–76%). Lastly, it is consistent with typical BMR values that would be aligned with the 1SD approach advocated by the US EPA (USEPA 2012), that is, 33% and 58%. Interestingly, several works have already adopted 50% as a BMR for the analysis and interpretation of in vivo mutagenicity dose–response data (i.e., Johnson et al. 2021; Chepelev et al. 2023; Douglas et al. 2022; Marchetti et al. 2021; Gollapudi et al. 2020; Zhang et al. 2024; Lynch et al. 2024; Powley et al. 2024; Bercu et al. 2023). Figure 5 shows examples of the application of a BMR of 50% for analyses of dose–response data for the TGR and *Pig-a* mutagenicity endpoints.

It is important to note that the Slob (2017) ES approach has hitherto not been employed to determine a robust BMR value for the in vivo comet assay. Nevertheless, the work of Zeller et al. (2017) did indicate that the minimum endpoint-specific BMR values for in vivo comet endpoints should be in the range of 39%–75%. Going forward, it will be necessary to collect, collate, curate, and analyze comet assay dose–response data in the same fashion as that conducted herein for in vivo mutagenicity endpoints. Although Slob (2017) examined data for the in vivo micronucleus (MN) endpoint, determining a *var* value that yields a BMR of 75% (i.e., 71%–79%), it may also be useful to revisit that endpoint using a more extensive dose–response dataset. In addition, the Slob (2017) ES approach could be used to determine robust BMR values for novel in vivo genetic toxicity endpoints. This is perhaps most important for ecNGS endpoints used to precisely determine mutation frequency at any locus in virtually any tissue of any organism (Marchetti et al. 2023). Bercu et al. (2023) and Zhang et al. (2024) have already employed a BMR of 50% in their analysis of ecNGS mutagenicity dose–response data.

Further work should also examine dose–response data for in vitro genotoxicity endpoints; robust BMR values for those endpoints could be employed to derive BMD values that can be used to calculate human Administered Equivalent Doses (AEDs) via in vitro-to-in vivo extrapolation (IVIVE) (Beal et al. 2022, 2023; Kuo et al. 2022). AED values can subsequently be compared with in vivo BMD values, human exposure values, and, more generally, can be used to support the interpretation of in vitro genotoxicity dose–response data in a risk assessment context.

This study used the 1SD approach to calculate mean BMR values based on background SD, and the results were used to support the choice of an endpoint-specific BMR of 50% for mutagenicity dose–response data. However, this should not be interpreted as a recommendation to use the 1SD approach for BMD analysis. Firstly, it should be noted that the background SD in a specific study is subject to sampling error. Thus, even if the studies could be perfectly replicated, their SD values would still differ. Moreover, the background SD for a study also includes unintended sources of variation, i.e., experimental conditions that were not perfectly homogeneous. Consequently, the background SD does not just reflect natural variation and experimental errors. Secondly, using the 1SD approach makes the extrapolation

of animal BMD values to human BMD values problematic since animals and humans do not have the same interindividual variation. Thus, analyses of human and animal data will yield different BMD values for the same value of the BMR expressed relative to the SD of any dose group (Slob 2017).

It should be acknowledged that the choice of endpoint-specific BMR could be based on different rationales (e.g., biological rationales), and/or different analytical approaches (EFSA Scientific Committee 2022; USEPA 2022; Slob 2017). The latter refers to the ES theory approach or the mean BMR based on background SD across a large number of studies. It could be argued that these approaches involve assumptions that may lead to some doubt about the calculated BMR values. However, it is important to realize that recommending BMR values using these approaches is operationally analogous to the situation associated with EFSA's default recommendation of 5% (Hardy et al. 2017), that is, although there is uncertainty associated with *var* and SD, there is no uncertainty associated with any specified BMR value. Importantly, although the ES theory approach gives a similar BMR to the 1SD approach (i.e., $BMR = \left(e^{z \cdot SD}\right)^{\alpha} = \left(e^{7 \cdot SD}\right)^{\frac{1}{8}} \approx e^{SD}$), and the approaches are mathematically aligned, the ES theory and 1SD approach are fundamentally different, as explained above.

In conclusion, the use of the Slob (2017) ES Theory and coincident *var* values yielded endpoint-specific BMR values for the TGR and *Pig-a* mutagenicity endpoints of 47% and 60%, respectively. Endpoint-specific BMR values determined using the truncated distribution of study-specific SD values for concurrent controls were 33% and 58% for the TGR and *Pig-a* endpoints, respectively. Considering the results obtained, which are based on quantitative and transparent analyses, and the in vivo genetic toxicity BMR values described in the literature (see summary in White et al. 2020), we propose uniform use of a BMR value of 50% for in vivo mutagenicity endpoints. This value can be used to derive robust BMD values (i.e., PoD values) for the interpretation of mutagenicity dose–response data in a human health risk assessment context. Nevertheless, it is important to note that analogous BMR determinations will be required for in vivo genetic toxicity endpoints that have not yet been appropriately scrutinized, for example, the comet assay and mutagenicity assays that employ ecNGS technologies. With respect to the MN assay, analyses that examined 139 datasets to determine *var* and *s* (Slob 2017) indicate a BMR of 71%–79% (White et al. 2020). Comparable analyses based on an extensive dataset of trimmed historical control values indicate a BMR of 34% to 49%. Provisional BMR estimates for the comet assay, which were based on a more limited database of trimmed historical control values, range from 39% to 75% (Zeller et al. 2017). With respect to ecNGS mutagenicity endpoints, all available dose–response data (i.e., 39 datasets) have recently been collected and collated, and the required analyses are underway (Mulugeta et al. in preparation). Lastly, additional analyses will be required to determine appropriate BMR values for in vitro genetic toxicity endpoints. Results to date for four in vitro endpoints indicate a BMR range of 20% to 61% (Beal et al. 2023).

Author Contributions

Paul White, George Johnson, Andreas Zeller, and Guangchao Chen conceived of the project and initiated the analyses. Nikolai Chepelev, Madison Bell, and Lauren Gallant collated and curated the data required for the analyses. Paul White and Guangchao Chen analyzed the data and prepared the figures and tables. Paul White and Guangchao Chen wrote the draft manuscript; all authors participated in manuscript review, revision, and finalization. All authors provided intellectual input and approved the final version of the manuscript.

Acknowledgments

The authors are grateful for helpful input from the Quantitative Analysis Workgroup (QAW) of the HESI-GTTC (Health and Environmental Sciences Institute-Genetic Toxicology Technical Committee), including HESI-GTTC managers Connie Chen and Raechel Puglisi. Bhaskar Gollapudi provided insightful comments on an early version of the manuscript. Additionally, we thank Gordon Barett, Christopher Rowat, Matthew Meier, Emily McVey, Judith de Heer, and 3 anonymous HESI reviewers for comments and suggested revisions. The work was financially supported by the Government of Canada's Chemicals Management Plan (CMP), the Dutch Ministry of Public Health, Welfare and Sport, the Dutch Ministry of Agriculture, Nature and Food Quality, and the Netherlands Food and Consumer Product Safety Authority (I/133002/02/DR). Open Access funding provided by the Health Canada library.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Allemang, A., C. Mahony, C. Lester, and S. Pfuhler. 2018. "Relative Potency of Fifteen Pyrrolizidine Alkaloids to Induce DNA Damage as Measured by Micronucleus Induction in HepaRG Human Liver Cells." *Food and Chemical Toxicology* 121: 72–81.

Bachmanov, A. A., D. R. Reed, G. K. Beauchamp, and M. G. Tordoff. 2002. "Food Intake, Water Intake, and Drinking Spout Side Preference of 28 Mouse Strains." *Behavior Genetics* 32, no. 6: 435–443.

Beal, M., M. Audebert, T. Barton-Maclaren, et al. 2022. "Quantitative In Vitro to In Vivo Extrapolation (IVIVE) of Genotoxicity Data Provides Protective Estimates of In Vivo Dose." *Environmental and Molecular Mutagenesis* 64: 105–122.

Beal, M. A., G. Chen, K. L. Dearfield, et al. 2023. "Interpretation of In Vitro Concentration-Response Data for Risk Assessment and Regulatory Decision-Making: Report From 2022 IWGT Quantitative Analysis Expert Working Group Meeting." *Environmental and Molecular Mutagenesis* 1: 22582.

Bercu, J. P., S. Zhang, Z. Sobol, P. A. Escobar, P. Van, and M. Schuler. 2023. "Comparison of the Transgenic Rodent Mutation Assay, Error Corrected Next Generation Duplex Sequencing, and the Alkaline Comet Assay to Detect Dose-Related Mutations Following Exposure to N-Nitrosodiethylamine." *Mutation Research* 891: 503685.

Chepelev, N., A. S. Long, M. Beal, et al. 2023. "Establishing a Quantitative Framework for Regulatory Interpretation of Genetic Toxicity Dose-Response Data: MOE (Margin of Exposure) Case Study of 48 Compounds With Both In Vivo Mutagenicity and Carcinogenicity Dose-Response Data." *Environmental and Molecular Mutagenesis* 64: 4–15.

Crump, K. S. 1984. "An Improved Procedure for Low-Dose Carcinogenic Risk Assessment From Animal Data." *Journal of Environmental Pathology, Toxicology and Oncology* 5: 339–348.

Douglas, G. R., C. Beevers, B. B. Gollapudi, et al. 2022. "Impact of Sampling Time on the Detection of Mutations in Rapidly Proliferating

Tissues Using Transgenic Rodent Gene Mutation Models: A Review." *Environmental and Molecular Mutagenesis* 63: 376–388.

Dragsted, L. O., B. Daneshvar, U. Vogel, et al. 2002. "A Sucrose-Rich Diet Induces Mutations in the Rat Colon." *Cancer Research* 62: 4339–4345.

EFSA (European Food Safety Authority). 2009. "Use of Benchmark Dose Approach in Risk Assessment: Guidance of the Scientific Committee." *EFSA Journal* 1150: 1–72.

EFSA (European Food Safety Authority) Scientific Committee. 2022. "Guidance on the Use of the Benchmark Dose Approach in Risk Assessment." *EFSA Journal* 20: 7584.

Gollapudi, B. B., S. Su, A. A. Li, G. E. Johnson, R. Reiss, and R. J. Albertini. 2020. "Genotoxicity as a Toxicologically Relevant Endpoint to Inform Risk Assessment: A Case Study With Ethylene Oxide." *Environmental and Molecular Mutagenesis* 61: 852–871.

Guo, X., R. H. Heflich, S. L. Dial, M. De, P. A. Richter, and N. Mei. 2018. "Quantitative Differentiation of Whole Smoke Solution-Induced Mutagenicity in the Mouse Lymphoma Assay." *Environmental and Molecular Mutagenesis* 59: 103–113.

Guo, X., R. H. Heflich, S. L. Dial, P. A. Richter, M. M. Moore, and N. Mei. 2016. "Quantitative Analysis of the Relative Mutagenicity of Five Chemical Constituents of Tobacco Smoke in the Mouse Lymphoma Assay." *Mutagenesis* 31: 287–296.

Haber, L. T., M. L. Dourson, B. C. Allen, et al. 2018. "Benchmark Dose (BMD) Modeling: Current Practice, Issues, and Challenges." *Critical Reviews in Toxicology* 48: 387–415.

Hardy, A., D. Benford, T. Halldorsson, et al. 2017. "Update: Use of the Benchmark Dose Approach in Risk Assessment." *EFSA Journal* 15: 4658.

Johnson, G. E., K. Dobo, B. Gollapudi, et al. 2021. "Permitted Daily Exposure Limits for Noteworthy N-Nitrosamines." *Environmental and Molecular Mutagenesis* 62: 293–305.

Johnson, G. E., L. G. Soeteman-Hernández, B. B. Gollapudi, et al. 2014. "Derivation of Point of Departure (PoD) Estimates in Genetic Toxicology Studies and Their Potential Applications in Risk Assessment." *Environmental and Molecular Mutagenesis* 55: 609–623.

Jones, C. P., K. L. Boyd, and J. M. Wallace. 2016. "Evaluation of Mice Undergoing Serial Oral Gavage While Awake or Anesthetized." *Journal of the American Association for Laboratory Animal Science* 55, no. 6: 805–810.

Kirkland, D., D. D. Levy, M. J. LeBaron, et al. 2019. "A Comparison of Transgenic Rodent Mutation and In Vivo Comet Assay Responses for 91 Chemicals." *Mutation Research* 839: 21–35.

Kuo, B., M. A. Beal, J. W. Wills, et al. 2022. "Comprehensive Interpretation of In Vitro Micronucleus Test Results for 292 Chemicals: From Hazard Identification to Risk Assessment Application." *Archives of Toxicology* 96: 2067–2085.

Lambert, I. B., T. M. Singer, S. E. Boucher, and G. R. Douglas. 2005. "Detailed Review of Transgenic Rodent Mutation Assays." *Mutation Research* 590: 1–280.

Long, A. S., J. W. Wills, D. Krolak, et al. 2018. "Benchmark Dose Analyses of Multiple Genetic Toxicity Endpoints Permit Robust, Cross-Tissue Comparisons of MutaMouse Responses to Orally Delivered Benzo[a]Pyrene." *Archives of Toxicology* 92: 967–982.

Lutz, W. K. 2009. "The Viracept (Nelfinavir)–ethyl Methanesulfonate Case: A Threshold Risk Assessment for Human Exposure to a Genotoxic Drug Contamination?" *Toxicology Letters* 190: 239–242.

Lynch, A. M., J. Howe, D. Hildebrand, et al. 2024. "N-Nitrosodimethylamine Investigations in Muta™Mouse Define Point-Of-Departure Values and Demonstrate Less-Than-Additive Somatic Mutant Frequency Accumulations." *Mutagenesis* 39: 96–118.

Manjanatha, M. G., A. Aidoo, S. D. Shelton, et al. 2006. "Genotoxicity of Acrylamide and Its Metabolite Glycidamide Administered in Drinking Water to Male and Female Big Blue Mice." *Environmental and Molecular Mutagenesis* 47: 6–17.

Manjanatha, M. G., L. W. Guo, S. D. Shelton, and D. R. Doerge. 2015. "Acrylamide-Induced Carcinogenicity in Mouse Lung Involves Mutagenicity: cII Gene Mutations in the Lung of Big Blue Mice Exposed to Acrylamide and Glycidamide for up to 4Weeks." *Environmental and Molecular Mutagenesis* 56: 446–456.

Marchetti, F., R. Cardoso, C. L. Chen, et al. 2023. "Error-Corrected Next-Generation Sequencing to Advance Nonclinical Genotoxicity and Carcinogenicity Testing." *Nature Reviews. Drug Discovery* 22: 165–166.

Marchetti, F., G. Zhou, D. LeBlanc, et al. 2021. "The 28 + 28 Day Design Is an Effective Sampling Time for Analyzing Mutant Frequencies in Rapidly Proliferating Tissues of MutaMouse Animals." *Archives of Toxicology* 95: 1103–1116.

Mauderly, J. L. 1986. "Respiration of F344 Rats in Nose-Only Inhalation Exposure Tubes." *Journal of Applied Toxicology* 6: 25–30.

Menz, J., M. E. Götz, U. Gündel, et al. 2023. "Genotoxicity Assessment: Opportunities, Challenges and Perspectives for Quantitative Evaluations of Dose-Response Data." *Archives of Toxicology* 97: 2303–2328.

Mittelstaedt, R. A., J. G. Shaddock, J. A. Bhalli, et al. 2021. "Differentiating Between Micronucleus Dose-Responses Induced by Whole Cigarette Smoke Solutions With Benchmark Dose Potency Ranking." *Mutation Research* 866: 503351.

Müller, L., and E. Gocke. 2009. "Considerations Regarding a Permitted Daily Exposure Calculation for Ethyl Methanesulfonate." *Toxicology Letters* 190: 330–332.

OECD (Organization for Economic Cooperation and Development). 2022a. *Transgenic Rodent Somatic and Germ Cell Mutation Assays*. OECD Guidelines for the Testing of Chemicals TG488.

OECD (Organization for Economic Cooperation and Development). 2022b. *Mammalian Erythrocyte Pig-a Gene Mutation Assay*. OECD Guidelines for the Testing of Chemicals TG470.

Powley, M. W., Z. Sobol, G. E. Johnson, et al. 2024. "N-Nitrosamine Impurity Risk Assessment in Pharmaceuticals: Utilizing in Vivo Mutation Relative Potency Comparison to Establish an Acceptable Intake for NTTP." *Regulatory Toxicology and Pharmacology* 152: 105681.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. www.R-project.org.

Rocha, J. L., E. J. Eisen, L. D. Van Vleck, and D. Pomp. 2004. "A Large-Sample QTL Study in Mice: I. Growth." *Mammalian Genome* 15: 83–99.

RStudio Team. 2023. "RStudio: Integrated Development for R. 2023.12.0 Build 369."

Shemansky, J. M., L. P. McDaniel, C. Klimas, et al. 2019. "Pig-a gene mutation database." *Environmental and Molecular Mutagenesis* 60: 759–762.

Singer, T. M. 2006. *Trangenic Rodent Gene Mutation Assays: Performance Characteristics and Exploration of the Effects of Critical Variables Affecting the Development of a Standardized Experimental Protocol*. Carleton University.

Slob, W. 2002. "Dose-Response Modeling of Continuous Endpoints." *Toxicological Sciences* 66: 298–312.

Slob, W. 2017. "A General Theory of Effect Size, and Its Consequences for Defining the Benchmark Response (BMR) for Continuous Endpoints." *Critical Reviews in Toxicology* 47: 342–351.

Slob, W., and W. Setzer. 2014. "Shape and Steepness of Toxicological Dose-Response Relationships of Continuous Endpoints." *Critical Reviews in Toxicology* 44: 270–297.

Strohl, K. P., A. J. Thomas, P. St Jean, E. H. Schlenker, R. J. Koletsky, and N. J. Schork. 1997. "Ventilation and Metabolism Among Rat Strains." *Journal of Applied Physiology (Bethesda, MD: 1985)* 82: 317–323.

Sucko, M. A., P. Danneman, and C. Brayton. 2001. *The Laboratory Mouse*. CRC Press.

Turner, P. V., C. Pekow, M. A. Vasbinder, and T. Brabb. 2011. "Administration of Substances to Laboratory Animals: Equipment Considerations, Vehicle Selection, and Solute Preparation." *Journal of the American Association for Laboratory Animal Science* 50: 614–627.

USEPA (United States Environmental Protection Agency). 1988. *Recommendations for and Documentation of Biological Values for Use in Risk Assessment*. Integrated Risk Information System (IRIS), United States Environmental Protection Agency. EPA/600/6–87.

USEPA (United States Environmental Protection Agency). 2012. *Benchmark Dose Technical Guidance*. Risk Assessment Forum, United States Environmental Protection Agency.

USEPA (United States Environmental Protection Agency). 2022. *Benchmark Dose Software. Version 3.3 Users Guide*. United States Environmental Protection Agency.

White, P. A., G. R. Douglas, D. H. Phillips, and V. M. Arlt. 2017. "Quantitative Relationships Between lacZ Mutant Frequency and DNA Adduct Frequency in MutaMouse Tissues and Cultured Cells Exposed to 3-Nitrobenzanthrone." *Mutagenesis* 32: 299–312.

White, P. A., A. S. Long, and G. E. Johnson. 2020. "Quantitative Interpretation of Genetic Toxicity Dose-Response Data for Risk Assessment and Regulatory Decision-Making: Current Status and Emerging Priorities." *Environmental and Molecular Mutagenesis* 61: 66–83.

White, P. A., A. Zeller, S. Pfuhler, and G. E. Johnson. 2019. "Re: Gi Et al. 2018, in Vivo Positive Mutagenicity of 1,4-Dioxane and Quantitative Analysis of Its Mutagenicity and Carcinogenicity in Rats." *Archives of Toxicology* 92: 3207–3221.

Wignall, J. A., A. J. Shapiro, F. A. Wright, et al. 2014. "Standardizing Benchmark Dose Calculations to Improve Science-Based Decisions in Human Health Assessments." *Environmental Health Perspectives* 122: 499–505.

Wills, J. W., G. E. Johnson, H. L. Battaion, and P. A. White. 2017. "Comparing BMD-Derived Genotoxic Potency Estimations Across Variants of the Transgenic Rodent Gene Mutation Assay." *Environmental and Molecular Mutagenesis* 58: 632–643.

Wills, J. W., G. E. Johnson, S. H. Doak, L. G. Soeteman-Hernandez, W. Slob, and P. A. White. 2016a. "Empirical Analysis of BMD Metrics in Genetic Toxicology Part I: In Vitro Analyses to Provide Robust Potency Rankings and Support MOA Determinations." *Mutagenesis* 31: 255–263.

Wills, J. W., A. S. Long, G. E. Johnson, et al. 2016b. "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: 265–275.

Zeller, A., G. Duran-Pacheco, and M. Guerard. 2017. "An Appraisal of Critical Effect Sizes for the Benchmark Dose Approach to Assess Dose-Response Relationships in Genetic Toxicology." *Archives of Toxicology* 91: 3799–3807.

Zhang, S., S. L. Coffing, W. C. Gunther, et al. 2024. "Assessing the Genotoxicity of N-Nitrosodiethylamine With Three In Vivo Endpoints in Male Big Blue® Transgenic and Wild-Type C57BL/6N Mice." *Environmental and Molecular Mutagenesis* 65: 190–202.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.