Toxicology Reports 6 (2019) 674–682

Contents lists available at ScienceDirect

Toxicology Reports

journal homepage: www.elsevier.com/locate/toxrep

Quantifying association between liver tumor incidence and early-stage liver weight increase – An NTP data analysis

Kan Shao*, Qiran Chen, Zemin Wang

Department of Environmental and Occupational Health, School of Public Health, Indiana University, Bloomington, IN, 47405, USA

ARTICLE INFO

Keywords:
Risk assessment
Liver tumor
Liver weight
Carcinogenicity
Benchmark dose
NTP

ABSTRACT

Two-year toxicology and carcinogenesis rodent studies conducted at the National Toxicology Program (NTP) are used to identify potential adverse health effects in humans due to chemical exposure, including cancer. Liver tumor, the most frequently diagnosed tumor type of chemically induced neoplastic effects documented in NTP's carcinogenicity studies, is usually difficult to be detected at early stage due to the inconspicuous symptoms. However, the abnormal growth of liver cells can lead to liver weight increase, so it is hypothesized that liver tumor incidence is associated with early stage liver weight increase. In this study, the association between liver weight increase and liver tumor incidence are quantified by (1) calculating the correlation coefficient of and (2) building quantitative linear relationship between benchmark dose estimates derived from these two types of data collected from NTP studies. Together with 151 chemical/species/sex combinations of liver tumor data showing positive evidence collected from 76 NTP long-term studies, short-term liver weight data reported in the same NTP report were extracted to be paired with the liver tumor data for the analyses. Results show that the estimated correlation coefficients (as high as 0.78) along with the adequately fitted linear models suggest that the association between relative liver weight increase and aggregated liver tumor incidence are relatively strong. Additional analyses focused on some more specific situations (e.g., specific tumor type or specific strain/sex combination) further confirmed the strong association. Given the design of this study, the interpretation of the findings is not that liver weight increase can be used to predict liver tumor incidence, instead, evident increase in liver weight might be used as a reason to prioritize the test article for a two-year toxicology and carcinogenesis study.

1. Introduction

Two-year toxicological and carcinogenesis rodent studies conducted at the National Toxicology Program (NTP) are often used to identify chemical-induced adverse health effects in humans, including choric diseases like cancer. In a traditional two-year bioassay study at NTP, both sexes of rats (e.g., F344/N or Sprague Dawley rats) and mice (e.g., B6C3F1/N hybrid mice) are exposed to a chemical at a number of dose levels (including an control group) in groups of 50 animals/sex/species to elicit toxicological responses. At the end of the study, all of the tested animals are sacrificed for a comprehensive pathological examination to identify any non-neoplastic and neoplastic lesions. Due to the large number of animals used in the experiments and time-consuming procedure, a typical two-year bioassay study can cost up to several millions of dollars.

Primary liver cancer is the sixth most commonly occurring cancer in the world and is among the leading cause of cancer deaths globally [1]. Studies have shown that a number of risk factors and comorbidities are associated with hepatic carcinogenesis, such as non-alcoholic fatty liver disease [2], and exposure to chemicals [3–5] and nanomaterials [6]. On the other hand, liver tumor represents the most frequently diagnosed tumor types of chemically induced neoplastic effects documented in NTP's carcinogenicity studies [7]. Consequently, a number of studies have investigated the feasibility and reliability to use early stage toxicological or pathological effects as an indicator to predict liver tumor incidences through the analyses of NTP data. Allen et al [8] assessed the effectiveness of correlating the presence of sub-chronic liver lesions with the occurrence of liver cancer by encompassing multiple NTP studies. The study found that the integrated consideration of hepatocellular necrosis, hepatocellular hypertrophy, and hepatocellular cytomegaly could be a good predictor of carcinogenicity in the 2-year study. Boobis et al [9], based on a retrospective analysis of sixteen chemicals

* Corresponding author at: Department of Environmental and Occupational Health, Indiana University School of Public Health, 1025 E. Seventh Street, Bloomington, IN, 47405, USA.
E-mail address: kshao@indiana.edu (K. Shao).

https://doi.org/10.1016/j.toxrep.2019.07.001
Received 19 February 2019; Received in revised form 28 June 2019; Accepted 6 July 2019
Available online 08 July 2019

with liver, lung, or kidney tumors in the NTP database, pointed out that cellular changes indicative of a tumor endpoint could be identified utilizing short-term conventional endpoints for many of the examined chemicals. Particularly, this study found that the combined consideration of the three liver lesions used in Allen et al. [8] together with increase relative liver weight successfully identified eleven of eleven liver carcinogens. Later, Ring and Eskofier [10] employed patterns recognized by data mining methods to accurately predict over 80% of the incidence of liver tumors using short-term liver weight data. However, the extracted patterns simultaneously indicated potential bias in liver tumor prediction which may depend on test agent and some other factors in study design (such as species and sex). All the studies have confirmed the feasibility to use early stage endpoints in liver to predict the incidence of liver tumor in rodents, even though unbiased prediction requires more detailed case-by-case analyses.

All the three studies focused on evaluating the correlation of short-term liver effects and long-term liver tumors by measuring the agreement between qualitative indicators (e.g., the presence of increased liver weight vs. the incidence of liver tumor), but none took the dose of test agent into account. As pointed out in Boobis et al [9], analyzing dose response for these effects will be important for improving the predictivitiy. Consequently, dose-response information is integrated into the analyses performed in this study, with an aim to investigate the correlation between the early-stage liver weight increase and liver tumor incidence in the long-term NTP rodent carcinogenicity studies in a more quantitative way. Benchmark dose (BMD) will be calculated for these two focused endpoints (i.e., liver tumor incidence and liver weight increase) and then the quantitative association between the endpoints will be assessed at two levels: (1) estimate the correlation coefficient of the dose levels that can cause critical effects (i.e., BMD) in the two focused endpoints; and (2) perform a regression analysis to predict long-term BMD based on the short-term BMD estimated from liver weight increase. The novel and advanced aspect of this method is two-fold: first, dose-response information (rather than dichotomous indicator of with or without tumor) included in the data is better utilized to identify changes in the endpoints; second, the prediction of long-term of BMD using short-term BMD is more favored by regulatory risk assessment than simply an indicator of correlation.

2. Methods

2.1. Data collection

The data of liver tumor incidence and liver weight were extracted from National Toxicological Program (NTP)'s database[1] . NTP is an interagency program that generates, shares, and also interprets toxicological information about potentially hazardous substances in the environment. Toxicological studies conducted at NTP meet all applicable health and safety requirements and are subjected to retrospective quality assurance audits before publication [11], therefore, the data published in NTP reports are of high quality. NTP provides a thorough database of study reports in two main series: (1) technical report on toxicology and carcinogenesis studies (TR series), which documents a 2-year toxicological and carcinogenesis study and reports both non-neoplastic and neoplastic effects observed in the study. Commonly, short-term studies (typically 2-week and 3-month) are conducted prior to the 2-year studies to determine the treatments to be used in the 2-year study and to identify some potential organs or systems of interest. The results of the short-term studies are usually available in the TR report. (2) Technical report on toxicity studies (TOX series), which reports non-neoplastic lesions and genetic toxicology observed in 2-week and 3-month studies that are mainly used to characterize and evaluate the toxicological potential of testing articles. In most cases, the test chemicals in the TR and TOX reports do not overlap with each other. Therefore, when searching for short-term liver weight data corresponding to the liver tumor incidence, the data reported in the same TR report are used as a preferred source. If no such data are available, then equivalent data (i.e., exposure to the same chemical through the same route) reported in the TOX report are considered.

Three steps were used to collect data for the analyses in this study, and the steps are graphically shown using a flow chart in Fig. 1. First, screening all 593 currently published TR reports using NTP's Chemical Effects in Biological Systems (CEBS)[2] . According to the results reported in the "Organ Sites with Neoplasia" section, there are 174 test articles of which at least one chemical/species/sex combination is associated with positive evidence (including clear and some evidence[3]) of liver carcinogenicity. Among these combinations, there are 50 combinations for male rats, 55 for female rats, 99 for male mice and 122 for female mice.

Second, reviewing all of these 174 test articles to find short-term (in this step, any duration shorter than or equal to 60 weeks was considered as short-term) liver weight data (including absolute liver weight, relative liver weight, or both) from its own TR. Consequently, adequate data for 71 test chemicals were collected from TRs. For the chemicals whose short-term data cannot be found in its own TR reports, we then attempted to find the corresponding liver weight data in the TOX reports. In this study, five sets of liver weight data obtained from the TOX reports were paired with long-term liver tumor incidence data. Although the corresponding studies were conducted separately, all these five TOX studies used the same treatment methods and strains of rodents as the corresponding TR studies. Eventually, 76 test chemicals including 153 chemical/species/sex combinations of long-term liver tumor incidence data were identified. The main reason why a large number of studies was screened out is that it was not a standard practice to report short-term liver weight data in the TRs published before 1990. There are a little more sets of liver weight data (261 sets in total) corresponding to these combinations because for some chemicals liver weights were measured at various time points. The identified 76 test chemicals are listed in Table 1 where the last column contains the number of chemical/species/sex combinations with positive evidence of liver tumor.

The third step is collecting and transforming the data to a consistent format for the identified combinations, and preprocessing the data to make them ready for analyses. For the liver tumor data, in addition to the dose levels and number of total subjects in the dose groups, the number of incidences of different types of neoplastic effects were recorded, such as hepatocellular adenoma, hepatocellular carcinoma, hepatoblastoma, cholangiocarcinoma, hepatocholangiocarcinoma, and hepatocholangioma, for each of the chemical/species/sex combinations if the data are available. Besides the individual tumor types, we also tried our best to record the number of animals with any of these types of tumor in each dose group. When individual liver incidence status is available in TRs, these number can be accurately counted (in majority of the cases), however, when such information was not reported, we then used the aggregated counts that cover most tumor types reported in TRs (e.g., "Hepatocellular Adenoma, or Carcinoma, or Hepatoblastoma" sometimes reported in TRs). Consequently, in our analyses presented below, four different types of endpoints were considered, i.e., adenoma, carcinoma, adenoma or carcinoma (i.e., the two most common types), and all neoplastic effects (i.e., number of animals with any types of liver tumor considered in NTP study). The count of individual and aggregated liver tumor incidence for each dose group was recorded in different columns in the "longterm_livertumor.xlsx" file as a supplemental material. Two combinations were removed because

[1] All of the NTP long-term technical reports are available at: https://ntp.niehs.nih.gov/results/pubs/longterm/reports/longterm/index.html.

[2] Available at https://tools.niehs.nih.gov/cebs3/ui/.

[3] The definition of clear evidence and some evidence is provided in each TR report.

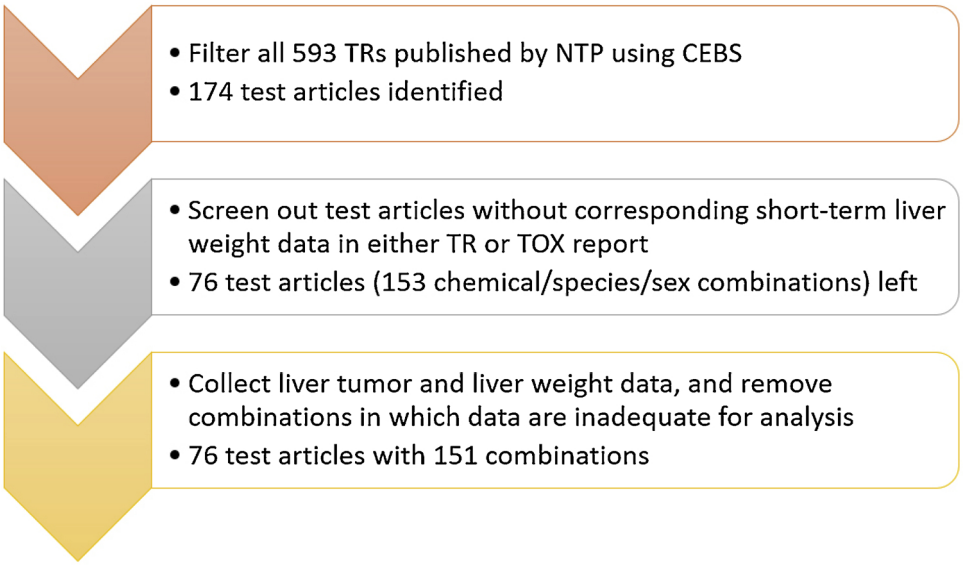

Fig. 1. The flow chart of the process of data collection.

only one exposure group (in addition to the control group) was used in the test (in TR-508). Therefore, 76 test chemicals with 151 combinations were eventually included in the final dataset for analyses. For liver weight data, both absolute and relative live weight data were collected if they are available (in some cases only one type was reported). The weight data were mostly reported in summary data as mean ± standard error. To make the datasets consistent, a few sets of data reported as standard deviation were converted to standard error by dividing the square root of the sample size. A few relative liver weight datasets expressed in percentage were converted to the unit of mg (liver weight) / g (body weight) (i.e., ‰ body weight) by multiplying 10 to make the data in the same unit (i.e., converting % body weight to ‰ body weight). Responding to the action taken in the liver tumor dataset, the two corresponding combinations were excluded in the liver weight data as well. The detailed liver weight data used in this study are stored in “shortterm_liverweight.xlsx” available in the supplemental material.

2.2. Benchmark dose methodology

The benchmark dose (BMD) methodology originally proposed by Crump in 1984 [19] is adopted in the study to estimate a toxicity value (i.e., BMD) that induces a predetermined change in the response. The BMD methodology has a few important advantages over the traditional pair-wise comparison approach to determine the dose that causes chemical-related adverse effect (i.e., the NOAEL/LOAEL method). For the purposes of this study, the key features that make the BMD method preferable are its ability (1) to take data at all dose levels into account (so that background response and potency of response can be considered), (2) to sensitively detect the dose-response trend in the data, and (3) to compare toxicity values among different studies and data types. Because liver tumor incidence and liver weight are two different types of endpoint data, i.e., dichotomous data and continuous data respectively, the BMD method for these two data types is briefly introduced below.

2.2.1. BMD analysis for dichotomous data

For dichotomous data (like the liver tumor incidence data indicating whether liver tumor is present or not), a BMD based on the extra risk definition (US EPA’s default option) is defined by the equation below:

$$\frac{f(BMD) - f(0)}{1 - f(BMD)} = BMR \quad (1)$$

where $f(\bullet)$ represents an empirical dose-response model commonly used in risk assessment. For dichotomous data, a benchmark response (i.e., BMR) value is chosen to indicate a toxicologically meaningful adverse change, typically on the lower end of response (such as 1% or 10%). Therefore, the estimated benchmark dose is the dose level that satisfies the function above.

In risk assessment practice, a number of empirical models are fit to dose-response data and then a BMD estimated from the most adequate model [12] or a model averaged BMD estimated from a suite of models [13] is chosen as the starting point for low dose extrapolation. In this study, the recently developed model averaging BMD estimation approach [14] has been applied to take model uncertainty into account for BMD estimation: eight dose-response models, including the Quantal-linear model, Logistic model, Probit model, 2nd degree multistage model, Weibull model, Loglogistic model, Logprobit model, and Dichotomous Hill model, are fit to the data individually, then a weight (calculated based on how well the model fits the data) averaged BMD is estimated by integrating BMD estimates from each individual model. Please refer to Shao and Shapiro [14] for detailed description of the format of the models, as well as the model averaging methodology. In the present study, for dichotomous data, BMDs are estimated based on two BMR values, i.e., 1% and 10% extra risk, for the reason that the response in the low dose range is our main focus.

2.2.2. BMD analysis for continuous data

For continuous data (such as the liver weight data that are on a continuous scale), the BMD is defined based on the relative change of central tendency as expressed by the Eq. (2) below:

$$\frac{f(BMD) \pm f(0)}{f(0)} = BMR \quad (2)$$

where $f(\bullet)$ is a continuous dose-response model to represent the central tendency of response. Unlike the dichotomous data, continuous data can either increase or decrease to be considered as adverse, consequently there is a “±” sign in the numerator in Eq. (2). Given the relative change definition, the BMR is commonly set at a small percentage (e.g., 1% or 10%) to define adverse effect in the response in low dose

Table 1
Chemicals Selected for This Analysis.

TR Report Year	Liver Tumor Data Source	Liver Weight Data Source	Test Article	CAS No.	Exposure Route	Count of Combinations
1978	TR-027	TOX-49	1,1,2,2-Tetrachloroethane	79-34-5	Gavage	2
1983	TR-244	TR-244	Polybrominated biphenyl mixture	67774-32-7	Gavage	4
1986	TR-308	TR-308	Chlorinated paraffins: C12, 60% chlorine	108171-26-2	Gavage	4
1989	TR-349	TR-349	Pentachlorophenol	87-86-5	Feed	4[c]
1989	TR-351	TR-351	p-Chloroaniline hydrochloride	20265-96-7	Gavage	2
1989	TR-352	TR-352	N-Methylolacrylamide	924-42-5	Gavage	2
1990	TR-382	TR-382	Furfural	98-01-1	Gavage	2
1991	TR-390	TR-390	3,3'-Dimethylbenzidine dihydrochloride	612-82-8	Drinking water	2
1991	TR-395	TR-395	Probenecid	57-66-9	Gavage	1
1991	TR-405	TR-405	C.I. Acid red 114	6459-94-5	Drinking water	2
1992	TR-397	TR-397	C.I. Direct blue 15	2429-74-5	Drinking water	2
1992	TR-407	TR-407	C.I. Pigment red 3	2425-85-6	Feed	2
1993	TR-384	TR-384	1,2,3-Trichloropropane	96-18-4	Gavage	2
1993	TR-400	TR-400	2,3-Dibromo-1-propanol	96-13-9	Dermal	3
1993	TR-402	TR-402	Furan	110-00-9	Gavage	4
1993	TR-416	TR-416	o-Nitroanisole	91-23-6	Feed	2
1993	TR-420	TR-420	Triamterene	396-01-0	Feed	2
1993	TR-422	TR-422	Coumarin	91-64-5	Gavage	1
1993	TR-423	TR-423	3,4-Dihydrocoumarin	119-84-6	Gavage	2
1993	TR-434	TR-434	1,3-Butadiene	106-99-0	Inhalation	2
1993	TR-443	TR-443	Oxazepam	604-75-1	Feed	4
1994	TR-430	TR-430	C.I. Direct blue 218	28407-37-6	Feed	2
1995	TR-439	TR-439	Methylphenidate hydrochloride	298-59-9	Feed	2
1996	TR-383	TR-383	1-Amino-2,4-dibromoanthraquinone	81-49-2	Feed	4
1997	TR-450	TR-450	Tetrafluoroethylene	116-14-3	Inhalation	4
1997	TR-457	TR-457	Salicylazosulfapyridine	599-79-1	Gavage	2
1997	TR-461	TR-461	Nitromethane	75-52-5	Inhalation	1
1997	TR-463	TR-463	D & C yellow no. 11	8003-22-3	Feed	2
1998	TR-467	TR-467	Chloroprene	126-99-8	Inhalation	1
1998	TR-475	TR-475	Tetrahydrofuran	109-99-9	Inhalation	1
1999	TR-466	TOX-10	Ethylbenzene	100-41-4	Inhalation	1
1999	TR-478	TOX-20	Diethanolamine	111-42-2	Dermal	2
1999	TR-480	TR-480	Lauric acid diethanolamine condensate	120-40-1	Dermal	1
1999	TR-485	TR-485	Oxymetholone	434-07-1	Gavage	1
2000	TR-470	TR-470	Pyridine	110-86-1	Drinking Water	2
2000	TR-476	TR-476	Primidone	125-33-7	Feed	2
2000	TR-479	TR-479	Coconut oil acid diethanolamine condensate	68603-42-9	Dermal	2
2000	TR-491	TR-491	Methyleugenol	93-15-2	Gavage	4
2001	TR-496	TR-496	Fumonisin B1	116355-83-0	Feed	1
2001	TR-499	TR-499	Indium phosphide	22398-80-7	Inhalation	2
2003	TR-503	TR-503	Chloral hydrate	302-17-0	Gavage	2
2003	TR-508	TOX-27	Riddelliine	23246-96-0	Gavage	4[d]
2004	TR-510	TR-510	Urethane	51-79-6	Drinking Water	6[c]
2004	TR-512	TR-512	Elmiron	37319-17-8	Gavage	2
2004	TR-515	TR-515	Propylene glycol mono-t-butyl ether	57018-52-7	Inhalation	2
2004	TR-516	TR-516	2-Methylimidazole	693-98-1	Feed	2
2005	TR-494	TR-494	Anthraquinone	84-65-1	Feed	3
2005	TR-527	TR-527	Leucomalachite Green	129-73-7	Feed	1
2006	TR-520	TR-520	PCB126	57465-28-8	Gavage	1
2006	TR-521	TR-521	TCDD	1746-01-6	Gavage	1
2006	TR-525	TR-525	Pentachlorodibenzofuran	57117-31-4	Gavage	1
2006	TR-526	TR-526	Mixture of TCDD, PeCDF, PCB126	1746-01-6 57117-31-4 57465-28-8[a]	Gavage	1
2006	TR-530	TR-530	Binary Mixture of PCB 126, PCB 153	57465-28-8 35065-27-1[a]	Gavage	1
2006	TR-531	TR-531	Binary Mixture of PCB 126, PCB 118	57465-28-8 31508-00-6[a]	Gavage	1
2006	TR-533	TOX-61	Benzophenone	119-61-9	Feed	1
2007	TR-537	TR-537	Dibromoacetic acid	631-64-1	Drinking Water	2
2007	TR-543	TR-543	alpha-Methylstyrene	98-83-9	Inhalation	1
2008	TR-541	TR-541	Formamide	75-12-7	Gavage	1
2009	TR-542	TR-542	Cumene	98-82-8	Inhalation	1
2009	TR-549	TR-549	Bromochloroacetic acid	5589-96-8	Drinking Water	2
2010	TR-551	TR-551	Isoeugenol	97-54-1	Gavage	1
2010	TR-554	TR-554	5-(Hydroxymethyl)-2-furfural	67-47-0	Gavage	1
2010	TR-557	TR-557	beta-Myrcene	123-35-3	Gavage	1
2010	TR-558	TR-558	3,3',4,4'-Tetrachloroazobenzene	14047-09-7	Gavage	2
2010	TR-559	TR-559	PCB 118	31508-00-6	Gavage	1
2010	TR-560	TR-560	Androstenedione	63-05-8	Gavage	2
2010	TR-562	TR-562	Goldenseal root powder	GOLDENSEALRT[b]	Feed	3

(*continued on next page*)

Table 1 (*continued*)

TR Report Year	Liver Tumor Data Source	Liver Weight Data Source	Test Article	CAS No.	Exposure Route	Count of Combinations
2011	TR-561	TR-561	Tetralin	119-64-2	Inhalation	1
2011	TR-563	TR-563	Pulegone	89-82-7	Gavage	2
2012	TR-571	TR-571	Kava kava extract	9000-38-8	Gavage	2
2012	TR-575	TR-575	Acrylamide	79-06-1	Drinking Water	1
2012	TR-579	TR-579	N,N-Dimethyl-p-toluidine	99-97-8	Gavage	4
2013	TR-578	TR-578	Ginkgo biloba extract	90045-36-6	Gavage	2
2014	TR-580	TR-580	beta-Picoline	108-99-6	Drinking Water	1
2014	TR-587	TR-587	Tetrabromobisphenol A	79-94-7	Gavage	1
2016	TR-589	TR-589	Pentabromodiphenyl Ether Mixture	32534-81-9	Gavage	4

a. The testing articles are mixtures of multiple chemicals.
b. A CAS number was not assigned for Goldenseal Root Powder, so we labeled it “GOLDENSEALRT”.
c. Three different levels of Ethanol were mixed in the test article.
d. Two combinations in this test article were removed later due to the lack of dose groups.

range. In other words, the BMD is the dose where the central tendency of the response has changed a certain percent from its counterpart in the control group [20]. 1% and 10% relative change are used as BMR in this study. Model averaged BMD estimation techniques for continuous data introduced in Shao and Shapiro [14] with the default settings are applied as well to take model uncertainty into account.

2.2.3. Individual model BMD estimation

In addition to the model averaged BMD estimates, we also investigate the situation where the difference in the dose-response models from which the BMDs are estimated can be mitigated as much as possible. Accordingly, only the Quantal-linear model, i.e., $f(d) = a + (1 - a) \times [1 - \exp(-b \times d)]$, for BMD estimation from dichotomous data, and the Linear model, i.e., $f(d) = a + b \times d$, for BMD estimation from continuous data were used. In both equations, d represents dose as the independent variable. There are two main reasons for this choice: (1) both models are one of the simplest models in the corresponding category (each has only two parameters), so that they have the adaptability to fit the data sets collected from the NTP studies, including those with only three dose levels; and (2) these two models have a similar shape (i.e., linear) in the low dose region, so that the discrepancy in BMD estimation and further the disturbance in correlation of BMD estimates caused by the formats of dichotomous and continuous dose-response models can be minimized.

2.3. Quantitative measure the association

After the BMDs are estimated separately from the liver weight data and liver tumor incidence data, the correlation and dependence of the two sets of BMD estimates at two levels were quantified: (1) calculating the Pearson correlation coefficient to measure how closely the BMDs are associated; (2) applying linear regression to quantitatively describe the relationship between the two sets of BMDs with the goal of predicting long-term liver tumor BMD using short-term liver weight BMD. Such a procedure will help us establish the qualitative correlation relationship between early stage liver weight increase and liver tumor incidence, and potentially develop the quantitative connection between short-term and long-term BMDs which may have significant impact on reducing the burden of animal use for risk assessment of chronic effect. The limitation of the analyses is discussed in Section 4.

For both the calculation of correlation coefficient and linear regression, the BMDs are first converted to log scale. The doses administered in short-term studies are commonly used to determine the maximum tolerant dose to be used in long-term study, so the doses in these two types of toxicological experiments often have differences larger than one order of magnitude. Consequently, log-scale conversion, a widely used operation in practice, is employed to mitigate the difference and to rationalize the analysis of the correlation of BMDs [15,16]. The Pearson correlation coefficient is calculated as $\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$, where $COV(X, Y)$ is the covariance of the two vectors of BMD estimates, and σ_X and σ_Y are the standard deviation of the first and second vector of BMD estimates respectively. The correlation coefficient is a number between -1 and 1. The correlation coefficient of 1, 0, and -1 respectively means two quantities are positively correlated with each other perfectly, two quantities are completely random, and two quantities are negatively correlated with each other perfectly. For linear regression, the linear model, $f(x) = a + bx$ is fit to the two sets of log-transformed BMDs where the short-term BMDs estimated from liver weight data are considered as the independent variable and the long-term BMDs estimated from the liver tumor data are considered as the dependent variable. The coefficient of determination (also known as R-square) is calculated to evaluate how well the linear model fit the data. The R-square is a value between 0 and 1, with high values generally suggesting a good fit and prediction capability. It is important to point out that the linear model here has the same mathematical format as the Linear dose-response model mentioned in Section 2.2.3, but the independent and dependent variables in these two scenarios are different.

3. Data analyses and results

The analysis results are presented in two subsections: (1) general results derived from a relatively comprehensive set of available data; and (2) a number of subsets of the data to illustrate the association in those specific cases. The BMD analyses were conducted using the recently developed Bayesian benchmark dose estimation system, i.e., the BBMD system [14].

3.1. General comparison

For the general comparison, all of the available liver tumor data (i.e., the aggregated liver tumor counts represented in the “Neoplastic_effects” column) were one-to-one paired with the corresponding relative liver weight data respectively through the following three common steps:

(1) Remove liver weight data measured at extra time points. The liver weight data measured at 13-week time point or closest to 13-week were kept, and extras were removed. If the only available liver weight data were from a study with shorter or longer durations (e.g., 4-week or 52-week), then the liver weight data were still kept. We didn’t exclude the chemicals that has liver weight measures at

Table 2
Correlation Coefficients of log-BMDs from Liver Weight & Liver Tumor Incidence.

	MA Continuous BMDs vs MA Dichotomous BMDs	Linear BMDs vs Quantal Linear BMDs
BMR = 10%	0.740	0.784
BMR = 1%	0.660	0.784

multiple time points, but excluded the extra measures that are far away from the 13-week time point. The reason to choose 13-week (or 3-month) is that this is the most commonly used study design to investigate sub-chronic effects.

(2) Remove the combinations with inadequate liver weight data. There are a few studies in which only mean value of liver weight was reported (no standard deviation or standard error data) resulting incomplete data for BMD modeling. Consequently, there are 144 combinations in the liver tumor paired with relative liver weight. Given both toxicological and statistical considerations, we excluded absolute liver weight from the correlation analyses and only focused on relative live weight.

(3) For correlation coefficient estimation and linear regression, the BMDs estimated from liver tumor and weight data given BMR = 10% were paired, and the BMDs estimates based on BMR = 1% were paired.

To ensure the BMD estimates are reliable to be used in the analyses performed in this study, the posterior predictive p-value [17], a goodness-of-fit indicator, is used to evaluate how well the dose-response models fit the data. The posterior predictive p-values reported by the BBMD system suggest that the Linear model for continuous data, the Quantal-linear model for dichotomous data, and the model averaging method for both types of data can produce valid BMD estimates given adequate model fitting. The correlation coefficient estimates for the BMDs from liver weight and liver tumor data are listed in Table 2, and the corresponding linear regression results are graphically shown in Figs. 2 to 5. The correlation coefficients listed in Table 2 as well as the fitted linear curves indicate that the association between the two sets of BMD estimates from the liver weight data and liver tumor data are relatively strong.

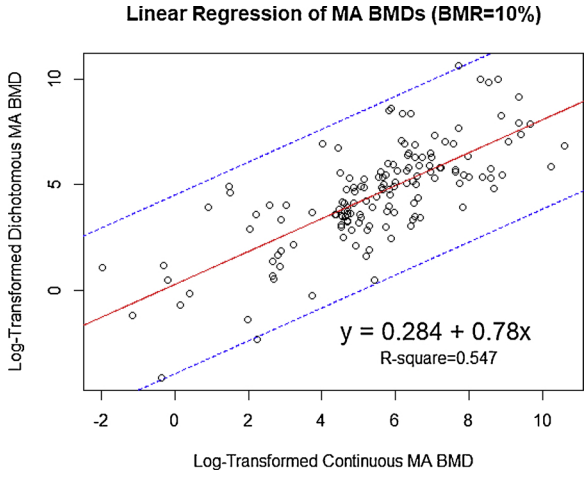

Fig. 2. Fitted linear model to the BMDs estimated using model averaging method given BMR = 10%. X-axis represents the log-transformed BMD estimated from the continuous liver weight data, and y-axis represents the log-transformed BMD estimated from the dichotomous liver tumor data. The red line in the graph represents the maximum likelihood estimated linear model with the 95th confidence interval represented by the two blue dashed lines. The equation of the fitted linear model and estimated R^2 are shown on the lower right corner.

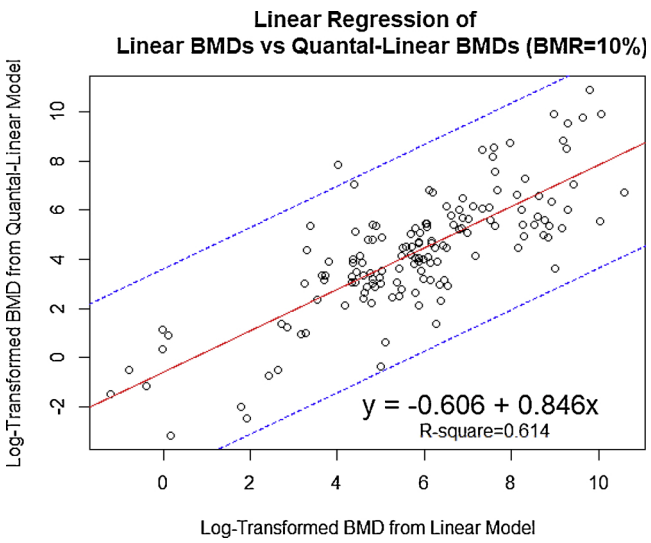

Fig. 3. Fitted linear model to the BMDs estimated using the Linear and Quantal-linear model for continuous data and dichotomous data respectively, given BMR = 10%. X-axis represents the log-transformed BMD estimated from the continuous liver weight data, and y-axis represents the log-transformed BMD estimated from the dichotomous liver tumor data. The red line in the graph represents the maximum likelihood estimated linear model with the 95th confidence interval represented by the two blue dashed lines. The equation of the fitted linear model and estimated R^2 are shown on the lower right corner.

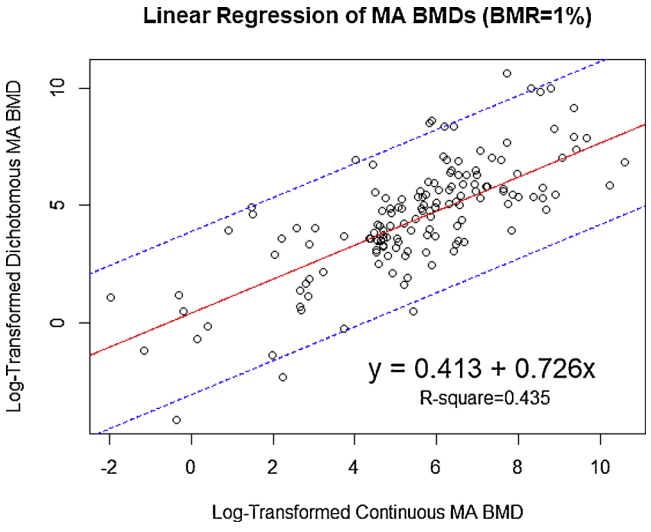

Fig. 4. Fitted linear model to the BMDs estimated using model averaging method given BMR = 1%. X-axis represents the log-transformed BMD estimated from the continuous liver weight data, and y-axis represents the log-transformed BMD estimated from the dichotomous liver tumor data. The red line in the graph represents the maximum likelihood estimated linear model with the 95th confidence interval represented by the two blue dashed lines. The equation of the fitted linear model and estimated R^2 are shown on the lower right corner.

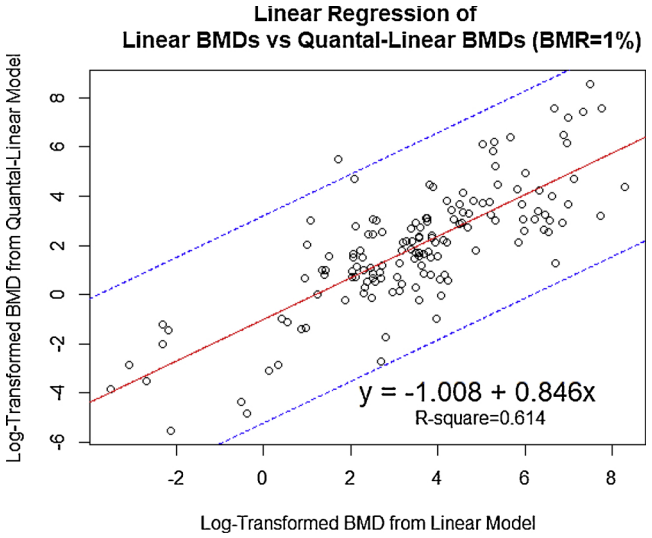

Fig. 5. Fitted linear model to the BMDs estimated using the Linear and Quantal-linear model for continuous data and dichotomous data respectively, given BMR = 10%. X-axis represents the log-transformed BMD estimated from the continuous liver weight data, and y-axis represents the log-transformed BMD estimated from the dichotomous liver tumor data. The red line in the graph represents the maximum likelihood estimated linear model with the 95th confidence interval represented by the two blue dashed lines. The equation of the fitted linear model and estimated R^2 are shown on the lower right corner.

3.2. Specific comparison

A few specific situations were explored to investigate the liver weight / tumor association in some subgroups of interest: (1) a) association between hepatocellular adenoma incidence and liver weight increase, b) between hepatocellular carcinoma incidence and liver weight increase, and c) between hepatocellular adenoma or carcinoma and liver weight increase; (2) association in subgroups male rats, female rats, male mice and female mice. The same 3-step procedure was employed to preprocess the BMDs from various subsets of liver tumor data for the correlation coefficient calculation. The results of tumor type specific correlation coefficient are shown in Table 3 below.

For the species/sex specific correlation coefficient calculation, we basically use the BMDs calculation from the relative liver weight and liver tumor incidence data (the same as the BMDs used in Table 2) and separate them into four subgroups for calculating the correlation coefficients. The results are presented in Table 4 below.

In addition to the correlation coefficients, linear regression has also been conducted for these specific cases mentioned above, and the parameter estimates of the linear model and R^2 estimates are listed in Table A1 in the Appendix. The results shown in the Tables 3, 4 and A1 demonstrate that in some cases (e.g., male mice) the association might be a little stronger, but the general pattern of associations illustrated in these specific cases is very similar to the overall situation discussed in Section 3.1.

Table 4
Correlation Coefficients of BMDs from Relative Liver Weight & Liver Tumor Incidence in four species/sex groups.

	MA Continuous BMDs vs MA Dichotomous BMDs	Linear BMDs vs Quantal Linear BMDs
Male Rats (15)	0.775 / 0.807	0.835 / 0.835
Female Rats (27)	0.766 / 0.758	0.798 / 0.798
Male Mice (49)	0.842 / 0.780	0.830 / 0.830
Female Mice (53)	0.776 / 0.705	0.808 / 0.808

Note: the numbers in the first column are the numbers of combinations used to calculate the correlation coefficient; two correlation coefficients given BMR = 10% (left) and BMR = 1% (right) are listed in each cell.

4. Discussion

This study innovatively applied the benchmark dose methodology to extract valuable information from short-term liver weight data and long-term liver tumor data: the BMD estimated based on fitted dose-response curve contains toxicological information (e.g., background response and potency) and is more sensitive to detect biological changes induced by chemical exposure. Given the quantification of the toxicity value of the short-term endpoint (i.e., liver weight increase) and long-term endpoint (i.e., liver tumor incidence), we are able to not only assess the concordance between the sub-chronic effect and the chronic effect in a relatively qualitative manner like what many previous studies did [8,10,18], but also build the quantitative relationship of the two sets of BMD estimates as a first step to predicting long-term toxicity value using short-term counterpart in support of human health risk assessment. The estimated correlation coefficients ranging from 0.627 to 0.842 together with the linear regression results confirm that the quantitative association between liver weight increase and liver tumor incidence is relatively strong, and also suggest that the estimated association indicators depend on a number of factors.

Model uncertainty is an important factor that can substantially influence the correlation estimation. The results listed in Tables 2 to 4 show that the correlation coefficients estimated from model averaged BMDs are consistently lower than the counterparts estimated using BMDs from the Linear and Quantal-linear model. The main reason is that the model averaged BMD estimation takes model uncertainty into account by integrating individual BMDs estimated from different dose-response models, consequently, the heterogeneity becomes larger due to the uncertainty and variability in the model-dependent BMD estimates. In addition, the results show that the correlation coefficients are generally lower at the BMR = 1% level than at the BMR = 10% level. This situation is mainly caused by the fact that the dose-response

Table 3
Correlation Coefficients of BMDs from Relative Liver Weight & Adenoma/Carcinoma Incidence.

	MA Continuous BMDs vs MA Dichotomous BMDs	Linear BMDs vs Quantal Linear BMDs
hepatocellular adenoma (130)	0.726 / 0.644	0.766 / 0.766
hepatocellular carcinoma (129)	0.767 / 0.674	0.740 / 0.740
Hepatocellular adenoma or carcinoma (144)	0.725 / 0.627	0.759 / 0.759

Note: the numbers in the first column are the numbers of combinations used to calculate the correlation coefficient; two correlation coefficients given BMR = 10% (left) and BMR = 1% (right) are listed in each cell.

relationship is getting more and more uncertain as the dose level getting lower. On the other hand, BMDs estimated from the Linear model and Quantal-linear model are more closely correlated when model uncertainty is excluded from the analysis. It is worth mentioning that, because of the linear shape of the Linear model and Quantal-linear model in the low-dose region, the estimated correlation coefficients are the same at the BMR levels of 10% and 1%.

The graphical results of linear regression demonstrated in Figs. 2–5 together with the estimated R^2 statistic suggest that a linear model can generally effectively describe the quantitative relationship between log-transformed BMD estimates from the liver weight and liver tumor incidence data: most of the dots are within the 95th confidence interval and the R^2 is within the range from 0.435 to 0.614. Similar to the reason for the relatively low correlation coefficient mentioned in the previous paragraph, model uncertainty also disturbs the performance of linear regression. Consequently, the R^2 estimates are a little smaller for the situations where model averaged BMD estimates were used. The fitted linear model also allows us to approximate the chronic BMD using a short-term BMD estimate. For example, for the data sets analyzed in the present study, the sub-chronic BMD is about 2.69 (with a 90th percentile interval from 0.51 to 5.69) times higher than the liver tumor BMD. It is still too premature to use the developed linear model predict chronic toxicity value using a short-term toxicity value, but the analyses and practice performed in this study can be a value first step towards that direction.

The correlation coefficients presented in Table 3 are relatively close, within the range from 0.644 to 0.767, indicating a relatively strong association between the early stage liver weight increase and three common categories of liver neoplastic effects used in risk assessment, hepatocellular adenoma, hepatocellular carcinoma, and hepatocellular adenoma or carcinoma. The hepatocellular adenoma illustrated a little stronger correlation with liver weight increase than hepatocellular carcinoma. One possible explanation is that adenomas usually appear earlier than carcinomas, which is consistent with the general belief that hepatic tumors in rodents progress from adenomas to carcinomas. Accordingly, the correlation coefficients estimated from adenoma or carcinoma are generally between the counterparts estimated from adenoma and carcinoma individually. All the three sets of correlation coefficients are quite close to the results presented in Table 2, which is consistent with the fact that in most of the combinations investigated, the aggregated tumor incidences are mostly composed of adenoma and carcinoma.

Results in Table 4 shown that male rats and male mice (especially male rats) might be a more reliable predictive model. In other words, the toxicological changes in short-term studies might be more consistent with the changes in long-term studies in male rats and male mice. This finding can have potential implications for future toxicity study design and the interpretation of short-term study findings: for example, to reduce the cost, male rats or male mice can be chosen as the default animal model to be tested first; and the liver weight increase observed in short-term studies in those rodents can be given a little more attention and serve as an indicator for a lifetime carcinogenicity study. Among these four species/sex combinations, a slightly weaker association has been observed in female rats, which might be mainly due to the facts that multiple strains of female rats have been used in recent NTP two-year carcinogenicity studies (mainly after 2005). Due to the limited number of studies using rats as the animal model (15 and 27 combinations for male and female rats respectively) included in the analysis, additional data analyses are needed to confirm the above findings.

It is also very important to understand the limitations of this study. The analyses in this study started with the collection of NTP studies with positive evidence of liver tumor, and followed by matching early stage liver weight data with these tumor incidences. Therefore, the positive correlation coefficients calculated and linear regression performed in this study are based on one important prerequisite: liver tumors exist. However, there are a few situations in NTP study where showed liver weight increase leads to no clear evidence of liver tumor, which may result in bias or false positives in short-term to long-term prediction as pointed out in a few previous studies [9,10]. Therefore, instead of asserting that liver weight increase can be an early stage indicator for liver tumor development, the study emphasizes that evident increase in liver weight might be used as a reason to prioritize the test article for a two-year toxicology and carcinogenesis study. Similarly, the linear equations illustrated in Figs. 2 to 5 should not be used as a model for quantitative prediction of chronic toxicity values, but for a rough approximation when chronic liver tumor data are missing or very limited for dose-response assessment. Our next step is to expand the database to investigate the association between early liver weight increase and no clear evidence of liver tumor formation, and the association between non-evident liver weight change and liver tumor development. Once the next step is finished, it will be more appropriate and plausible to use liver weight change as a predictor for liver tumor incidence. It is worth mentioning that the utility of the results is also limited by some other factors, such as the relatively small number of chemicals/combinations available for the data analyses (especially for the species-sex specific categories).

5. Conclusion

This study investigated the quantitative association between liver tumor incidence and early stage liver weight increase by innovatively calculating the correlation coefficient of and establishing linear model for BMDs estimated from toxicological data collected from NTP studies. The results suggest that early-stage liver weight increase has evident association with liver tumor formation, but this association can be affected by a number of factors such as the dose-response model used for the analysis and the specific tumor type of interest. Given the limitations in the design of this study, results suggest that early-stage liver weight increase can be considered as a risk factor for liver tumor development in rodents and as a reason to prioritize the research resources for two-year toxicology and carcinogenesis study, but not ready to be used as a qualitative or quantitative predictor for liver tumor.

Acknowledgement

Research reported in this publication was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number R41TR002567. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A

Table A1
Parameter and R^2 Estimates in Linear Regression Analysis for Specific Cases.

Specific Case	BMR	BMDs	a	b	R^2
hepatocellular adenoma (130)	10%	MA vs MA	0.668	0.762	0.528
		Lin vs QL	0.22	0.782	0.586
	1%	MA vs MA	0.788	0.712	0.415
		Lin vs QL	−0.33	0.782	0.586
hepatocellular carcinoma (129)	10%	MA vs MA	1.154	0.782	0.588
		Lin vs QL	0.679	0.833	0.548
	1%	MA vs MA	1.469	0.74	0.455
		Lin vs QL	0.248	0.833	0.548
Hepatocellular adenoma or carcinoma (144)	10%	MA vs MA	0.628	0.735	0.525
		Lin vs QL	−0.191	0.796	0.577
	1%	MA vs MA	0.714	0.669	0.393
		Lin vs QL	−0.708	0.796	0.577
Male Rats (15)	10%	MA vs MA	1.559	0.785	0.601
		Lin vs QL	−0.384	0.963	0.698
	1%	MA vs MA	1.845	0.834	0.652
		Lin vs QL	−0.516	0.963	0.698
Female Rats (27)	10%	MA vs MA	0.747	0.889	0.587
		Lin vs QL	−0.263	0.958	0.638
	1%	MA vs MA	1.341	0.915	0.575
		Lin vs QL	−0.407	0.958	0.638
Male Mice (49)	10%	MA vs MA	−1.089	0.913	0.71
		Lin vs QL	−0.994	0.824	0.689
	1%	MA vs MA	−1.114	0.886	0.609
		Lin vs QL	−1.448	0.824	0.689
Female Mice (53)	10%	MA vs MA	0.157	0.754	0.603
		Lin vs QL	−0.418	0.772	0.654
	1%	MA vs MA	0.183	0.696	0.496
		Lin vs QL	−0.99	0.772	0.654

Note: The seven specific cases in this table are corresponding to the cases shown in Tables 3 and IV. The abbreviations in the column named “BMDs” represent how the BMDs used in linear regression were estimated: “MA vs MA” means that both short-term and long-term BMDs were estimated using model averaging method, and “Lin vs QL” means that the short-term and long-term BMDs were estimated from the Linear model and Quantal-linear model respectively. “a” and “b” are the estimated intercept and slope parameter in the linear model, and the coefficient of determination estimates, R^2, are listed in the last column.

References

[1] L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, Global cancer statistics, 2012, CA Cancer J. Clin. 65 (2) (2015) 87–108.

[2] R. Cioboată, A. Găman, D. Traşcă, A. Ungureanu, A.O. Docea, P. Tomescu, F. Gherghina, A.L. Arsene, A.M. Tsatsakis, D. Spandidos, N. Drakoulis, D. Călina, Pharmacological management of non-alcoholic fatty liver disease: atorvastatin versus pentoxifylline, Exp. Ther. Med. 13 (2017) 2375–2381.

[3] J.K. Dunnick, A.R. Pandiri, B.A. Merrick, G.E. Kissling, H. Cunny, E. Mutlu, S. Waidyanatha, R. Sills, H.L. Hong, T.V. Ton, T. Maynor, L. Recio, S.L. Phillips, M.J. Devito, A. Brix, Carcinogenic activity of pentabrominated diphenyl ether mixture (DE-71) in rats and mice, Toxicol. Rep. 5 (2018) 615–624.

[4] L.N. Ramalho, L.D. Porta, R.E. Rosim, T. Petta, M.J. Augusto, D.M. Silva, F.S. Ramalho, C.A. Oliveira, Aflatoxin B1 residues in human livers and their relationship with markers of hepatic carcinogenesis in São Paulo, Brazil. Toxicology Reports 5 (2018) 777–784.

[5] E. Vakonaki, V.P. Androutsopoulos, J. Liesivuori, A.M. Tsatsakis, D.A. Spandidos, Pesticides and oncogenic modulation, Toxicology 307 (2013) 42–45.

[6] Z. Piperigkou, K. Karamanou, A.B. Engin, C. Gialeli, A.O. Docea, D.H. Vynios, M.S. Pavão, K.S. Golokhvast, M.I. Shtilman, A. Argiris, E. Shishatskaya, A.M. Tsatsakis, Emerging aspects of nanotoxicology in health and disease: from agriculture and food sector to cancer therapeutics, Food Chem. Toxicol. 91 (2016) 42–57.

[7] NTP (National Toxicology Program). 2018. https://manticore.niehs.nih.gov/organsites/. Retrieved on December 29, 2018.

[8] D.G. Allen, G. Pearse, J.K. Haseman, R.R. Maronpot, Prediction of rodent carcinogenesis: an evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogenicity studies, Toxicol. Pathol. 32 (4) (2004) 393–401.

[9] A.R. Boobis, S.M. Cohen, N.G. Doerrer, S.M. Galloway, P.J. Haley, G.C. Hard, F.G. Hess, J.S. MacDonald, S. Thibault, D.C. Wolf, J. Wright, A data-based assessment of alternative strategies for identification of potential human cancer hazards, Toxicol. Pathol. 37 (2009) 714–732.

[10] M. Ring, B.M. Eskofier, Data mining in the U.S. National toxicology program (NTP) database reveals a potential Bias Regarding liver tumors in rodents irrespective of the test agent, PLoS One 10 (2) (2015) e0116488.

[11] NTP (National Toxicology Program). 2017. https://ntp.niehs.nih.gov/results/pubs/longterm/reports/longterm/index.html. Retrieved on October 18, 2017.

[12] US EPA (US Environmental Protection Agency), Benchmark Dose Technical Guidance Document. EPA/100/R-12/001, U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC, 2012.

[13] EFSA Scientific Committee, A. Hardy, D. Benford, T. Halldorsson, M.J. Jeger, K.H. Knutsen, S. More, A. Mortensen, H. Naegeli, H. Noteborn, C. Ockleford, A. Ricci, G. Rychen, V. Silano, R. Solecki, D. Turck, M. Aerts, L. Bodin, A. Davis, L. Edler, U. Gundert-Remy, S. Sand, W. Slob, B. Bottex, J.C. Abrahantes, D.C. Marques, G. Kass, J.R. Schlatter, 2017. Update: Guidance on the use of the benchmark dose approach in risk assessment, Efsa J. 15 (1) (2017) 4658 41 pp.

[14] K. Shao, A. Shapiro, A web based system for bayesian benchmark dose estimation, Environ. Health Perspect. 26 (1) (2018) 017002.

[15] D. Krewski, D.W. Gaylor, A.P. Soms, M. Szyszkowicz, An overview of the report: correlation between carcinogenic potency and the maximum tolerated dose: implications for risk assessment, Risk Anal. 13 (4) (1993) 383–398.

[16] R.S. Thomas, S.C. Wesselkamper, N.C.Y. Wang, Q.J. Zhao, D.D. Petersen, J.C. Lambert, I. Cote, L. Yang, E. Healy, M.B. Black, H.J. Clewell, B.C. Allen, M.E. Andersen, Temporal concordance between apical and transcriptional points of departure for chemical risk assessment, Toxicol. Sci. 134 (1) (2013) 180–194.

[17] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second edition, Chapman and Hall/CRC Press, Boca Raton, FL, 2004.

[18] B. Wang, G. Gray, Concordance of noncarcinogenic endpoints in rodent chemical bioassays, Risk Anal. 35 (6) (2015) 1154–1166.

[19] K.S. Crump, A new method for determining allowable daily intakes, Toxicol. Sci. 4 (5) (1984) 854–871.

[20] W. Slob, Dose-response modeling of continuous endpoints, Toxicol. Sci. 66 (2) (2002) 298–312.