


ORIGINAL ARTICLE

Benchmark dose profiles for bivariate exposures

Tugba Akkaya Hocagil^{1,2}  | Louise M. Ryan³ | Richard J. Cook¹ | Khue-Dung Dang⁴ | R. Colin Carter⁵ | Gale A. Richardson⁶ | Nancy L. Day⁶ | Claire D. Coles⁷ | Heather Carmichael Olson⁸ | Sandra W. Jacobson⁹ | Joseph L. Jacobson⁹

¹Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

²Department of Biostatistics, School of Medicine, Ankara University, Ankara, Turkiye

³School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney, New South Wales, Australia

⁴School of Mathematics and Statistics, University of Melbourne, Melbourne, Victoria, Australia

⁵Division of Pediatric Emergency Medicine, Columbia University Medical Center, New York, New York, USA

⁶Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

⁷Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia, USA

⁸Department of Psychiatry and Behavioral Sciences, University of Washington, Washington, District of Columbia, USA

⁹Department of Psychiatry and Behavioral Neurosciences, School of Medicine, Wayne State University, Detroit, Michigan, USA

Correspondence

Tugba Akkaya Hocagil, Department of Statistics and Actuarial Science, University of Waterloo, Building M3 Room 4229, 200 University Ave West, Waterloo, Ontario, N2L 3G1, Canada. Email: takkayahocagil@uwaterloo.ca

Funding information

Foundation for the National Institutes of Health, Grant/Award Number: R01 AA025905

Abstract

While benchmark dose (BMD) methodology is well-established for settings with a single exposure, these methods cannot easily handle multidimensional exposures with nonlinear effects. We propose a framework for BMD analysis to characterize the joint effect of a two-dimensional exposure on a continuous outcome using a generalized additive model while adjusting for potential confounders via propensity scores. This leads to a dose–response surface which can be summarized in two dimensions by a contour plot in which combinations of exposures leading to the same expected effect are identified. In our motivating study of prenatal alcohol exposure, cognitive deficits in children are found to be associated with both the frequency of drinking as well as the amount of alcohol consumed on each drinking day during pregnancy. The general methodological framework is useful for a broad range of settings, including combinations of environmental stressors, such as chemical mixtures, and in explorations of the impact of dose rate rather than simply cumulative exposure on adverse outcomes.

KEYWORDS

benchmark dose analysis, benchmark dose profile, bivariate exposures, cognition, prenatal alcohol exposure

1 | INTRODUCTION

Benchmark dose (BMD) analysis aims to estimate a “dose” or level of exposure associated with an increase in the risk of an adverse response by a prespecified amount (called the benchmark response or BMR) (Budtz-Jørgensen et al., 2001; Crump, 1984; Erdreich & Stara, 1984). This so-called BMD is estimated by first modeling the dose–response relation-

ship between the exposure and response and then finding the exposure level leading to the expected increase in risk. Uncertainty is addressed by the computation of a BMDL, which corresponds to a lower confidence limit on the true BMD. Government and private entities employ BMDs and BMDLs for risk assessment in both the United States and the European Union (General Accounting Office, 2001; OECD, 2006, 2014; U.S. Environmental Protection Agency,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/), 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.

© 2024 The Authors. *Risk Analysis* published by Wiley Periodicals LLC on behalf of Society for *Risk Analysis*.

2012). Following the introduction of the initial safe dose for methylmercury onto the US Environmental Protection Agency (EPA)'s Integrated Risk Information System (IRIS) in 1995, guidance and broad explanations regarding the application of BMD methodology emerged (European Centre for Disease Prevention and Control, 2011; European Food Safety Authority (EFSA), 2017; Klaassen, 2012; U.S. Environmental Protection Agency, 2000). Subsequently, BMD analysis became the standard method for establishing safe exposure thresholds within environmental toxicology and regulatory frameworks (National Research Council (US) Committee on the Toxicological Effects of Methylmercury, 2000). Haber et al. (2018) provide an excellent summary, including a discussion of the recent trend to use model averaging techniques to address issues of model uncertainty.

Methods for BMD analysis are well-established for settings involving a single exposure but methods for multi-dimensional exposures are less developed. Chen et al. (1990) developed a method to compute the BMDL for studies of the chemical mixture using a likelihood ratio technique (Chen et al., 1990), but BMDLs were obtained separately for each chemical so joint effects were not addressed. Zhu (2005) developed estimators for the BMD for quantal data, with two factors modeled in tandem, but where one of these factors was taken as time and viewed as a secondary nuisance variable (Zhu, 2005). Deutsch and Piegorsch (2012, 2013) described a methodology for conducting a full BMD analysis with two exposures for quantal and nonquantal data. All of these approaches require the specification of a parametric form for nonlinear effects and do not accommodate covariate adjustment. Several researchers have argued that parametric dose–response modeling is too restrictive for BMD estimation, but little work has been carried out for BMD analysis involving flexible dose–response models. Wheeler and Bailer (2012) use Bayesian methods and Gibbs sampling to fit monotone P-splines, while Piegorsch et al. (2012) developed nonparametric methods for BMD estimation for quantile modeling. Lin et al. (2015) developed nonparametric methods for BMD estimation with continuous responses. Wheeler et al. (2015) present a nonparametric estimator based on quantile regression. All these nonparametric and semiparametric methods are developed for assessing the risk of a single exposure.

In our motivating study, there is evidence that cognitive deficits associated with the prenatal alcohol exposure (PAE) do not simply depend on the total amount of alcohol consumed during pregnancy, but rather on both the amount of alcohol consumed on drinking occasions, and the frequency of drinking (days/week) during pregnancy (Jacobson et al., 2024). Moreover, exploratory analyses suggest the presence of nonlinear, interacting effects of these variables.

In this paper, we propose a flexible framework for BMD analysis in which the dose–response surface characterizing the effect of two continuous exposure variables on a continuous response is fitted based on a generalized additive model (GAM; Wood, 2017); confounding variables can be accom-

modated in the regression model directly or using propensity scores. We then show how to estimate BMD contours from the dose–response surface. An innovative parametric bootstrap method is then employed to compute corresponding BMDL contours. We use this flexible modeling approach to gain insights into the effects of dose/occasion and drinking frequency on the child's cognition and derive BMD and BMDL contours.

The remainder of the paper is organized as follows. In Section 2, we describe six US cohort studies designed to assess the relationship between maternal alcohol intake during pregnancy and cognitive function measured in the offspring. Dose–response modeling involving harmonized data from these six studies motivates our work. After briefly reviewing how BMDs are calculated for continuous outcomes assessed in epidemiological settings, Section 3 extends this standard analysis to a setting involving GAMs and describes a bootstrap approach for calculating BMDLs derived from the fitted semiparametric model. In Section 4, we describe how this can then be extended to characterize the joint effects of a bivariate exposure variable. In all our analyses, the effects of potential confounders are addressed through regression adjustment on propensity scores. In Section 5, we apply the methods to our motivating data set and draw insights, with further discussion, conclusions, and topics warranting further research presented in Section 6.

2 | PRENATAL ALCOHOL AND CHILDHOOD COGNITION

Animal models and epidemiological studies have linked PAE to a broad range of long-term cognitive and behavioral deficits (Jacobson et al., 2024). However, there is relatively little consensus regarding the levels and patterns of PAE that are associated with clinically significant cognitive deficits. For these reasons, the US Office of the Surgeon General's (2005) recommendation (National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention (Internet), 2005) that all women abstain from alcohol during pregnancy continues to be the most appropriate guideline. However, there are compelling diagnostic reasons for studying the impact of dose/occasion and frequency of drinking on childhood cognition. This information, for example, can help clinicians differentiate between children affected by PAE and those with attention-deficit hyperactivity disorder (ADHD) and other conditions that give a similar clinical presentation (Peadon & Elliott, 2010). This distinction is critically important since the preferred treatment options are quite different (Petrenko & Alto, 2017). For example, methylphenidate is often used in treating ADHD, but it is generally ineffective for children impacted by the PAE (Mela et al., 2018). A more complete understanding of the etiology of cognitive limitations will help ensure timely and appropriate treatments are administered. This information may also provide some reassurance to women who may have inadvertently been exposed to low levels of alco-

hol before realizing they were pregnant. Traditionally, for diagnostic purposes, problematic exposure has been crudely characterized as a “pattern of excessive intake characterized by substantial, regular intake or heavy episodic drinking,” or the presence of social, legal, or medical problems related to drinking (Chudley et al., 2005; Cook et al., 2016; Hoyme, 2005; Stratton et al., 1996). Astley and Clarren (2000) suggested a quantitative criterion but emphasized that there is no “clear consensus on the amount of alcohol that can actually be toxic to the fetus.” More recently, risky drinking was defined as at least six drinks/week for >2 weeks, or three drinks on at least two occasions, but there is no empirical evidence of adverse effects on the offspring of women with these relatively low levels of exposure (e.g., a single drink/day over a 2-week period) or of women who did not binge drink (at least four to five drinks) on multiple occasions.

To help inform more evidence-based decision-making related to potentially risky drinking during pregnancy, the National Institutes of Health (NIH) funded a project in 2017 to combine data from six longitudinal cohort studies in order to obtain a sufficiently rich and large data set to rigorously explore the dose–response relationship between the PAE and child development. The cohorts, conducted in Detroit (Jacobson et al., 1993), Pittsburgh (Day et al., 1989; Richardson et al., 1999), Atlanta (Brown et al., 1998; Coles et al., 2006), and Seattle (Streissguth et al., 1981) had all been funded by the NIH between 1975 and 1993 and each recruited expectant women who were interviewed regarding their drinking behavior during pregnancy. Their offspring were then followed longitudinally from infancy through childhood and assessed periodically using a battery of neuropsychological tests to assess IQ and four domains of cognitive function: learning and memory, executive function, and academic achievement in reading and mathematics. The number of mother/infant pairs recruited for these cohorts ranged from 138 to 720, with a total combined sample size of 2226. Over the past several years, our team has published several articles focused on different aspects of the project. There have been several statistical methods publications addressing some of the analysis challenges, which have required nonstandard solutions. Topics addressed include how to apply meta-analysis methods to combine multiple outcome data from multiple studies (Akkaya Hocagil et al., 2022), how to use propensity scores in meta-analysis (Akkaya Hocagil et al., 2024) as well as several others (Dang, Ryan, Akkaya Hocagil et al., 2023; Dang, Ryan, Cook et al., 2023; Li et al., 2023). Akkaya Hocagil et al. (2021) extended the computation of propensity scores to account for the possibility that abstainers might differ from even very low-level drinkers and also allowed for long-tailed exposure distributions. All these various methodologies have been used to report on the substantive findings in the alcohol epidemiology literature. Jacobson et al. (2021) used the meta-analysis framework to establish that alcohol exposure had a similar magnitude of effect across a range of outcomes measured within all four cognitive domains. To address the dose–response question, Jacobson et al. (2024) describe the use of second-order confirmatory factor analysis, conducted

separately within each cohort, to obtain an overall cognitive function score for each child, which was then used as the outcome measure in a dose–response analysis approach. This paper found significant evidence that cognitive deficits associated with the PAE depend not simply on the total amount of alcohol consumed during pregnancy but on both the amount of alcohol consumed on each drinking occasion and the frequency of drinking (days/week) during pregnancy. Furthermore, the paper found evidence of nonlinearities in the dose–response pattern. This present paper builds on that previous work to derive a BMD that can be used to guide clinical guidelines.

Here, we summarize some of the key features of the six cohorts and refer readers to two earlier articles from our group (Jacobson et al., 2021, 2024) for further remarks on important epidemiological issues and limitations of these studies. A critically important issue is the reliability of the methods used to assess maternal alcohol usage during pregnancy. All interviews took place during pregnancy, with the exception of one study in which mothers were interviewed shortly after delivery. The number of maternal interviews varied by study. In these interviews, detailed information was obtained regarding their alcohol consumption, as well as various other baseline characteristics, including family socioeconomic status, maternal education (years), and cognitive ability. Because the covariates were assessed and reported slightly differently among the cohorts, we computed propensity scores for each study by fitting separate models for alcohol exposure (Akkaya Hocagil et al., 2024). In computing these propensity scores, we followed the recommended practice (Austin, 2011) of including all relevant, available covariates. To avoid potential overadjustment bias (Schisterman et al., 2009), only covariates evaluated at the time of study enrollment, during pregnancy, or around the time of the child’s birth were included (See Table S1 in Supplementary Materials). Propensity scores were then used as a covariate in a global response model characterizing the effects of alcohol on cognitive function. To reflect the fact that propensity scores were based on slightly different sets of covariates for each study, we also allowed the effect of the propensity score on the outcome to vary by the study by including a study by propensity score interaction (Akkaya Hocagil et al., 2024). Additional modeling details are provided in Section 5, as well as in Jacobson et al. (2024) who discuss the goodness-of-fit considerations and our rationale for using nonlinear models via GAMs.

All six cohorts reported continuous measures of maternal alcohol consumption during pregnancy. In particular, two exposure variables were reported: (1) alcohol dose per occasion, measured as the average number of standard drinks consumed per day on which drinking occurred, and (2) frequency of drinking, measured as the proportion of days during pregnancy when alcohol was consumed. Average absolute alcohol measured in ounces (oz) and consumed per day during pregnancy is then computed from the information collected directly in the interim for the other two exposure variables. For the purpose of our discussion, we consider all

three of these exposure variables as measures of “dose” and retain the terminology benchmark dose.

In this paper, we present a new procedure for BMD estimation that leads to a benchmark dose contour quantifying levels and patterns of exposure that are associated with the development of clinically important cognitive deficits in children born to mothers who consume alcohol during pregnancy.

3 | BMD ANALYSIS, SCALAR EXPOSURE VARIABLE

We begin with a review of a traditional BMD analysis involving a continuous response, a continuous exposure variable of interest, and several confounding variables. We formulate a linear model for the response as a function of the exposure while adjusting for confounders through the inclusion of a propensity score as an additional term in the regression. We then consider BMD analysis based on a GAM (Wood, 2017).

3.1 | BMD estimation based on a linear model

3.1.1 | Point estimation of the BMD

Let Y denote the response of interest, X a scalar continuous exposure variable, and \mathbf{Z} a vector of potential confounding variables. To mitigate the effect of the potential confounders, we estimate propensity scores for each individual by modeling the exposure X as a function of the potential confounders \mathbf{Z} through a linear regression model:

$$X = \alpha_0 + \alpha_1 \mathbf{Z} + R_x, \quad (1)$$

with α_0 and α_1 parameters to be estimated and R_x a mean zero error term, independent of \mathbf{Z} and with constant variance. While propensity score methodology was originally developed for the context of a binary exposure (e.g., treated vs. nontreated), a number of relatively recent developments have extended the methods to the context of continuous exposures (Hirano & Imbens, 2004; Kosuke & van Dyk, 2004). Upon fitting (1), let $S = \hat{E}(X|\mathbf{Z}) = \hat{\alpha}_0 + \hat{\alpha}_1 \mathbf{Z}$ denote the propensity score, corresponding to the predicted value of the exposure X , given the measured covariates \mathbf{Z} . While there are a number of ways propensity scores can be used, we regress upon the propensity score in the linear outcome model.

$$Y = \beta_0 + \beta_1 X + \beta_2 S + R_y, \quad (2)$$

where R_y is an error term, independent of both X and S and variance $\text{var}(Y|X, S) = \text{var}(R_y) = \sigma^2$. Provided there are no unmeasured confounders and that the propensity score model is correct, fitting model (2) yields a consistent estimate for

the causal effect of X , represented by β_1 . See Kosuke and van Dyk (2004) for further discussion about this approach. In our motivating study, Y is a measure of cognitive function. High values of Y are therefore desirable and hence if X represents some measure of alcohol intake we anticipate $\beta_1 < 0$. For convenience in what follows, we let $\mu(X, S) = \beta_0 + \beta_1 X + \beta_2 S$ denote the predicted outcome given exposure X and propensity score S .

Crump (1984) originally proposed the estimation of a BMD in the context of a binary outcome but generalized the method to accommodate continuous outcomes. Specifically, he assumed that an individual has had an adverse effect if their outcome falls in the extreme tail (in our application this will be the lower tail) of the response distribution (Crump, 1995). In practice, one can specify the tail area directly (e.g., IQ scores below 75 are considered adverse) or indirectly (e.g., the lower 5% of the response distribution of unexposed individuals are considered to have an adverse response). There are computational and analytical advantages of the indirect approach (Budtz-Jørgensen et al., 2001), especially in settings where the definition of an abnormal response depends on the confounders and we will use this approach. We let p_o denote the proportion of unexposed individuals considered to be in the adverse range. The threshold $\tau_o(S)$ is the 100 p_o th percentile of the response distribution for a subject with propensity score S ; thus the probability a response falls below $\tau_o(S)$ for unexposed individuals ($X = 0$) with propensity score S is p_o . In Figure 1, we display quantities p_o and $\tau_o(S)$ on the left side with the fitted dose–response line depicting a decreasing mean response with increasing levels of exposure, consistent with what would be expected when Y is cognitive function and X is PAE.

Under the response model (2) and under the additional assumption that the error terms (R_y) are normally distributed,

$$P(Y < \tau_o(S) | X = 0, S) = P\left(\frac{Y - \mu(0, S)}{\sigma} < \frac{\tau_o(S) - \mu(0, S)}{\sigma}\right),$$

which if $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable, gives the following relation between p_o and $\tau_o(S)$:

$$p_o = \Phi((\tau_o(S) - \mu(0, S))/\sigma).$$

Solving for $\tau_o(S)$ gives

$$\tau_o(S) = \mu(0, S) + \sigma \Phi^{-1}(p_o). \quad (3)$$

Computing a BMD requires specification of the BMR, or the specified increase in the proportion of individuals with responses in the abnormal range (i.e., with $Y < \tau_o(S)$). Choosing a suitable value for the BMR generally requires judgment and consideration regarding the nature of the data set and the context in which the BMD will be used. A BMR will usually be chosen so that the corresponding BMD estimate can be reliably estimated, but then additional uncer-

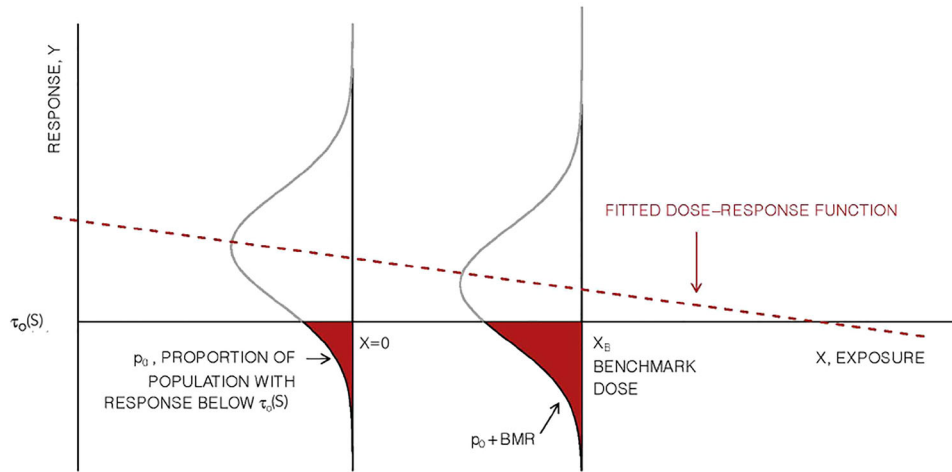


FIGURE 1 A schematic illustrating the derivation of a benchmark dose with continuous exposure and response; exposure level is represented on the horizontal axis and the distribution of the response is depicted in the vertical axis. BMR, benchmark response.

tainty factors might be applied to linearly extrapolate to the risk levels of specific regulatory interest (U.S. Environmental Protection Agency, 2012). When analysis is based on epidemiological data, a value of 1% or 5% is typically used for the BMR. The BMD, denoted here by X_B , is the dose of exposure that results in an increase in the proportion of responses in the abnormal range in the population corresponding to the BMR. That is, the BMD satisfies:

$$P(Y < \tau_0(S) | X = X_B, S) = p_0 + \text{BMR}, \quad (4)$$

where the right-hand side of (4) is the proportion of the unexposed population ($X = 0$) with abnormally low responses plus the additional proportion in this range (BMR) due to the BMD of exposure. The normal distribution displayed on the right-hand side of Figure 1 is the response distribution at X_B , the BMD. The increase in risk of falling below the threshold $\tau_0(S)$ is evident from the larger shaded area. The decreasing mean response with increasing exposure X is depicted on the right is determined by the linear dashed plot in Figure 3.

Again based on the linear model, if we further assume Gaussian errors we have

$$P(Y < \tau_0(S) | X = X_B, S) = \Phi\left(\frac{\tau_0(S) - \mu(X_B, S)}{\sigma}\right),$$

giving the relation

$$\Phi^{-1}(p_0 + \text{BMR}) = (\tau_0(S) - \mu(X_B, S))/\sigma. \quad (5)$$

Replacing $\tau_0(S)$ in (5) with the expression in (3), we obtain

$$\Phi^{-1}(p_0 + \text{BMR}) - \Phi^{-1}(p_0) = \frac{\mu(0, S) - \mu(X_B, S)}{\sigma}. \quad (6)$$

For the additive linear model in (2), the right-hand side reduces to $-\beta_1 X_B/\sigma$ which does not depend on S , and hence the confounding variables. It is this simplification which

gives the appeal of the indirect approach to computing $\tau_0(S)$ (Budtz-Jørgensen et al., 2001). Letting

$$A = A(p_0, \text{BMR}) = \text{fl}^{-1}(p_0 + \text{BMR}) - \text{fl}^{-1}(p_0), \quad (7)$$

we can explicitly write

$$X_B = -A\sigma/\beta_1. \quad (8)$$

The quantity in (8) is estimated by replacing the unknown parameters, β_1 and σ , by their respective estimated values to obtain $\hat{X}_B = -A\hat{\sigma}/\hat{\beta}_1$.

3.1.2 | Estimation of the BMDL in the linear model

The BMDL is a lower confidence limit for the BMD and reflects the statistical uncertainty associated with the use of estimated parameters instead of their true values. In practice, a lower one-sided 95% or 97.5% confidence interval (CI) is often used. This can be obtained in a number of ways. Crump (1984, 1995) proposed using likelihood-based limits to calculate the BMDL. Budtz-Jørgensen et al. (2001) derive the exact sampling distribution of the estimated BMD and from this are able to find an exact formula for a BMDL in the context of a linear model. For more complex dose–response functions, there will generally be no simple closed-form expression for the BMD. We therefore describe and implement an alternative approach that exploits the duality between hypothesis testing and CI estimation. CIs for a parameter of interest can be constructed by the set of values that are not rejected by a corresponding hypothesis test. All parameter values not rejected by profile likelihood ratio tests at the 5% significance level, for example, form 95% likelihood ratio-based CIs (Sprott, 2008); profile likelihood ratio intervals can be likewise constructed in multiparameter problems. An exact CI for a binomial probability is another example, and in

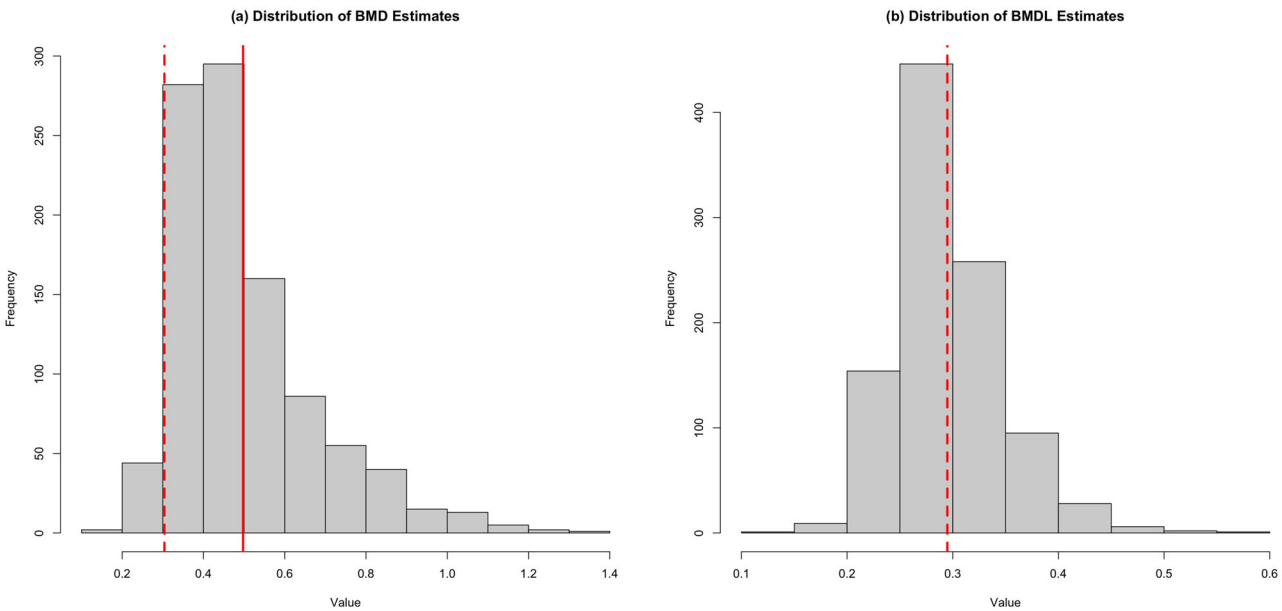


FIGURE 2 Results based on 1000 simulated data sets. Panel (A) shows a histogram of the 1000 estimated benchmark doses (BMDs) while Panel (B) shows a histogram of a lower confidence limit on the true BMD (BMDL) estimates based on 2000 parametric bootstrap replications for each of the 1000 simulated data sets.

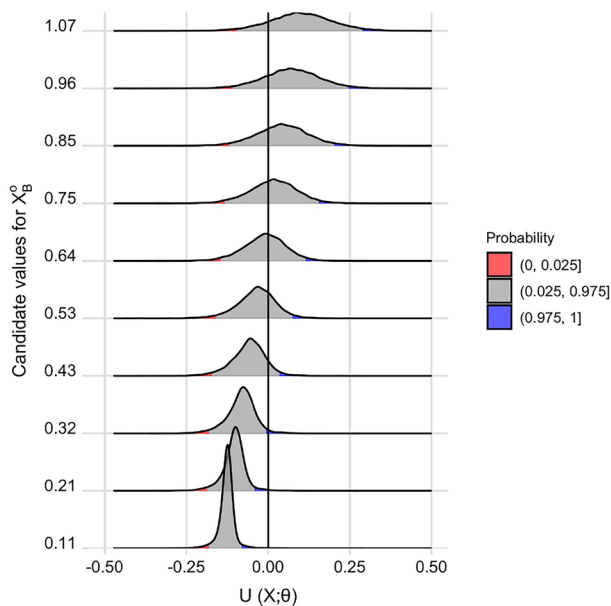


FIGURE 3 Smoothed histograms of the bootstrap distribution of the estimated benchmark dose (BMD) functional $U(X_B^o; \hat{\theta})$ computed on a grid of candidate BMD values, X_B^o . Here, X_B^o is measured as the logarithm of average absolute alcohol per day, averaged across pregnancy. The BMD is estimated as ≈ 0.64 and the lower confidence limit on the true BMD (BMDL) is estimated as ≈ 0.32 oz of absolute alcohol (log scale).

failure time analysis, a 95% CI for the median survival time is defined implicitly by the survival times at which a test that the survival probability is 0.5 is not rejected. More generally, Hu and Kalbfleisch (2000) discuss the use of estimating functions to test hypotheses, and use resampling methods to carry out such tests and define corresponding CIs.

The BMD for the linear model is the solution to $U(X, \theta) = 0$ where

$$U(X; \theta) = \frac{(-X)\beta_1}{\sigma} - A \tag{9}$$

is a functional, with θ a vector containing the unknown parameters β_1 and σ . The BMD is the value of X such that $X_B\beta_1 = -\sigma A$ and the estimated BMD, \hat{X}_B , is the solution to $U(X; \hat{\theta}) = 0$, where θ has been replaced by $\hat{\theta}$. It is the use of these estimated parameters in $U(X; \theta)$ that leads to uncertainty regarding the BMD and, hence, the need to compute a lower confidence limit defining the BMDL. The BMD defined in terms of (9) is a function of the maximum likelihood estimates and so is invariant to reparameterization.

Note that the null hypothesis $H_0 : X_B = X_B^o$ is equivalent to the null hypothesis $H_0 : U(X_B^o; \theta) = 0$. It follows that a 95% CI for X_B can be defined as the set of X_B^o values such that this latter null hypothesis is not rejected based on the test statistic $U(X_B^o; \hat{\theta})$. There are several ways to do this. If $U(X_B^o; \hat{\theta})$ is assumed to be asymptotically normal and a variance estimate of $U(X_B^o; \hat{\theta})$ is known, then the null hypothesis $H_0 : U(X_B^o; \theta) = 0$ would not be rejected if

$$-1.96\sqrt{\text{var}\{U(X_B^o; \hat{\theta})\}} \leq U(X_B^o; \hat{\theta}) \leq 1.96\sqrt{\text{var}\{U(X_B^o; \hat{\theta})\}}. \tag{10}$$

That is a two-sided 95% CI for the BMD corresponds to the set of

X_B^o values such that (10) is satisfied and the lower BMDL (lower 2.5% confidence limit) corresponds to the smallest value within this set. It is straightforward to establish that (10)

holds if the following interval includes 0:

$$\left(U(X_B^o; \hat{\theta}) - 1.96\sqrt{\text{var}\{U(X_B^o; \hat{\theta})\}}, U(X_B^o; \hat{\theta}) + 1.96\sqrt{\text{var}\{U(X_B^o; \hat{\theta})\}} \right). \quad (11)$$

Because the functional $U(X; \theta)$ equals $-A$ by definition when X equals zero, the right-hand side of interval (11) will first cross 0. Hence, the BMDL will correspond to the smallest value of X_B^o such that $U(X_B^o; \hat{\theta}) + 1.96\sqrt{\text{var}\{U(X_B^o; \hat{\theta})\}} = 0$.

In complex semiparametric models, there are considerable computational advantages to deriving CIs for the BMD indirectly through the functional $U(X_B^o; \hat{\theta})$ since this quantity may generally be much easier to calculate than the BMD itself. As we show shortly, however, it may be difficult to derive an analytic expression for $\text{var}\{U(X_B^o; \hat{\theta})\}$. We circumvent this problem through use of a parametric bootstrap procedure (Efron, 2000). For the linear model, the first step is to fit (2) and compute the predicted response $\hat{\mu}(X_i, S_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 S_i$ for each individual i , where X_i and S_i are the observed values of exposure and propensity score for that individual. Next, using a parametric bootstrap, we generate M samples of size n where for the m th sample we:

1. For each individual i , we generate a new residual $R_{y_i}^{(m)}$ from a Gaussian distribution with mean zero and variance $\hat{\sigma}^2$ where $\hat{\sigma}^2$ is the estimated error variance based on the model fitted to the original data;
2. Form a new outcome for each individual by adding the predicted value and residual from Step 1 to get $Y_i^{(m)} = \hat{\mu}(X_i, S_i) + R_{y_i}^{(m)}$;
3. Given the m th bootstrap sample we fit the model (2) and record $\hat{\theta}^{(m)} = (\hat{\beta}^{(m)}, \hat{\sigma}^{(m)})$ and compute the value of the functional $U(X_B^o; \hat{\theta}^{(m)})$ at a candidate BMD value X_B^o on a grid of points.

Once the M bootstrap samples have been generated a bootstrap variance estimate $\widehat{\text{var}}\{U(X_B^o; \hat{\theta})\}$ can be obtained and (10) or (11) can be used.

An alternative approach to interval estimation is to base it on the quantiles of the bootstrap distribution of the functional—this has the advantage of relaxing the normality assumption required for CI construction based on $\widehat{\text{var}}\{U(X_B^o; \hat{\theta})\}$ and therefore handles a wider range of problems. The procedure is as follows. Let $Q_p(X_B^o; \hat{\theta})$ be the 100 p th percentile of the bootstrap distribution of $U(X_B^o; \hat{\theta})$ and consider an approximate 95% CI for the functional $U(X_B^o; \theta)$ given by

$$\{Q_{2.5}(X_B^o) < U(X_B^o; \hat{\theta}) < Q_{97.5}(X_B^o)\}.$$

Because the quantity $U(X; \theta)$ equals $-A$ when $X = 0$, the BMDL corresponds to the smallest value of X_B^o for which

$Q_{97.5}(X_B^o) = 0$. This value can be obtained by a simple grid search. A large number of bootstrap samples is recommended when estimating extreme quantiles to ensure accurate coverage probabilities.

In Section 5, we will see that in the linear model setting, the parametric bootstrap approach yields almost identical BMDL estimates as the analytic approach of Budtz-Jørgensen et al. (2001), giving confidence that the bootstrap approach will also work well in settings where the simpler solution is not available. In particular, the next section describes how the methodology described above can be used to estimate a BMD and BMDL based on a flexible nonlinear model where no simple analytic solution is available. Section 3.3 will report on a small simulation study designed to assess the performance of the approach.

3.2 | Estimation of the BMD and BMDL with GAMs

To relax the assumption of a linear effect of the covariates in the response model in (2), we consider a GAM (Hastie & Tibshirani, 1986) allowing the effects of X and S to be expressed through

$$Y = \beta_0 + f(X) + g(S) + R_y, \quad (12)$$

where $f(\cdot)$ and $g(\cdot)$ are smooth weakly parametric functions and R_y is an error term, independent of X and S with $R_y \sim N(0, \sigma^2)$. In practice, these smooth terms are estimated using weakly parametric functions such as splines. We used the implementation developed by Simon Wood through his R package `mgcv` (Wood, 2017; Wood et al., 2016) which uses B-splines as the default, though many other options are available (Harrell, 2015). GAMs accommodate nonlinear relations between covariates (here exposure X or propensity score S) and the response; the functions accommodate quite general relationships but reduce in special cases to familiar forms (i.e., $f(X) = X\beta_1$ in the linear case).

We somewhat informally refer to these functions as parameters to be estimated in what follows. The derivation of the previous section applies here unchanged up until Equation (6), but when the effect of X is expressed through is replaced by

$$[\Phi^{-1}(p_o + \text{BMR}) - \Phi^{-1}(p_o)] = [f(0) - f(X)]/\sigma. \quad (13)$$

While it is logical to think that $f(0) = 0$, this is not necessarily so, depending on the specific formulation and model fitting strategy. For the `gam` function from the `mgcv` package (Wood, 2017), $f(x)$ is estimated in such a way that the estimate is 0 at the average x value. Indeed, it is the requirement to estimate $f(0)$ that complicates the computation of a BMD in our setting and which motivates our use of the parametric bootstrap. We let $A = A(p_o, \text{BMR})$ denote the fixed left-hand side of (13) as defined previously at (7). If we again let $\theta = (f(\cdot), \sigma)$ denote the set of parameters to be estimated for the compu-

tation of the right-hand side of (13), we can define a new functional

$$U(X; \theta) = (f(0) - f(X))/\sigma - A$$

such that setting $U(X; \hat{\theta}) = 0$ yields an estimated BMD, \hat{X}_B . Note that $\hat{\theta} = (\hat{f}(\cdot), \hat{\sigma})$ now represents the estimated function characterizing the effect of X and the residual standard error: it is the sampling variation in these that leads to the uncertainty in the estimate \hat{X}_B as an estimate of the true BMD, X_B .

To construct a 95% CI for the BMD, we use the parametric bootstrap approach described in Section 3.1.2. Specifically, for each data point (Y_i, X_i, S_i) in the sample, we generate a predicted value based on the fitted model and generate a new residual to obtain a new value of the outcome Y_i , according to (12) using $\hat{f}(\cdot)$, $\hat{g}(\cdot)$, and $\hat{\sigma}$ in place of their true values and let $\hat{\theta}^{(m)} = (\hat{f}^{(m)}(\cdot), \hat{\sigma}^{(m)})$ be the fitted dose–response curve and error standard deviation from the m th bootstrap sample. Next, we consider a grid of candidate values for X_B . We let $U(X_B^{\circ}; \hat{\theta}^{(m)})$ denote the estimated functional value at a candidate BMD value of X_B° , for $m = 1, \dots, M$. From this, we can obtain a bootstrap distribution of $U(X_B^{\circ}; \hat{\theta})$.

Using the quantile approach, we let $Q_{\alpha}(X_B^{\circ})$ denote the 100α th percentile of the bootstrap distribution of $U(X_B^{\circ}; \hat{\theta})$ generated at a specific value, X_B° . and define a 95% CI for $U(X_B^{\circ}; \theta)$ as:

$$Q_{2.5}(X_B^{\circ}) < U(X_B^{\circ}; \theta) < Q_{97.5}(X_B^{\circ}). \quad (14)$$

The BMDL is then the smallest value of X_B° for which this interval includes 0. As before, because the quantity $U(X; \theta)$ degenerates to $-A$ by definition when $X = 0$, the BMDL will correspond to the smallest value of X_B° for which $Q_{97.5}(X_B^{\circ}) = 0.025$. As described in the previous section involving the linear model, this can be obtained by a grid search. An alternative is to fit an auxiliary generalized additive logistic regression model with response $\Delta^m(X_B^{\circ}) = I(U(X_B^{\circ}; \hat{\theta}^{(m)}) > 0)$, an indicator of whether the functional for the m th bootstrap sample and evaluated at X_B° exceeds 0. The covariate of this auxiliary model is the exposure and the purpose of fitting it is to understand how the quantiles of the bootstrap distribution vary as a function of the exposure to identify the BMDL.

Specifically for each X_B° in a set of grid points we take the data as $\{(\Delta^m(X_B^{\circ}), X_B^{\circ}), m = 1, \dots, M\}$, to which we fit an auxiliary generalized additive logistic regression model with a smooth function of X_B° to model the relation between X_B° and $P(\Delta^m(X_B^{\circ}) = 1)$. If $\hat{\pi}(X_B^{\circ})$ denotes the fitted value from this auxiliary GAM, we find the value of X_B° that yields

$$\hat{\pi}(X_B^{\circ}) = 0.025$$

to define the BMDL.

We describe the application of this approach more detail in the application, but first we present a small simulation study before discussing the use of GAMs with bivariate exposures.

3.3 | An illustrative simulation study

Here, we report on a small simulation study to evaluate the empirical performance of the proposed approach for estimating the BMD and associated BMDL.

We specify the true dose–response function as $f(x) = \alpha_1 * \exp(\alpha_2 * (x + 1))$ and generate $Y_i = f(x_i) + R_{yi}$ with $R_{yi} \sim N(0, \sigma^2)$ with $\sigma^2 = 1.5$ for $i = 1, 2, \dots, n$ where $n = 2000$. We set $\alpha_1 = 0.1$ and $\alpha_2 = 0.3$, corresponding to a true BMD value of 0.5. The values of x were generated from a zero-inflated gamma distribution to mimic the long-tailed exposure distribution seen in our application. After fitting the model to obtain $(\hat{f}(\cdot), \hat{\sigma}^2)$, we then create $M = 2000$ bootstrap samples and obtain a bootstrap distribution for $U(X_B^{\circ}; \hat{\theta})$ for each value on a grid of points over 5001 grid points ranging from the minimum (0.00) to the maximum (5.30) values of x . This is used to obtain the estimated BMDL for the simulated sample via an auxiliary generalized additive logistic model as described in Section 3.2. We then repeated this for $nsim = 1000$ data sets to obtain empirical distributions of the estimated BMDs and BMDLs.

Figure 2A illustrates the empirical distribution of the BMD estimated for $nsim = 1000$ samples and Figure 2B illustrates the empirical distribution of the corresponding BMDLs. We see from Panel (A) that the BMD estimates across simulations are dispersed about the true BMD value of 0.5, suggesting that the use of the GAM can characterize the dose–response curve well. The red-dashed vertical line in Panel (A) is the empirical fifth percentile of the BMD estimates across the 1000 simulated samples. This value, 0.29, can thus be interpreted as an estimate of the true BMDL for the given setting and sample size. Panel (B) shows the empirical distribution of the estimated BMDLs across 1000 samples with vertical red-dashed showing the average of the 1000 estimated BMDLs. The latter is very close to 0.29, which corresponds to the empirical fifth percentile of BMD estimates shown in Panel (A). These results provide assurance that the generalized additive modeling approach to estimating BMD and BMDL can be implemented and yields estimates with good empirical properties.

4 | BMD VIA GAMS WITH A BIVARIATE EXPOSURE

In this section, we extend our approach to the setting where the interest lies in the bivariate exposure variable denoted by $X = (X_1, X_2)$, with flexible modeling of the interaction between the components. When interest lies in evaluating the causal effects of multiple exposures, Li et al. (2023) argue that it is necessary to consider propensity scores based

on the joint distribution of these exposures, given potential confounders. More details of our approach are provided in Supplementary Materials, but for here, it suffices to denote by $\mathbf{S}(\mathbf{Z})$ a vector-valued propensity score and we then specify a GAM with the following functional form for the mean of Y :

$$\mu(X, \mathbf{Z}) = \beta_0 + f(X_1, X_2) + \mathbf{g}(\mathbf{S}(\mathbf{Z})), \quad (15)$$

where $f(\cdot, \cdot)$ is now a smooth, flexible bivariate function and $\mathbf{g}(\cdot)$ is a vector of terms that adjust for the vector-valued propensity score (see Supplementary Materials for more details). We use the `gam` function available through the R package `mgcv` (Wood, 2017), which uses tensor products to produce the bivariate smooth. For a fixed set of confounders whose effects are captured through the propensity scores $\mathbf{S}(\mathbf{Z})$, we can plot the three-dimensional surface $f(X_1, X_2)$ to give insight into the nature of the dose–response surface. Contour plots of this surface may also be constructed by fixing one exposure variable and examining the trend in mean response as a function of the other. In the context of the study of chemical mixtures or multidrug interactions in health research, such contours are referred to as isobolograms (Huang et al., 2019). The BMD which is a scalar for univariate exposure, now becomes a two-dimensional benchmark curve corresponding to sets of values of the two exposure variables. To describe this in more detail, we first define the functional that needs to be set to 0 to find the BMD. This is

$$U(X; \theta) = [f(0, 0) - f(X_1, X_2)]/\sigma - A, \quad (16)$$

where $f(0, 0)$ refers to the predicted mean for unexposed individuals ($X_1 = X_2 = 0$) and A is defined as in (7). Then, the BMD curve corresponds to the set of values $\mathbf{X}_B = (X_{1B}, X_{2B})$ that satisfy $U(\mathbf{X}_B; \theta) = 0$.

As in the univariate case, unknown functions in (16) and parameters must be estimated and we do so using the `gam` function in R (Wood, 2017) (see Supplementary Material for further detail). The estimated BMD curve corresponds to the set of points that satisfy $U(\mathbf{X}_B; \hat{\theta}) = 0$, where

$$U(\mathbf{X}_B; \hat{\theta}) = [\hat{f}(0, 0) - \hat{f}(X_{1B}, X_{2B})]/\hat{\sigma} - A.$$

We achieved this by a grid search in the \mathbf{X} -space.

Estimation of the BMDL contour for bivariate exposure is best carried out via the parametric bootstrap approach described in Section 3.2. Let $\hat{\theta}^{(m)} = (\hat{f}^{(m)}(\cdot, \cdot), \hat{\sigma}^{(m)})$ be the estimated parameters from the m th bootstrap sample and let $U(\mathbf{X}; \hat{\theta}^{(m)})$ denote the estimating function evaluated at an exposure point \mathbf{X} for the m th bootstrap sample. We consider a grid of points in the X -plane, denoting a specific location in this grid by $\mathbf{X}_B^\circ = (X_{1B}^\circ, X_{2B}^\circ)$, and consider the set of values of the estimating function evaluated at this grid point, for each of the bootstrap samples: $U(\mathbf{X}_B^\circ; \hat{\theta}^{(m)})$, $m = 1, \dots, M$. As in Section 3.2, we let $Q_\alpha(\mathbf{X}_B^\circ)$ be the 100α th percentile of the bootstrap distribution of $U(\mathbf{X}_B^\circ; \hat{\theta})$ evaluated at \mathbf{X}_B° . The

goal is now to find the contour corresponding to the smallest sets of X_{1B}° and X_{2B}° such that $Q_{97.5}(\mathbf{X}_B^\circ) = 0$. Finding this contour is more complicated now due to the bivariate nature of \mathbf{X}_B° . We use the strategy of creating a set of binary indicators and then performing an auxiliary analysis. Specifically, let $\Delta^m(\mathbf{X}_B^\circ) = I\left(U(\mathbf{X}_B^\circ; \hat{\theta}^{(m)}) > 0\right)$ be an indicator of whether the estimating function, computed for the m th bootstrap sample and evaluated at \mathbf{X}_B° , exceeds zero. This results in a data set $\{(\mathbf{X}_B^\circ, \Delta^m(\mathbf{X}_B^\circ)); m = 1, \dots, M\}$ for each \mathbf{X}_B° in a chosen set of grid points. Based on these data, we then fit a generalized additive logistic regression model with a tensor product used to model the relation between \mathbf{X}_B° and $\hat{\pi}(\mathbf{X}_B^\circ) = P(\Delta^m(\mathbf{X}_B^\circ) = 1 | \mathbf{X}_B^\circ)$. Here, we aim to find the combinations of $\mathbf{X}_B^\circ = (X_{1B}^\circ, X_{2B}^\circ)$ that ensure

$$\hat{\pi}(\mathbf{X}_B^\circ) = 0.025.$$

Now we will illustrate how the methods all work for our motivating example.

5 | APPLICATION TO THE PAE STUDY

Here, we apply the proposed methodology in Section 2 to examine the effect of PAE on child cognition. The data are from six US longitudinal cohort studies aiming to quantify levels and patterns of exposure that are associated with the development of clinically important cognitive deficits in children born to mothers who consume alcohol during pregnancy. In all six cohorts, data on maternal alcohol consumption were quantified in terms of daily alcohol intake (standard drinks or oz of absolute alcohol per day averaged across pregnancy as well as the frequency of drinking days and the amount of alcohol consumed on each drinking day during pregnancy. The distributions of daily alcohol intake are summarized by cohort in the Supplementary Materials. The daily alcohol intake (standard drinks or oz of absolute alcohol per day averaged across pregnancy and the amount of alcohol consumed on each drinking day during pregnancy were highly skewed (skewness > 3.0). We adopted a $\log(X + 1)$ transformation to improve model stability. Further discussion related to our decision to use this logarithmic transformation may be found in Supplementary Materials. We also provide there some more detailed discussion regarding the distributions of daily alcohol intake averaged across pregnancy by cohort, showing histograms of these variables in the original scale and in the log scale (see Figures S1 and S2). In our recent article, Jacobson et al. (2024) discuss the results of exploratory analyses on the harmonized data set and highlight the need for models accommodating nonlinear interaction effects when considering bivariate exposure.

We consider a BMD analysis for a continuous response, cognitive function score, based on a linear model involving a single continuous exposure, daily alcohol intake averaged across pregnancy (log scale), and the estimated propensity

TABLE 1 Benchmark dose (BMD) and the lower confidence limit on the true BMD (BMDL) estimates for the effect of prenatal exposure measured in ounces of absolute alcohol consumption per day during pregnancy.

Dose–response model	BMD	BMDL	
		Budtz-Jørgensen et al. (2001)	Functional bootstrap
Linear model	0.55	0.34	0.35
Generalized additive model	0.67		0.31

score. We set $BMR = 0.01$ and $p_0 = 0.025$ and obtain a point estimate of the BMD and corresponding 97.5% lower confidence limit BMDL. We estimated the BMDL using the two approaches described in Section 3.1.2. In the first approach, we approximated the 97.5% lower confidence limit for the BMD using the analytical formula derived by Budtz-Jørgensen et al. (2001). For the parametric bootstrap approach, a lower confidence limit for the BMD is approximated by the 2.5th percentile from the bootstrapped distribution of the estimated BMD, based on simulating 10,000 new parametric bootstrap samples from the fitted model. Since we use a parametric bootstrap approach, cohort sample sizes for all six cohort studies and exposure distributions are unchanged across the bootstrapped data sets. The first row of Table 1 shows the estimated BMD and two BMDL estimates which are seen to be very similar.

To relax the assumption of a linear effect of the average alcohol consumption per day across pregnancy and the propensity score, we then fitted a GAM using the `gam` function from the `mgcv` package in R (Wood, 2017) and a BMDL was estimated using the parametric bootstrap approach described in Section 3.2, setting $M = 10,000$. We considered a grid search consisting of 100 candidate values for X_B° , evenly distributed between the minimum (0.00) and maximum (3.54) amount of alcohol consumed on each drinking day during pregnancy recorded in the data set and generated a bootstrap distribution of the functional $U(X^\circ; \hat{\theta})$ at each grid point. As discussed previously, the BMDL was chosen as smallest exposure level where the 97.5th percentile of the bootstrap distribution equaled 0. The resulting BMD and BMDL estimates appear in the second row of Table 1. The BMD estimate from the GAM is higher than what those of the linear model but the corresponding BMDL estimates are similar.

To provide a visualization of how the bootstrap approach works, Figure 3 shows smoothed histograms of the bootstrap distribution of $U(X^\circ; \hat{\theta})$ at a discrete set of values of X° . The vertical line represents the point 0 and for each histogram, tail areas have been shaded below the 2.5th percentile, and above the 97.5th percentile. The BMDL corresponds to the lowest value of X_B° for which 97.5th percentile of the bootstrap distribution equals 0. In the figure, we can see that this indeed

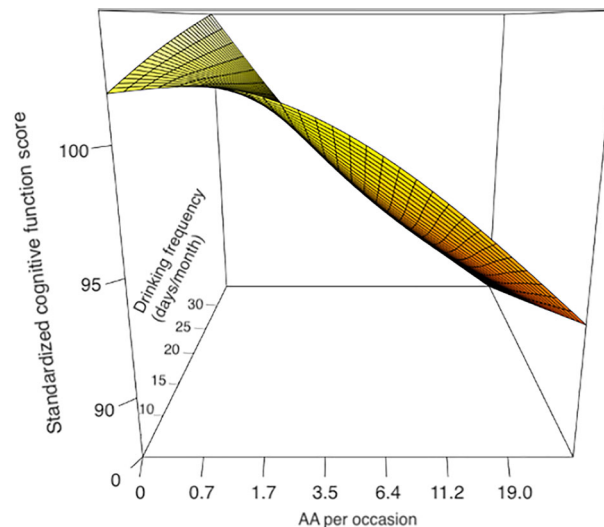


FIGURE 4 Surface plot depicting effects of maternal alcohol dose/occasion and drinking frequency (adjusted for potential confounders) on offspring cognitive function derived from the bivariate smooth generalized additive model (GAM). AA, absolute alcohol.

happens right around $X_B^\circ = 0.32$ and this is consistent with our analysis which yields the value of 0.31 for the BMDL.

We next consider estimation of a BMD curve and associated BMDL in terms of a bivariate exposure. To demonstrate our proposed method, we investigate the role of the pattern of exposure where we distinguish between consistent low levels of exposure and episodic periods of higher exposure (binge drinking) that may result in a similar total alcohol intake over a given period of time. We fitted the model described in Section 4 to predict cognition, with X_1 and X_2 representing alcohol dose per occasion (log scale) and frequency of drinking, respectively. Using the approach described in Section 4, we used a bivariate GAM in order to construct a flexible three-dimensional surface predicting cognitive function in relation to alcohol dose per occasion (log scale) and drinking frequency, while adjusting for confounders via propensity scores. The bivariate GAM approach was justified on the basis of previous exploratory data analysis that suggested the presence of both nonlinearities and interactions between these two different dimensions of alcohol exposure (Jacobson et al., 2024). Further discussion may be found in Supplementary Materials.

Figure 4 presents the estimated dose–response surface obtained from the fitted GAM. The curve depicted at the front edge of the surface plot shows a nonlinear effect of dose per occasion among infrequent drinkers. At low levels of dose per occasion, the curve is quite flat, suggesting little adverse effect at these drinking levels, for infrequent drinkers. As the dose per occasion increases beyond about 1.7 oz, the curve decreases as the frequency increases. At the highest levels of drinking frequency (i.e., daily or 30 days per month), the curve slopes sharply downward in a relatively linear manner, suggesting that for frequent drinkers, any increase in

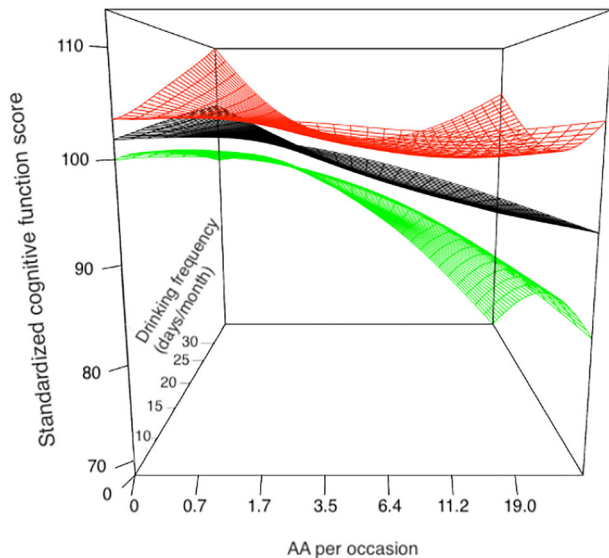


FIGURE 5 Surface plot with confidence interval surfaces at ± 2 standard errors depicting effects of maternal alcohol dose/occasion and drinking frequency (adjusted for potential confounders) on offspring cognitive function derived from the bivariate smooth generalized additive model (GAM). AA, absolute alcohol.

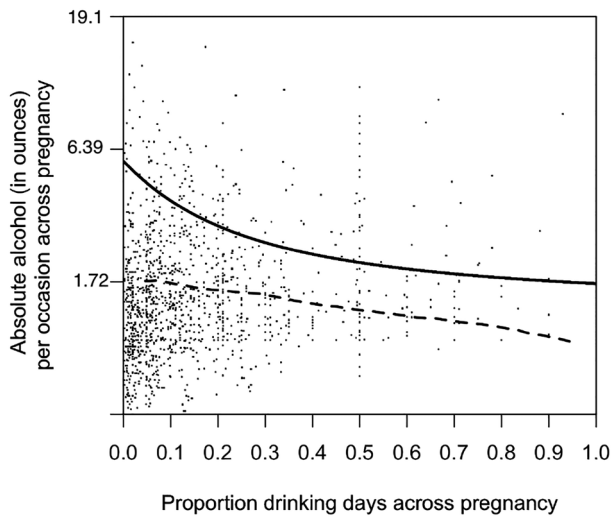


FIGURE 6 Estimated benchmark dose (BMD) and benchmark dose lower bound curves. The solid line represents the estimated BMD curve and the dashed line represents the estimated lower confidence limit on the true BMD (BMDL) curve using the parametric bootstrap approach.

dose/occasion is associated with an appreciable adverse effect on cognitive function.

Figure 5 shows the estimated dose–response surface of the exposure effects with CI surfaces at ± 2 standard errors. These wider pointwise confidence surfaces reflect greater uncertainty in the model at higher levels of drinking, where there were fewer women in the cohorts.

The solid line in Figure 6 shows the estimated BMD curve overlaid on a scatter plot of alcohol per occasion (y-axis) versus drinking frequency in the proportion of days per month

(x-axis). This BMD curve shows the combinations of dose per occasion and drinking frequency associated with the effects defining the BMD. We obtained a point estimate for the BMD curve for the combination of values of alcohol per occasion and drinking frequency by a grid search in the X -space. We estimated the BMDL contour using the parametric bootstrap approach described in Section 4. We searched over a grid containing 100 different candidate values of X_{1B} and X_{2B} , yielding a total of 10,000 grid points to search. We used $M = 10,000$ bootstrap samples for the estimated functional $U(\mathbf{X}_B^\circ; \hat{\theta})$ at each candidate point in the grid, $\mathbf{X}_B^\circ = (X_{1B}, X_{2B})$. Next, for each bootstrapped sample, we computed the set of indicators, $\Delta^m(\mathbf{X}_B^\circ) = I(U(\mathbf{X}_B^\circ; \hat{\theta}^{(m)}) > 0)$ of whether the m th bootstrapped estimated estimating function at each grid point exceeded zero. Finally, we fitted a generalized additive logistic regression model with a tensor product of the logarithm of absolute alcohol per occasion and drinking frequency to find the combinations of the logarithm of absolute alcohol per occasion and drinking frequency where 97.5% of the bootstrap distribution exceeds the threshold. Figure 6 shows the estimated BMDL (dashed line) that is obtained from the parametric bootstrap approach. The results suggest an estimated BMDLs range from around 1.5 oz for infrequent drinkers to just over 0.6 oz for frequent or daily drinkers.

All analyses were conducted in R (R Core Team, 2021). The R Code for our bivariate modeling can be found in the Appendix and the data used in the analysis have been uploaded to the Harvard Dataverse (<https://doi.org/10.7910/DVN/SCLH9W>).

6 | DISCUSSION AND CONCLUSIONS

We have developed a flexible joint dose–response model for estimating a BMD contour for bivariate exposures in quantitative risk assessment. Specifically, we estimated joint BMD points via GAMs, while adjusting for confounding variables through regression adjustment based on propensity scores, and derived the corresponding lower confidence limit using an innovative parametric bootstrap approach based on generating a bootstrap distribution of a functional, then inverting this to obtain the lower limit of interest, which corresponds to the BMDL. Our methodology has provided important insights into the effects of PAE on child cognition and in particular regarding the interacting effects of dose per occasion and frequency of drinking. It also can be used in settings directed at more nuanced questions regarding the timing and intensity of exposure (Kriebel et al., 2007) when such detailed information is available.

An important caveat is that our semiparametric analysis based on generalized additive modeling or GAMs may encounter problems for data sets that have only a small number of control individuals, or those with zero exposure. This is because the method requires having a reliable estimate of the predicted cognitive score for unexposed individuals (see Equation 16). In our application, at least from a technical

perspective (see discussion below), this is not an issue since our data include a large number of nondrinkers or abstainers. In practice, however, it will be a problem for settings involving ubiquitous exposures such as methylmercury (Whitney & Ryan, 2013) where the majority of subjects will have at least low-level exposure levels. In such settings, it might be useful to consider modified approaches such as isotonic regression or monotone splines that impose some mild conditions on the shape of the dose–response for low levels of exposure (Piegorisch et al., 2012).

We demonstrated our approach using data from six large longitudinal cohort studies that aimed to assess the effects of PAE on child cognitive function. In order to assess the performance of the bootstrap methodology, we started with an analysis based on a single exposure variable, the logarithm of average daily alcohol consumption across pregnancy, estimating a BMD based on a linear model. This allowed us to compare BMDL estimates based on our bootstrap approach with the analytic solution described by Budtz-Jørgensen et al. (2001). We found the two BMDL estimates to be very close, suggesting that our bootstrap approach worked well. We then applied the extended methodology to the bivariate exposure setting where alcohol exposure was characterized by both dose per occasion and frequency of drinking. The estimated dose–response surface suggests a nonlinear interaction. Specifically, the dose–response curve appears to be quite flat at low levels of dose per occasion (less than 1.7 oz), for infrequent drinkers. But at higher levels of dose per occasion, the estimated curve starts sloping downward. By contrast, the dose–response curve at the highest levels of drinking frequency (i.e., daily or 30 days per month), the curve slopes sharply downward in a relatively linear manner, suggesting that for frequent drinkers, any increase in dose/occasion is associated with an increase in the adverse effect on cognitive function. This visual interpretation is supported by the BMD analysis. For infrequent drinkers, the estimated BMD is quite high (around 5.7 oz per occasion), falling to lower than 2.0 oz for frequent or daily drinkers. The BMDLs range from around 1.5 oz for infrequent drinkers to just over 0.6 oz for frequent or daily drinkers. These findings suggest that whether there are clinically significant adverse effects on a child's cognitive development depends in many cases on the pattern of the mother's drinking during pregnancy such that there appears to be little effect if the mother drinks <3 drinks/occasion infrequently (i.e., <2 days/month). In this context, it is important to emphasize that there are several critical caveats that need to be kept in mind. First of all, there are always uncertainties associated with epidemiological analysis. Also, a BMD analysis is not intended to identify levels where there is no effect, but rather where risk levels are not excessively raised relative to background. For these reasons, the US Office of the Surgeon General's (2005) recommendation that all women abstain from alcohol during pregnancy continues to be the most appropriate guideline. It also important to emphasize that the present paper has not set out to present a definitive risk analysis for the PAE, but rather to present a methodology for BMD and BMDL computation in the context of a bivari-

ate exposure. A more detailed applied analysis focusing on the interpretation and clinical implications of our analysis is underway. This follow-up paper will provide an opportunity to address in more depth some of the nuances and challenges associated with epidemiological explorations of the health effects of alcohol. For example, some argue (Mugavin et al., 2020) that abstainers can differ from low-level drinkers for a variety of reasons that are not always well captured by standard covariate adjustment.

In our motivating example, we addressed the challenge of modeling an outcome of interest as a function of exposure that can be measured in two dimensions. The methodology can also potentially apply in settings involving mixtures of different exposures. The methodology extends naturally beyond the two-dimensional setting to higher-order exposures, though there may be practical limitations in such cases. For example, tensor product terms with dimensions higher than two may be unstable due to sparse data. There are opportunities for further methodological developments in this area.

ACKNOWLEDGMENTS

This research funded by grants from the National Institutes of Health/National Institute on Alcohol Abuse and Alcoholism, Lycaki Young Fund from the State of Michigan to Sandra W. Jacobson and Joseph L. Jacobson. This work also supported by grants from the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers to Louise M. Ryan (CE140100049), the Natural Sciences and Engineering Research Council of Canada to Richard J. Cook (RGPIN2017-04207), the Canadian Institutes of Health Research to Richard J. Cook and Louise M. Ryan (FRN PJT-180551).

ORCID

Tugba Akkaya Hocagil  <https://orcid.org/0000-0003-3603-4491>

REFERENCES

- Akkaya Hocagil, T., Cook, R. J., Jacobson, S. W., Jacobson, J. L., & Ryan, L. M. (2021). Propensity score analysis for a semi-continuous exposure variable: A study of gestational alcohol exposure and childhood cognition. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(4), 1390–1413.
- Akkaya Hocagil, T., Hwang, H., Jacobson, S. W., Ryan, L. M., & Jacobson, J. L. (2024). Meta-analysis on studies with heterogeneous and partially observed covariates. *JBIC Evidence Synthesis*, 22(3), 413–433.
- Akkaya Hocagil, T., Ryan, L. M., Cook, R. J., Jacobson, S. W., Richardson, G. A., Day, N. L., Coles, C. D., Carmichael Olson, H., & Jacobson, J. L. (2022). A hierarchical meta-analysis for settings involving multiple outcomes across multiple cohorts. *Stat*, 11(1), e462.
- Astley, S., & Clarren, S. (2000). Diagnosing the full spectrum of fetal alcohol-exposed individuals: Introducing the 4-digit diagnostic code. *Alcohol and Alcoholism*, 35(4), 400–410. <https://doi.org/10.1093/alcal/35.4.400>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424.
- Brown, J., Bakeman, R., Coles, C., Sexson, W., & Demi, A. (1998). Maternal drug use during pregnancy: Are preterm and full-term infants affected

- differently? *Developmental Psychology*, 34, 540–554. <https://doi.org/10.1037/0012-1649.34.3.540>
- Budtz-Jørgensen, E., Keiding, N., & Grandjean, P. (2001). Benchmark dose calculation from epidemiological data. *Biometrics*, 57(3), 698–706. <http://www.jstor.org/stable/3068405>
- Chen, J., Gaylor, D., & Kodell, R. (1990). Estimation of the joint risk from multiple-compound exposure based on single-compound experiments. *Risk Analysis*, 10(2), 285–290. <https://doi.org/10.1111/j.1539-6924.1990.tb01049.x>
- Chudley, A., Conry, J., Cook, J. L., Look, C., Rosales, T., & LeBlanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *CMAJ*, 172(suppl 5), S1–S21. <https://doi.org/10.1503/cmaj.1040302>, https://www.cmaj.ca/content/172/5_suppl/S1
- Coles, C., Platzman, K., Raskind-Hood, C., Brown, R., Falek, A., & Smith, I. (2006). A comparison of children affected by prenatal alcohol exposure and attention deficit, hyperactivity disorder. *Alcoholism: Clinical and Experimental Research*, 21, 150–161. <https://doi.org/10.1111/j.1530-0277.1997.tb03743.x>
- Cook, J., Green, C., Lilley, C., Anderson, S., Baldwin, M., Chudley, A., Conry, J., LeBlanc, N., Look, C., Lutke, J., Mallon, B., McFarlane, A., Temple, V., & Rosales, T. (2016). Fetal alcohol spectrum disorder: A guideline for diagnosis across the lifespan. *CMAJ*, 188(3), 191–197. <https://www.cmaj.ca/content/188/3/191.full.pdf>, <https://doi.org/10.1503/cmaj.141593>
- Crump, K. (1984). A new method for determining allowable daily intakes. *Toxicological Sciences*, 4(5), 854–871. <https://doi.org/10.1093/toxsci/4.5.854>
- Crump, K. (1995). Calculation of benchmark doses from continuous data. *Risk Analysis*, 15(1), 79–89.
- Dang, K.-D., Ryan, L. M., Akkaya Hocagil, T., Cook, R. J., Richardson, G. A., Day, N. L., Coles, C. D., Carmichael Olson, H., Jacobson, S. W., & Jacobson, J. L. (2023). Bayesian modelling of effects of prenatal alcohol exposure on child cognition based on data from multiple cohorts. *Australian & New Zealand Journal of Statistics*, 65(3), 167–186.
- Dang, K.-D., Ryan, L. M., Cook, R. J., Akkaya Hocagil, T., Jacobson, S. W., & Jacobson, J. L. (2023). Bayesian outcome selection modeling. *Stat*, 12(1), e568.
- Day, N., Jasperse, D., Richardson, G., Robles, N., Sambamoorthi, U., Taylor, P., Scher, M., Stoffer, D., & Cornelius, M. (1989). Prenatal exposure to alcohol: Effect on infant growth and morphologic characteristics. *Pediatrics*, 84(3), 536–541.
- Deutsch, R., & Piegorsch, W. (2012). Benchmark dose profiles for joint-action quantal data in quantitative risk assessment. *Biometrics*, 68(4), 1313–1322. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2012.01811.x>
- Deutsch, R., & Piegorsch, W. (2013). Benchmark dose profiles for joint-action continuous data in quantitative risk assessment. *Biometrical Journal*, 55(5), 741–754. <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201300037>
- Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association*, 95(452), 1293–1296.
- Erdreich, L. S., & Stara, J. F. (1984). *Approaches to risk assessment for multiple chemical exposures*. U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH.
- European Centre for Disease Prevention and Control. (2011). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2009. *EFSA Journal*, 9(3), 2090. <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2011.2090>
- European Food Safety Authority (EFSA). (2017). Dietary reference values for nutrients. *EFSA Supporting Publications*, 14(12), e15121E. <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2017.e15121>
- General Accounting Office. (2001). *Chemical risk assessment: Selected Federal Agencies' procedures, assumptions, and policies*. General Accounting Office, Washington DC. <https://books.google.ca/books?id=ctinzWEACAAJ>
- Haber, L. T., Dourson, M. L., Allen, B. C., Hertzberg, R. C., Parker, A., Vincent, M. J., Maier, A., & Boobis, A. R. (2018). Benchmark dose (BMD) modeling: Current practice, issues, and challenges. *Critical Reviews in Toxicology*, 48(5), 387–415.
- Harrell, F. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer International Publishing. <https://books.google.ca/books?id=sQ90rgEACAAJ>
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–310. <https://doi.org/10.1214/ss/1177013604>
- Hirano, K., & Imbens, G. (2004). The propensity score with continuous treatments. In *Applied Bayesian modeling and causal inference from incomplete-data perspectives* (pp. 73–84). Wiley.
- Hoyme, H. E. (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 institute of medicine criteria: In reply. *Pediatrics*, 115(6), 1787–1788. <https://doi.org/10.1542/peds.2005-0702>
- Hu, F., & Kalbfleisch, J. D. (2000). The estimating function bootstrap. *Canadian Journal of Statistics*, 28(3), 449–481.
- Huang, R.-y., Pei, L., Liu, Q., Chen, S., Dou, H., Shu, G., Yuan, Z.-x., Lin, J., Peng, G., Zhang, W., & Fu, H. (2019). Isobologram analysis: A comprehensive review of methodology and current research. *Frontiers in Pharmacology*, 10, 1222.
- Jacobson, J., Akkaya Hocagil, T., Jacobson, S., Coles, C., Richardson, G., Olson, H., Day, N., Carter, R., Dodge, N., Dang, K., Cook, R., & Ryan, L. (2024). Effects of prenatal alcohol exposure on cognitive development: A dose-response analysis. *Alcohol: Clinical and Experimental Research*, 45(10), 2040–2058.
- Jacobson, J., Akkaya Hocagil, T., Ryan, L., Dodge, N., Richardson, G., Olson, H., Coles, C., Day, N., Cook, R. J., & Jacobson, S. (2021). Effects of prenatal alcohol exposure on cognitive and behavioral development: Findings from a hierarchical meta-analysis of data from six prospective longitudinal US cohorts. *Alcoholism: Clinical and Experimental Research*, 45(10), 2040–2058.
- Jacobson, J., Jacobson, S., Sokol, R., Martier, S., Ager, J., & Kaplan-Estrin, M. (1993). Teratogenic effects of alcohol on infant development. *Alcoholism: Clinical and Experimental Research*, 17(1), 174–183. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1530-0277.1993.tb00744.x>
- Klaassen, C. (2012). *Casarett and Doull's toxicology: The basic science of poisons* (8th ed.). McGraw-Hill Education. <https://accesspharmacy.mhmedical.com/content.aspx?aid=1100083589>
- Kosuke, I., & van Dyk, D. (2004). Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467), 854–866. <https://doi.org/10.1198/016214504000001187>
- Kriebel, D., Checkoway, H., & Pearce, N. (2007). Exposure and dose modelling in occupational epidemiology. *Occupational and Environmental Medicine*, 64(7), 492–498.
- Li, K., Akkaya Hocagil, T., Cook, R. J., Ryan, L. M., Carter, R. C., Dang, K.-D., Jacobson, J. L., & Jacobson, S. W. (2023). Use of generalized propensity scores for assessing effects of multiple exposures. *Statistics in Biosciences*, 1–30. <https://doi.org/10.1007/s12561-023-09403-8>
- Lin, L., Piegorsch, W., & Bhattacharya, R. (2015). Nonparametric benchmark dose estimation with continuous dose-response data. *Scandinavian Journal of Statistics*, 42(3), 713–731. <http://www.jstor.org/stable/24586891>
- Mela, M., Okpalauwaekwe, U., Anderson, T., Eng, J., Nomani, S., Ahmed, A., & Barr, A. M. (2018). The utility of psychotropic drugs on patients with fetal alcohol spectrum disorder (FASD): A systematic review. *Psychiatry and Clinical Psychopharmacology*, 28(4), 436–445.
- Mugavin, J., MacLean, S., Room, R., & Callinan, S. (2020). Adult low-risk drinkers and abstainers are not the same. *BMC Public Health*, 20(1), 1–7.
- National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention (Internet). (2005). Advisory on alcohol use in pregnancy from the U.S. Surgeon General. Retrieved from <https://www.cdc.gov/ncbddd/fasd/documents/sg-advisory.pdf>
- National Research Council (US) Committee on the Toxicological Effects of Methylmercury. (2000). *Toxicological effects of methylmercury*.

- The National Academies Press. <https://www.nap.edu/catalog/9899/toxicological-effects-of-methylmercury>, <https://doi.org/10.17226/9899>
- OECD. (2006). *Current approaches in the statistical analysis of ecotoxicity data: A guidance to application*. OECD Series on Testing and Assessment No. 54, Paris. <https://www.oecd-ilibrary.org/content/publication/9789264085275-en>, <https://doi.org/10.1787/9789264085275-en>
- OECD. (2014). *Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, Supporting Test Guidelines 451, 452 and 453* (2nd ed.). OECD Series on Testing and Assessment No. 54, Paris. <https://doi.org/10.1787/9789264221475-en>
- Peardon, E., & Elliott, E. J. (2010). Distinguishing between attention-deficit hyperactivity and fetal alcohol spectrum disorders in children: Clinical guidelines. *Neuropsychiatric Disease and Treatment*, 6, 509–515.
- Petrenko, C. L., & Alto, M. E. (2017). Interventions in fetal alcohol spectrum disorders: An international perspective. *European Journal of Medical Genetics*, 60(1), 79–91.
- Piegorsch, W., Xiong, H., Bhattacharya, R., & Lin, L. (2012). Nonparametric estimation of benchmark doses in environmental risk assessment. *Environmetrics*, 23(8), 717–728. <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2175>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Richardson, G., Hamel, S., Goldschmidt, L., & Day, N. (1999). Growth of infants prenatally exposed to cocaine/crack: Comparison of a prenatal care and a no prenatal care sample. *Pediatrics*, 104(2), e18.
- Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, 20(4), 488–495.
- Sprott, D. A. (2008). *Statistical inference in Science*. Springer Science & Business Media.
- Stratton, K., Howe, C., & Battaglia, F. (1996). *Fetal alcohol syndrome: Diagnosis, epidemiology, prevention, and treatment*. National Academia Press. <https://doi.org/10.17226/4991>
- Streissguth, A., Martin, D., Martin, J., & Barr, H. (1981). The Seattle longitudinal prospective study on alcohol and pregnancy. *Neurobehavioral Toxicology and Teratology*, 2(3), 223–233.
- U.S. Environmental Protection Agency. (2000). *Benchmark dose technical guidance document*. https://www.epa.gov/sites/default/files/2015-01/documents/benchmark_dose_guidance.pdf
- U.S. Environmental Protection Agency. (2012). *Benchmark dose technical guidance document*. <https://www.epa.gov/risk/benchmark-dose-technical-guidance>
- Wheeler, M. W., Shao, K., & Bailer, A. J. (2015). Quantile benchmark dose estimation for continuous endpoints. *Environmetrics*, 26(5), 363–372. <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2342>
- Wheeler, M., & Bailer, A. J. (2012). Monotonic Bayesian semiparametric benchmark dose analysis. *Risk Analysis*, 32(7), 1207–1218.
- Whitney, M., & Ryan, L. (2013). Uncertainty due to low-dose extrapolation: Modified BMD methodology for epidemiological data. *Environmetrics*, 24(5), 289–297.
- Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press. <https://books.google.ca/books?id=JTkkDwAAQBAJ>
- Wood, S., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563. <https://doi.org/10.1080/01621459.2016.1180986>
- Zhu, Y. (2005). Dose-time-response modeling of longitudinal measurements for neurotoxicity risk assessment. *Environmetrics*, 16(6), 603–617. <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.725>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Akkaya Hocagil, T., Ryan, L. M., Cook, R. J., Dang, K.-D., Carter, R. C., Richardson, G. A., Day, N. L., Coles, C. D., Carmichael Olson, H., Jacobson, S. W., & Jacobson, J. L. (2024). Benchmark dose profiles for bivariate exposures. *Risk Analysis*, 44, 2415–2428. <https://doi.org/10.1111/risa.14303>