OXFORD

Semi-parametric benchmark dose analysis with monotone additive models

Alex Stringer [1,*], Tugba Akkaya Hocagil[2], Richard J. Cook [1], Louise M. Ryan[3], Sandra W. Jacobson[4], Joseph L. Jacobson[4]

[1]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo N2L 3G1, Canada, [2]Department of Biostatistics, Ankara University, Ankara 06230, Turkey, [3]School of Mathematics and Statistics, University of Technology Sydney, Sydney, NSW 2007, Australia, [4]Department of Psychiatry and Behavioural Neurosciences, Wayne State University School of Medicine, Detroit, MI 48202, United States

*Corresponding author: Alex Stringer, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo N2L 3G1, Canada (alex.stringer@uwaterloo.ca).

ABSTRACT

Benchmark dose analysis aims to estimate the level of exposure to a toxin associated with a clinically significant adverse outcome and quantifies uncertainty using the lower limit of a confidence interval for this level. We develop a novel framework for benchmark dose analysis based on monotone additive dose-response models. We first introduce a flexible approach for fitting monotone additive models via penalized B-splines and Laplace-approximate marginal likelihood. A reflective Newton method is then developed that employs de Boor's algorithm for computing splines and their derivatives for efficient estimation of the benchmark dose. Finally, we develop a novel approach for calculating benchmark dose lower limits based on an approximate pivot for the nonlinear equation solved by the estimated benchmark dose. The favorable properties of this approach compared to the Delta method and a parameteric bootstrap are discussed. We apply the new methods to make inferences about the level of prenatal alcohol exposure associated with clinically significant cognitive defects in children using data from six NIH-funded longitudinal cohort studies. Software to reproduce the results in this paper is available online and makes use of the novel `semibmd` R package, which implements the methods in this paper.

KEYWORDS: additive model; benchmark dose analysis; Laplace approximation; marginal likelihood; monotone smoothing.

1 INTRODUCTION

1.1 Benchmark dose analysis

Benchmark dose analysis is a statistical technique used by environmental toxicologists to quantify the level of exposure to a harmful substance that is associated with an adverse response. The US Environmental Protection Agency (EPA, 2012) and the European Food Safety Authority (EFSA et al., 2022) use the lower limit of a 95% confidence interval of the *benchmark dose*—called the benchmark dose, lower (BMDL)—to set limits on acceptable levels of exposure to a wide variety of toxins. In this paper, we present an innovative general framework for benchmark dose analysis and inference based on monotone additive models. We apply the new methodology to the problem of estimating levels of prenatal alcohol exposure that are associated with clinically significant cognitive defects in children.

Crump (1984; 1995) introduced the concept of a benchmark dose (BMD) and BMDL. Methods were developed first for parametric dose-response modeling of data obtained from designed experiments. Budtz-Jorgensen et al. (2001) proposed use of a dose-response model with a continuous exposure variable arising in epidemiological studies. Such a model can adjust for confounders through regression on covariates or propensity scores. They derive an exact confidence interval for the BMD estimator based on a linear dose-response model with Gaussian errors.

In recent work on parametric dose-response models, Aerts et al. (2020) defined a family of parametric dose-response functions and recommended model-averaged estimates based on a computationally intensive bootstrap procedure.

Parametric models are needed for data with few unique exposure values, such as those often arising from experimental studies. Epidemiological studies often involve a broad range of exposure with many unique values. This type of data supports more flexible non- and semi-parametric modeling to characterize the dose-response function. In epidemiological setups, such models should also enable one to mitigate biases due to the effect of confounding variables. Accordingly, various semi- and non-parametric approaches to benchmark dose estimation have been developed for epidemiological data with continuous and binary outcomes, although they may also be useful for experimental data with sufficiently rich design. Wheeler and Bailer (2012) fit models based on monotone P-splines in a Bayesian framework using MCMC. For binary outcomes, non-parametric methods for dose-response modeling were developed by Piegorsch et al. (2012; 2014) using isotonic regression, again relying on intensive bootstrap calculations for BMDL estimation; methods to deal with continuous responses were developed by Lin et al. (2015). Wheeler et al. (2015) present a non-parametric quantile regression-based method. These methods are all based on

non-parametric estimation of the dose-response relation and associated computationally-intensive inference procedures, which typically require the dose-response model to be fit many times.

1.2 Contributions

We take a semi-parametric approach to benchmark dose analysis, combining the efficiency of the parameteric approaches with the flexibility of the non-parameteric methods. We base our approach on monotone generalized additive models to estimate the benchmark dose and calculate the BMDL; restricting attention to monotone dose-response curves means that in this framework, an increased exposure cannot be associated with a preferred response (Wheeler and Bailer, 2012). We describe a novel procedure for fitting monotone generalized additive models based on B-splines with coefficients parameterized in such a way that they yield estimates of a monotone curve. Laplace-approximate marginal likelihood (Wood, 2011) is applied for selection of the smoothing parameter (Wood et al., 2016). We then develop a fast and stable method for solving the non-linear equation defining the BMD estimate based on de Boor's algorithm for B-splines and their derivatives (de Boor, 2001). This is made feasible by deriving bounds on the estimated BMD and then applying a reflective Newton line search (Coleman and Li, 1994). Finally, we introduce a novel method for BMDL calculation based on the inversion of a hypothesis test based on an approximate pivot derived from the nonlinear equation solved by the estimated BMD. This BMDL is shown to have superior empirical and computational properties to those of the usual Delta method and bootstrap-based BMDLs in the settings considered. Significant computational improvements for the bootstrap are achieved by sampling from an approximate posterior distribution for the parameters, which enables full bootstrap-based BMDL calculation using only a single fit of the dose-response model. The speed of the novel BMD estimation algorithm makes this Bayesian parametric bootstrap approach computationally attractive in typical applications.

1.3 Motivating application

Prenatal alcohol exposure (PAE) has been linked to a broad range of long-term cognitive and behavioral deficits (Jacobson et al., 2024). However, there is relatively little information about the *level* of PAE that is associated with clinically significant cognitive deficits. Attempts to define what constitutes excessive drinking have been both qualitative (Stratton et al., 1996; Hoyme, 2005; Chudley et al., 2005; Cook et al., 2016) and quantitative (Astley and Clarren, 2000; Hoyme et al., 2016), although Astley and Clarren (2000) emphasize that there is no "clear consensus on the amount of alcohol that can actually be toxic to the fetus." Addressing this question has important clinical implications in terms of diagnosing children who have been adversely affected by prenatal alcohol exposure. We apply our method for semi-parameteric benchmark dose analysis to this problem using data from six United States National Institutes of Health-funded longitudinal studies in which expectant women were interviewed regarding their drinking behavior during pregnancy and their children followed through young adulthood and assessed using a variety of cognitive tests.

2 BENCHMARK DOSE ANALYSIS AND MONOTONE SPLINES

2.1 Benchmark dose analysis

Consider a response $Y_i \in \mathbb{R}$ and let $x_i \in \mathbb{R}$ represent the exposure for individual $i = 1, \ldots, n$. We consider the following dose-response model:

$$Y_i = \alpha + f(x_i) + \sum_{j=1}^{m} g_j(z_{ij})$$

$$+ \sigma \epsilon_i, \ \epsilon_i \overset{\text{iid}}{\sim} N(0, 1), \ i = 1, \ldots, n, \quad (1)$$

where α is an intercept, $f(x)$ is a strictly monotone decreasing function of the exposure, and $g_j(z_j)$ are m functions of other covariates, z_j, for $j = 1, \ldots, m$. This specification assumes that an increase in x reduces the mean response, appropriate for settings where smaller values of the response are undesirable, as is the case with cognition scores; see Sections 1.3 and 5. Adaptations to settings where an increase in exposure leads to a greater response (and hence $f(x)$ would be monotone increasing) are straightforward.

Let $p_0 \in (0, 1)$, let $p_+ \in (0, 1 - p_0)$, and let $\tau_0(x_0, z)$ satisfy $\mathbb{P}(Y < \tau_0(x_0, z); x_0, z) = p_0$ where $\mathbb{P}(\cdot; x, z)$ is the probability distribution for Y following the model (1) with covariates $(x, z) \equiv (x, z_1, \ldots, z_m)$. Here $x_0 \in \mathbb{R}$ is a baseline exposure representing an unexposed subject, and is almost always set at $x_0 = 0$ in practice if the data contain unexposed subjects. The benchmark dose (BMD, denoted as x_b) is defined as the value of x satisfying

$$\mathbb{P}(Y < \tau_0(x_0, z); x_b, z) = p_0 + p_+. \quad (2)$$

The value p_+ is called the benchmark response (BMR); see Crump (1995) and Akkaya Hocagil et al. (2024) for further details.

Under the additive model (1), x_b is defined by the nonlinear equation $U(x_b) = 0$, where

$$U(x) = (f(x_0) - f(x))/\sigma - c(p_0, p_+), \quad (3)$$

and $c(p_0, p_+) = \Phi^{-1}(p_0 + p_+) - \Phi^{-1}(p_0)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. It is clear from (3) that defined this way, x_b does not depend on $\tau_0(x_0, z)$ or z. An estimate, $\widehat{x_b}$, of x_b solves the nonlinear equation $U_n(\widehat{x_b}) = 0$, where

$$U_n(x) = (\widehat{f}(x_0) - \widehat{f}(x))/\widehat{\sigma} - c(p_0, p_+), \quad (4)$$

and $\widehat{f}(x)$ and $\widehat{\sigma}$ are estimates of $f(x)$ and σ.

2.2 Quantifying uncertainty in the benchmark dose

To address the uncertainty in estimation of the benchmark dose, it is common to base guidelines on the lower endpoint of an approximate 95% confidence interval for x_b; this is referred to as the benchmark dose-lower, denoted by BMDL. Budtz-Jorgensen et al. (2001) derive the exact sampling distribution of $\widehat{x_b}$ in linear dose-response models and use it to give a formula for a BMDL. More generally, an explicit sampling distribution of $\widehat{x_b}$ is not available, so a common approach is to use the Delta method for variance estimation, and rely on asymptotic normality of $\widehat{x_b}$. Computationally-intensive bootstrap-based BMDL estimation

is also possible by refitting the dose-response model many times; see Moerbeek et al. (2004) for a review of existing methods. In Section 3.4, we derive a novel method for uncertainty quantification of the BMD based on semi-parametric dose-response models and asymptotic normality of the approximate pivot in Equation 4.

2.3 Monotone B-splines

The dose-response function is represented as a monotone spline function. This proves useful for both fitting the dose-response model and rapid computation of the BMD and BMDL, and ensures that a BMD exists for some choice of the BMR; see Section 3. The function $f(x)$ can be written as

$$f(x) = \sum_{l=1}^{L} b_{l,p}(x)\beta_l, \qquad (5)$$

where $b_{l,p}(x)$ is the lth B-spline function of order p on the interval $[a, b] \subset \mathbb{R}$ with knots $\boldsymbol{t} = (t_1, \ldots, t_{L+p})$ such that $a \leq t_1 \leq \cdots \leq t_{L+p} \leq b$. Using cubic B-splines having $p = 4$, we follow the standard practice of choosing a large value of L and controlling over-fitting through penalized estimation; see Wood et al. (2016) and Wood (2017) for details.

B-splines are defined recursively with $b_{l,1}(x) = I([t_l, t_{l+1}))$ and $b_{l,p}(x) = \omega_{l,p-1}b_{l,p-1}(x) + (1 - \omega_{l+1,p-1})b_{l+1,p-1}(x)$ where $\omega_{l,p}(x) = (x - t_l)/(t_{l+p} - t_l)I(t_{l+p} \neq t_l)$. These definitions are used during model fitting (Section 3.2) to compute the design matrix where the x at which the splines are to be evaluated are specified in advance.

The numerical methods of Section 3 (see also Algorithms 2 and 3 in Supplementary Materials Section A) require computation of the spline function (5) within an iterative root-finding procedure at many values of x that cannot be known in advance. In this setting, computing $b_{1,p}(x), \ldots, b_{L,p}(x)$ naively and computing $f(x)$ using (5) is too computationally intensive. For each fixed value of x, the sum defining the value of the spline function $f(x)$ has only $p + 1$ nonzero terms. de Boor's algorithm (de Boor, 2001) computes the spline function $f(x)$ efficiently by ignoring computations known to add zero to the overall sum because the associated basis functions evaluate to zero; see Algorithm 1 in Supplementary Materials Section A. This greatly increases the speed of the computations involved in estimating the BMD and hence makes the parametric bootstrap of Section 3.4 practically feasible.

The derivative of a B-spline basis function is given by

$$b'_{l,p}(x) = (p - 1)(b_{l,p-1}(x)/(t_{l+p-1} - t_l) \\ - b_{l+1,p-1}(x)/(t_{l+p} - t_{l+1})),$$

and the derivative of the spline function (5) is:

$$f'(x) = (p - 1) \sum_{l=2}^{L} b_{l,p-1}(x) \frac{\beta_l - \beta_{l-1}}{t_{l+p-1} - t_l}. \qquad (6)$$

From this, it can be seen that $\beta_1 < \cdots < \beta_L$ is a sufficient condition for $f(x)$ to be strictly monotone decreasing, and this is straightforward to ensure during model fitting via reparameterization as introduced by Pya and Wood (2015); see Section 3.2. Further, because the derivatives of a spline func-

tion are also spline functions (linear combinations of spline basis functions), de Boor's algorithm is also applied to obtain the derivatives required to implement Newton's method, leading to further computational benefits.

3 INFERENCES ABOUT THE BENCHMARK DOSE

3.1 Existence of the benchmark dose

The benchmark dose x_b, defined as the solution to $U(x) = 0$ where $U(x)$ is given by (2), is not guaranteed to exist for every choice of x_0, p_0, p_+. The specification of x_0, p_0, p_+ should be context-dependent; see Haber et al. (2018). If the data involve a large number of unexposed subjects, as is the case in our motivating example on prenatal alcohol exposure, then setting $x_0 = 0$ is reasonable. Otherwise, this is an important modeling choice; see Whitney and Ryan (2013) for a detailed discussion.

Assumption 1 is sufficient to ensure that for some p_0, p_+, there exists a value of x_b satisfying $U(x_b) = 0$:

Assumption 1 *(a) The unknown function $f(x)$ is continuously differentiable with*
$\sup_{x \in \mathbb{R}} |f'(x)| < \infty$ *and* $f'(x) < 0$ *for every $x \in \mathbb{R}$; and (b) with probability 1 the estimated function $\widehat{f}(x)$ is continuously differentiable with $\sup_{x \in \mathbb{R}} |\widehat{f'}(x)| < \infty$ and $\widehat{f'}(x) < 0$ for every $x \in \mathbb{R}$.*

Assumption 1 (a) states that higher exposure cannot be associated with an equal or better expected response. Assumption 1 (b) can be satisfied by applying a monontonic smoother, as we do in Sections 2.3 and 3.2. Since $U(x_0) < 0$ from Assumption 1, by the intermediate value theorem $U(x_{max}) > 0$ is a sufficient condition for the existence of $x_b \in (x_0, x_{max})$ satisfying $U(x_b) = 0$. Further, a sufficient condition for the estimate $\widehat{x_b}$ to exist is $U_n(x_{max}) > 0$. In cases where the underlying dose-response curve is not monotonic or when $U_n(x_{max}) < 0$, a benchmark dose may not exist. This is important to consider in applications.

3.2 Fitting the dose-response model

Under the B-spline representation (5), the dose-response model (1) is:

$$Y_i = \alpha + \sum_{l=1}^{L} b_{l,4}(x_i)\beta_l^c + \sum_{j=1}^{m} \sum_{l=1}^{L} b_{l,4}(z_{ij})\gamma_{jl}$$

$$+ \sigma \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, 1), \quad i = 1, \ldots, n. \qquad (7)$$

Monotonicity can be enforced by reparameterization; see Pya and Wood (2015) for transformations that yield monotone increasing and decreasing functions as well as other shape constraints. We set $\beta_1^c = \beta_1$ and $\beta_l^c = \beta_{l-1}^c - \exp(\beta_l)$, $l = 2, \ldots, L$, which guarantees that $\beta_1^c < \cdots < \beta_L^c$ and hence that $f(x)$ and its estimate, $\widehat{f}(x) = b_{1,4}(x)\widehat{\beta}^c_1 + \cdots + b_{L,4}(x)\widehat{\beta}^c_L$, are monotone decreasing. The unknown parameters to be estimated are $\boldsymbol{\Psi} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma) \in \mathbb{R}^D$ where $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L) \in \mathbb{R}^L$, $\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{mL}) \in \mathbb{R}^{mL}$ and $\sigma > 0$, so that $D = 2 + L(1 + m)$.

Define the design matrices $\boldsymbol{B} = (B_{il})$ with $B_{il} = b_{l,4}(x_i)$ and $\boldsymbol{Z} = [\boldsymbol{Z}_1 : \cdots : \boldsymbol{Z}_m]$ with $\boldsymbol{Z}_j = (Z_{jil})$ and $Z_{jil} = b_{l,4}(z_{ij})$. Further define $\boldsymbol{\phi} = (\tau, \boldsymbol{\lambda}) \in \mathbb{R}^s$ where $s = m + 2$, $\tau = -2\log\sigma$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{m+1})$ where $\lambda_j \in \mathbb{R}$ are (log) smoothing penalty parameters to be estimated. Finally, let $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_{m+1}$ be matrices of integrated squared B-spline second derivatives, computed according to the algorithm of Wood (2017), and let $\boldsymbol{S}_\lambda = \text{diag}\left(e^{\lambda_2}\boldsymbol{S}_2, \ldots, e^{\lambda_{m+1}}\boldsymbol{S}_{m+1}\right)$. A penalized joint (negative) log-likelihood for the unknown parameters, $(\boldsymbol{\Psi}, \boldsymbol{\phi})$, is

$$\ell(\boldsymbol{\Psi}, \boldsymbol{\phi}) = \frac{1}{2}e^\tau \|\boldsymbol{y} - \alpha\boldsymbol{1}_n - \boldsymbol{B}\boldsymbol{\beta}_c - \boldsymbol{Z}\boldsymbol{\gamma}\|_2^2 + e^{\lambda_1}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{S}_1\boldsymbol{\beta}$$
$$+ \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{S}_\lambda\boldsymbol{\gamma}, \tag{8}$$

and a marginal likelihood for $\boldsymbol{\phi}$ is

$$\mathcal{L}(\boldsymbol{\phi}) = \int \exp\left\{-\ell(\boldsymbol{\Psi}, \boldsymbol{\phi})\right\} d\boldsymbol{\Psi}. \tag{9}$$

We define the profile maximum likelihood estimator as $\widehat{\boldsymbol{\Psi}}(\boldsymbol{\phi}) = \text{argmin}_{\boldsymbol{\Psi}}\ell(\boldsymbol{\Psi}, \boldsymbol{\phi})$, and the Hessian matrix $\boldsymbol{H}(\boldsymbol{\phi}) = -\partial_{\boldsymbol{\Psi}}^2\ell(\widehat{\boldsymbol{\Psi}}(\boldsymbol{\phi}), \boldsymbol{\phi})$. Inferences about $\boldsymbol{\phi}$ are to be based on the Laplace-approximate maximum marginal likelihood estimator, $\widehat{\boldsymbol{\phi}} = \text{argmax}_{\boldsymbol{\phi}}\mathcal{L}_{\text{LA}}(\boldsymbol{\phi})$, where

$$\mathcal{L}_{\text{LA}}(\boldsymbol{\phi}) = (2\pi)^{D/2} |\boldsymbol{H}(\boldsymbol{\phi})|^{-1/2} \exp\left\{-\ell(\widehat{\boldsymbol{\Psi}}(\boldsymbol{\phi}), \boldsymbol{\phi})\right\} \tag{10}$$

is a Laplace approximation to $\mathcal{L}(\boldsymbol{\phi})$. Point estimates of $\boldsymbol{\Psi}$ are then given as $\widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\phi}})$, and a point estimate of the unknown dose-response function at any $x \in \mathbb{R}$ is $\widehat{f}(x) = b(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_c$ with $b(x) = (b_{1,4}(x), \ldots, b_{L,4}(x))^{\mathrm{T}}$. Estimates $\widehat{g}_j$ for each g_j, $j = 1, \ldots, m$ are analogously obtained. Uncertainty quantification is discussed in Section 3.4.

Spline calculations are carried out using the Rcpp framework (Eddelbuettel and Francois, 2011). The Template Model Builder (TMB) framework (Kristensen et al., 2016) is used for automatic differentiation and Laplace approximation, yielding efficient computation of $\log\mathcal{L}_{\text{LA}}(\boldsymbol{\phi})$ and $\partial_{\boldsymbol{\phi}}\log\mathcal{L}_{\text{LA}}(\boldsymbol{\phi})$. We compute $\widehat{\boldsymbol{\phi}}$ via quasi-Newton optimization based on these quantities. The linear constraints $\sum_{i=1}^n \widehat{f}(x_i) = 0$ and $\sum_{i=1}^n \widehat{g}(z_{ij}) = 0$, $j = 1, \ldots, m$ are imposed directly, without "constraint-absorbing" reparameterizations (Stringer, 2024). Note that because $U_n(\cdot)$ is invariant to the addition of constant terms to $\widehat{f}$, estimation of x_b is invariant to the choice of linear constraints. However, such constraints are required to ensure the identifiability of α as well as more than one f and g together, so they still must be used.

3.3 Computational method for benchmark dose estimation

Given a fitted dose-response model, we obtain the estimate, $\widehat{x_b}$, defined as the solution to a non-linear equation, $U_n(\widehat{x_b}) = 0$, via a reflective Newton line search. The efficiency and stability of the method are critically important in the parametric bootstrap of Section 3.4, where we repeat the procedure thousands of times to calculate a benchmark dose lower limit. Efficient computation of the required B-spline functions and derivatives to implement the Newton iteration is facilitated by de Boor's algorithm (de Boor, 2001), and a stable iteration is achieved by bounding the solution to a known interval.

Bounds on $\widehat{x_b}$ are obtained by noting that Assumption 1 (b) guarantees that $U_n(\cdot)$ is continuous and strictly monotonic, and further that $U_n(x_0) = -c(p_0, p_+) < 0$. We therefore check the condition that $U_n(x_{\max}) > 0$ in applications. If this constraint is satisfied, then by definition of $\widehat{x_b}$ and the intermediate value theorem, $x_0 < \widehat{x_b} < x_{\max}$. We then augment the typical Newton iteration with the reflective transformation given by Coleman and Li (1994), using $(x_0, x_{\max})$ as the required interval within which the solution is known to lie. The full algorithm is given in Algorithm 2 in Supplementary Materials Section A, which depends further on de Boor's algorithm (Algorithm 1 in Supplementary Materials Section A).

3.4 Computation of benchmark dose lower limits

A method for computing a 95% confidence interval for x_b is required since the lower endpoint of a 95% confidence interval for x_b defines the benchmark dose lower limit (BMDL); see Section 2.2. We introduce a novel BMDL based on an approximate pivot obtained from U_n, and give a fast and stable method for computing it. This computational method draws on and extends the efficient framework for computation of $\widehat{x_b}$ (Section 3.3). Accordingly, we also provide a very efficient Bayesian parametric bootstrap that uses this framework to drastically speed up bootstrap BMDL calculation by avoiding the re-fitting of the dose-response model. We further compare to the more usual but undesirable Delta method based on asymptotic normality of $\widehat{x_b}$ directly. Further details for the bootstrap and Delta method are given in Supplementary Materials Section C.

Inspection of Equation 4 yields the approximate pivot $U_n(x_b)^2/V_n(\widehat{x_b}) \overset{\cdot}{\sim} \chi_1^2$, where

$$V_n(x) = \widehat{\text{Var}}\{U_n(x)\}$$
$$= \frac{1}{\widehat{\sigma}^2}\left(b(x_0) - b(x)\right)^{\mathrm{T}} \Sigma(\widehat{\boldsymbol{\beta}}_c)\left(b(x_0) - b(x)\right),$$

and $\Sigma(\widehat{\boldsymbol{\beta}}_c) = \text{Cov}(\widehat{\boldsymbol{\beta}})$. Denote the entire vector of spline weights and variance/smoothing parameters by $\boldsymbol{\theta} = (\boldsymbol{\Psi}, \boldsymbol{\phi})$, with estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\phi}})$. We follow a similar method to Wood et al. (2016) and draw samples, $\{\boldsymbol{\theta}_j\}_{j=1}^M$ where $M \in \mathbb{N}$, from the approximate posterior $\boldsymbol{\theta}|\boldsymbol{Y} \overset{\cdot}{\sim} \text{N}(\widehat{\boldsymbol{\theta}}, \boldsymbol{H}^{-1}(\widehat{\boldsymbol{\theta}}))$ using the method of Rue (2001). We use the sample covariance matrix of $\boldsymbol{\beta}_c^1, \ldots, \boldsymbol{\beta}_c^M$, $\widehat{\Sigma}(\widehat{\boldsymbol{\beta}}_c) = \text{Cov}(\boldsymbol{\beta}_c^1, \ldots, \boldsymbol{\beta}_c^M)$, to estimate $\Sigma(\widehat{\boldsymbol{\beta}}_c)$. With these quantities available, we define the pivot-based BMDL, $\widehat{x_1}^{\text{piv}}$, as $\widehat{x_1}^{\text{piv}} \in \inf \widehat{\mathcal{C}}_n(\alpha)$, where

$$\widehat{\mathcal{C}}_n(\alpha) = \left\{x \in \mathbb{R} : U_n(x)^2 < V_n(x)\chi_{1,1-\alpha/2}^2\right\}.$$

Computation of $\widehat{x_1}^{\text{piv}}$ is more involved than that of $\widehat{x_b}$. We utilize another application of the reflective Newton line search coupled with de Boor's algorithm as follows: Define the function $\kappa_n(x) = U_n(x)^2 - V_n(x)\chi_{1,1-\alpha/2}^2$ and note that $\widehat{x_1}^{\text{piv}} \in \inf\{x : \kappa_n(x) = 0\}$. Accordingly, we again apply a reflective Newton line search to find an appropriate zero of κ_n. The required derivative is $\kappa_n'(x) = 2U_n(x)U_n'(x) - V_n'(x)\chi_{1,\alpha}^2$ where $V_n'(x) = -2b'(x)^{\mathrm{T}}\Sigma(\widehat{\boldsymbol{\beta}}_c)\left(b(x_0) - b(x)\right)$. Both the B-spline basis function vectors, $b(x)$, and their vectors of derivatives, $b'(x)$, are computed directly using the recursions described in

Section 2.3. Bounds are obtained by observing that $\kappa_n(x_0) = c(p_0, p_+)^2 > 0$ and $\kappa_n(\widehat{x_b}) = -V_n(x)\chi^2_{1,\alpha} < 0$, and applying the intermediate value theorem to conclude that there exists $\widehat{x}_1^{\texttt{piv}} \in (x_0, \widehat{x_b})$. Although such a root may or may not not be unique, we have not observed problems with convergence or stability in our experiments (Section 4) or data analysis (Section 5). The full algorithm is given in Algorithm 3 in Supplementary Materials Section A.

With the sample $\{\boldsymbol{\theta}_j\}_{j=1}^M$ already efficiently drawn from the approximate parameter posterior $\boldsymbol{\theta}|Y \sim \text{N}(\widehat{\boldsymbol{\theta}}, \boldsymbol{H}^{-1}(\widehat{\boldsymbol{\phi}}))$, an efficient an parametric bootstrap BMDL, $\widehat{x}_1^{\texttt{boot}}$, is obtained immediately by repeating the procedure introduced in Section 3.3 for each sampled $\boldsymbol{\theta}_j$. This yields a bootstrap sample $\{(\widehat{x_b})_j\}_{j=1}^M$, and $\widehat{x}_1^{\texttt{boot}}$ is defined as the lower 2.5% quantile of this sample. A BMDL estimated by the Delta method is given by

$$\widehat{x}_1^{\Delta} = \widehat{x_b} - 2\frac{V_n(\widehat{x_b})^{1/2}}{|U_n'(\widehat{x_b})|}.$$

Further details about $\widehat{x}_1^{\texttt{boot}}$ and $\widehat{x}_1^{\Delta}$ are given in Supplementary Materials Section C.

4 EMPIRICAL PERFORMANCE AND COMPUTATIONAL CONSIDERATIONS

We compare via simulations the empirical performance of our proposed BMDL $\widehat{x}_1^{\texttt{piv}}$ with that of the parametric bootstrap $\widehat{x}_1^{\texttt{boot}}$ and the Delta method $\widehat{x}_1^{\Delta}$; see Section 3.4. All computations are performed using the semibmd R package, which implements the methods described in Section 3. We simulate 100 000 replicates from the true dose-response function $f(x) = \exp(-sx)$ for varying $s > 0$. This gives a monotone curve of varying steepness, where smaller s yields a flatter curve and hence a more difficult estimation problem. We report here selected results on empirical coverage and computation times; extended results are available in Supplementary Materials Section B.

We report empirical coverage probabilities and relative computation times for each simulation. Empirical coverage probabilities are defined as the proportion of simulated BMDLs that were less than the true BMD. The expected such proportion is 97.5%, reflecting the definition of the BMDL as the lower limit of an equal-tailed 95% confidence interval for x_b. Monte Carlo confidence intervals for the empirical coverages are also shown, and the number of simulations was chosen to be high enough for these to be very narrow.

Relative computation times are defined as the number of seconds of total computation time required to compute each of $\widehat{x}_1^{\texttt{piv}}$ and $\widehat{x}_1^{\texttt{boot}}$ divided by the number of seconds of total computation time required to compute $\widehat{x}_1^{\Delta}$ for each simulated set of data. Such times are therefore unitless. Monte Carlo confidence intervals are based on the central limit theorem and hence we simply report means and standard errors.

Figure 1 shows the empirical coverage rates of $\widehat{x}_1^{\Delta}$, $\widehat{x}_1^{\texttt{piv}}$, and $\widehat{x}_1^{\texttt{boot}}$. The latter based on 1000 parametric bootstrap replications for each of the 100 000 simulated replicates; this massive amount of computation was made feasible by the speed and stability of the $\widehat{x_b}$ computations described in Sections 3.3 and 3.4.

The performance of $\widehat{x}_1^{\texttt{piv}}$ is satisfactory: coverage is at or just above the nominal level in most cases. In the more difficult cases of flat dose-response function (low s, top 2 rows) or higher residual standard deviation σ (rightmost column), undercoverage is observed at the lower sample sizes $n = 200$ and $n = 500$, but this corrects in all cases for $n = 1000$. The bootstrap $\widehat{x}_1^{\texttt{boot}}$ is sometimes more conservative than $\widehat{x}_1^{\texttt{piv}}$ but otherwise generally agrees with it, at a cost of roughly $30\times$ more computation time (see Table 1). The Delta method $\widehat{x}_1^{\Delta}$ is highly conservative, and often gives coverage of 100% or very near. All methods were too conservative in the case of a flat dose-response curve with higher σ (top right corner).

Table 1 summarizes computational aspects of the procedure. Shown are the (unitless) relative computation times for $\widehat{x}_1^{\texttt{piv}}$ and $\widehat{x}_1^{\texttt{boot}}$ relative to $\widehat{x}_1^{\Delta}$, which should be the fastest. The proposed pivot-based BMDL $\widehat{x}_1^{\texttt{piv}}$ is essentially just as fast as the Delta method with relative computation time only 1% higher; the raw computation times for both are on the order of microseconds on the hardware used. Because the Bayesian parametric bootstrap given in Section 3.4 does not require re-fitting the dose-response model, it is also fast, and with 1000 samples, it takes only around 30 times more computing time than $\widehat{x}_1^{\Delta}$ and $\widehat{x}_1^{\texttt{piv}}$. However, its performance is broadly comparable to that of $\widehat{x}_1^{\texttt{piv}}$. The conclusion is that $\widehat{x}_1^{\texttt{piv}}$ achieves better results than $\widehat{x}_1^{\Delta}$ in the same computational time, and equally satisfactory results to $\widehat{x}_1^{\texttt{boot}}$ in about $30\times$ less computational time.

5 BENCHMARK DOSE ANALYSIS OF PRENATAL ALCOHOL EXPOSURE

We consider data from six longitudinal cohort studies of alcohol consumption by pregnant women conducted in Detroit (Jacobson et al., 1993), Pittsburgh (Day et al., 1991; Richardson et al., 1999), Atlanta (Coles et al., 2006; Brown et al., 1998), and Seattle (Streissguth et al., 1981). The research question of interest is to quantify levels of prenatal alcohol exposure (PAE) associated with the development of clinically important cognitive deficits in children. In these six cohort studies, children were followed from infancy through young adulthood and investigators administered a range of neuropsychological tests to assess IQ and four domains of cognitive function: learning and memory, executive function, and academic achievement in reading and mathematics. To obtain an overall cognitive function score for each child, a structural equation model was fitted for each cohort. We then used the estimated cognitive function score as the outcome measure in the below analysis. For details on this approach, see Jacobson et al. (2024) and Akkaya Hocagil et al. (2024).

Data on maternal alcohol consumption were summarized in terms of average alcohol intake per day (ounces of absolute alcohol, (AA)/day) during pregnancy, and data on a broad range of potential confounders were collected in these cohorts. Since each cohort provided a somewhat different set of confounding variables, for these data Jacobson et al. (2024) modeled the exposure variable, average alcohol intake per day, as a function of the potential confounders via a linear regression model and use this to estimate a propensity score (Rosenbaum and Rubin, 1983). We used these propensity scores as covariates
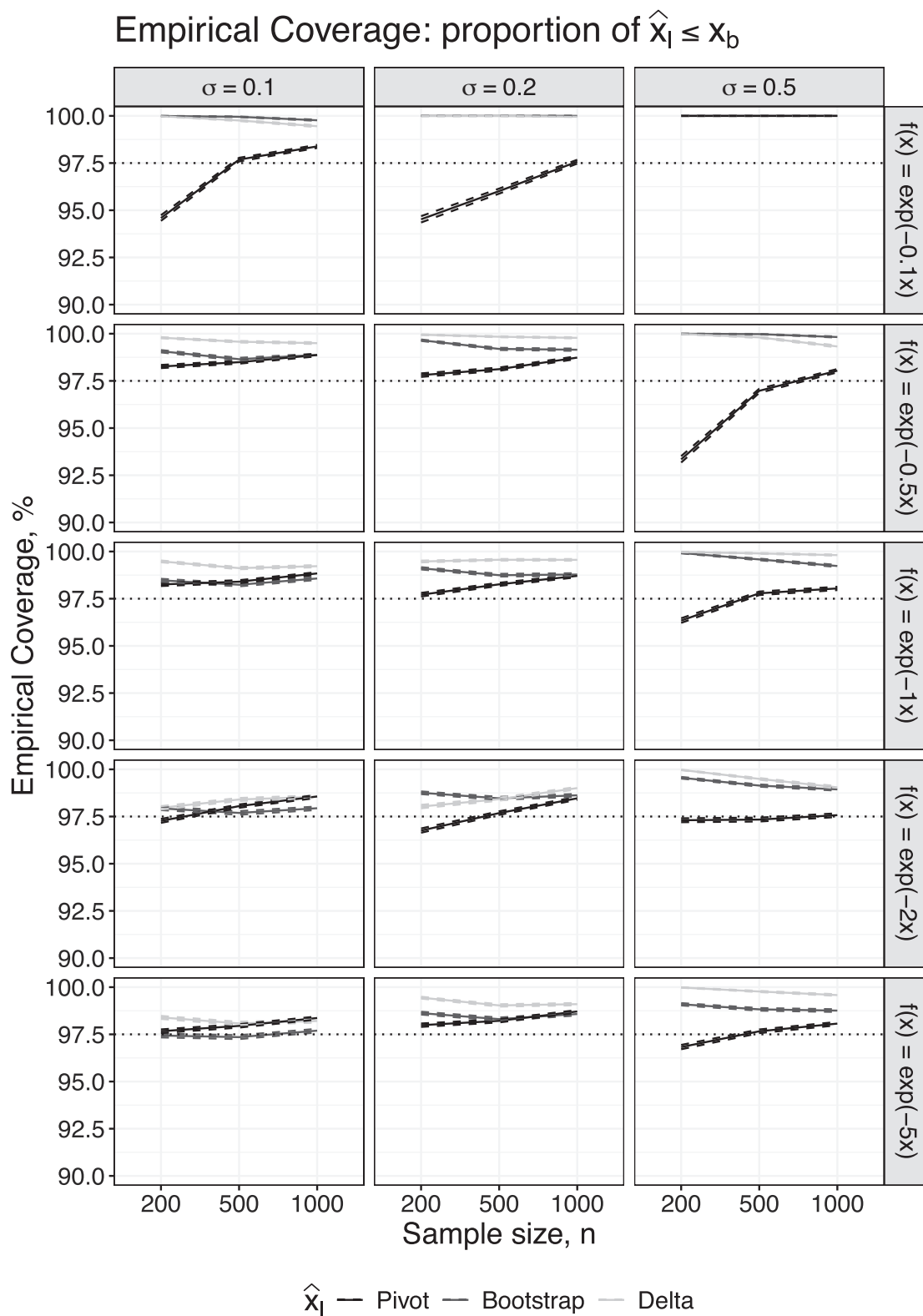
FIGURE 1 Empirical coverage proportions, $\widehat{x}_1 \leq x_{\mathrm{b}}$, from 100 000 simulations described in Section 4.

TABLE 1 Average and empirical standard error of computation time for $\widehat{x}_1^{\texttt{piv}}$, $\widehat{x}_1^{\texttt{boot}}$ relative to $\widehat{x}_1^{\Delta}$. Numbers shown are average and (standard error) of the computational time in seconds of $\widehat{x}_1^{\texttt{piv}}$ or $\widehat{x}_1^{\texttt{boot}}$ divided by that of $\widehat{x}_1^{\Delta}$, and are hence unitless relative computation times.

		$n = 200$		$n = 500$		$n = 1000$	
s	σ	$\widehat{x}_1^{\texttt{piv}}$	$\widehat{x}_1^{\texttt{boot}}$	$\widehat{x}_1^{\texttt{piv}}$	$\widehat{x}_1^{\texttt{boot}}$	$\widehat{x}_1^{\texttt{piv}}$	$\widehat{x}_1^{\texttt{boot}}$
.1	.1	1.02(0.00)	33.80(0.05)	1.01(0.00)	33.47(0.05)	1.01(0.00)	33.34(0.05)
	.2	1.02(0.00)	34.60(0.06)	1.02(0.00)	33.98(0.06)	1.02(0.00)	33.71(0.05)
	.5	1.02(0.00)	35.50(0.07)	1.02(0.00)	35.58(0.06)	1.02(0.00)	35.78(0.06)
.5	.1	1.01(0.00)	33.43(0.05)	1.01(0.00)	33.29(0.05)	1.01(0.00)	33.43(0.05)
	.2	1.01(0.00)	33.41(0.05)	1.01(0.00)	33.31(0.05)	1.01(0.00)	33.35(0.05)
	.5	1.02(0.00)	33.90(0.05)	1.01(0.00)	33.55(0.05)	1.01(0.00)	33.44(0.05)
1	.1	1.01(0.00)	33.54(0.05)	1.01(0.00)	33.38(0.05)	1.01(0.00)	33.39(0.05)
	.2	1.01(0.00)	33.44(0.05)	1.01(0.00)	33.36(0.05)	1.01(0.00)	33.22(0.05)
	.5	1.01(0.00)	33.52(0.04)	1.01(0.00)	33.46(0.05)	1.01(0.00)	33.48(0.05)
2	.1	1.01(0.00)	33.71(0.05)	1.01(0.00)	33.53(0.05)	1.01(0.00)	33.42(0.05)
	.2	1.01(0.00)	33.76(0.05)	1.01(0.00)	33.46(0.05)	1.01(0.00)	33.70(0.05)
	.5	1.01(0.00)	33.66(0.05)	1.01(0.00)	33.69(0.05)	1.01(0.00)	33.51(0.05)
5	.1	1.01(0.00)	33.91(0.05)	1.01(0.00)	33.79(0.05)	1.01(0.00)	34.01(0.05)
	.2	1.01(0.00)	34.14(0.05)	1.01(0.00)	33.77(0.05)	1.02(0.01)	33.88(0.05)
	.5	1.01(0.00)	34.07(0.05)	1.01(0.00)	33.97(0.05)	1.01(0.00)	34.03(0.05)

(Rosenbaum and Rubin, 1983; Imbens and Hirano, 2004) in the following dose-response model.

Let Y_i represent the cognitive function score, $x_i = \log(a_i + 1)$ where a_i is the average alcohol intake per day during pregnancy, $j_i \in \{1, \ldots, 6\}$ index the cohort to which subject i belongs, and $z_i = s_{j_i}$ denote the computed propensity scores for subject $i = 1, \ldots, 2226$ in cohort j_i. A log-transformation was applied to a_i to reduce the influence of very high exposure values on the fitted model. The dose-response model is then

$$Y_i = \alpha + f(x_i) + g_{j_i}(z_i) + \sigma \epsilon_i, \ \epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, 1), \quad (11)$$

where f is an unknown, smooth monotone dose-response function, and g_j, $j = 1, \ldots, 6$ are unknown smooth functions. We apply the basis expansions of Section 2.3 and fit the model using the methods described in Section 3.2. We compute the estimated BMD, $\widehat{x}_{\text{b}}$, using the method of Section 3.3, and the BMDL, $\widehat{x}_1^{\texttt{piv}}$, using the methods of Section 3.4. We also compute $\widehat{x}_1^{\Delta}$ and $\widehat{x}_1^{\texttt{boot}}$ for comparison. We used $p_+ = .01$ and $p_0 = .025$; see Akkaya Hocagil et al. (2024) for the rationale for these choices.

Figure 2 represents the estimated dose-response curve obtained from the fitted monotone additive model along with the estimated BMD and corresponding BMDLs; see Figure 1 in Supplementary Materials Section D for model diagnostics. The estimated BMD was $\widehat{x}_{\text{b}} = 0.764$ and the three BMDLs were $\widehat{x}_1^{\texttt{piv}} = 0.356$, $\widehat{x}_1^{\texttt{boot}} = 0.310$, and $\widehat{x}_1^{\Delta} = 0.175$. The Delta method estimate, $\widehat{x}_1^{\Delta}$, is much closer to zero than the other two, and hence communicates a much more conservative clinical recommendation; combined with its empirical coverage being too high in simulations, we remark that this can be regarded as a disadvantage of the Delta method over the two proposed BMDLs.

The total amount of time that the procedure took to fit the model, estimate the BMD, and compute the three BMDLs based on 100 000 bootstrap iterations was 396 seconds on a 2021 M1 Macbook Pro with 64 GB of RAM. Model fitting and estimating the BMD took 3.93 seconds and 30 microseconds, respectively. Estimating the covariance matrix via posterior sampling and computing the pivot BMDL took 1.18 and 0.4 seconds, re-

spectively. The computation time for 100 000 iterations of the parametric bootstrap for $\widehat{x}_{\text{b}}$ was the longest at 390 seconds, and produced a result close to $\widehat{x}_1^{\texttt{piv}}$ which was computed nearly 1000× faster. Fitting the model, estimating the BMD, and computing the pivot-based BMDL $\widehat{x}_1^{\texttt{piv}}$ took a total of 5.53 seconds.

6 DISCUSSION

The method described in Section 3.2 for fitting the dose-response model can be used to fit monotone generalized additive models in which the response is binary, count, categorical, and generally has any smooth density; see Pya and Wood (2015). The quadratic term in Equation 8 would simply be replaced by the negative log-likelihood, and then the Laplace approximation described in Section 3.2 applies unchanged; see Wood et al. (2016), Kristensen et al. (2016), and Brooks et al. (2017). For continuous data, Wood et al. (2016) show how this approach can also be used to relax the assumption of constant variance in additive models with Gaussian responses. Our proposed approach for fitting dose-response models is easily adapted to other settings.

Extension of the present methods for BMD(L) determination for continuous outcomes to binary, count, and categorical outcomes is complicated only by the BMD itself having different definitions in each of these types of data. For example, for a binary outcome Y_i, Crump (1995, Equation (1)) defines the BMD as the excess risk; if $P(Y_i = 1; x) = \mu(x)$ then the BMD x_{b} is defined by $\mu(x_{\text{b}}) - p_0 = p_+$. If our approach is applied to fit the logistic dose-response model $\log[\mu(x)/\{1 - \mu(x)\}] = f(x) + g_1(z_1) + \cdots + g_d(z_d)$, then the reflective Newton approach of Section 3.3 could be applied to solve $\widehat{\mu}(x_{\text{b}}) - p_0 = p_+$. A pivot-based BMDL may also be possible to derive. These questions are interesting for future work; one challenge to overcome is that the BMD will depend on the values of any other covariates in the model, complicating analysis and reporting.

We recommend avoiding the Delta method ($\widehat{x}_1^{\Delta}$; Section 3.4) for forming confidence intervals for $\widehat{x}_{\text{b}}$ in general. The estimate

(a) Estimated dose-response curve, $\widehat{f}$.

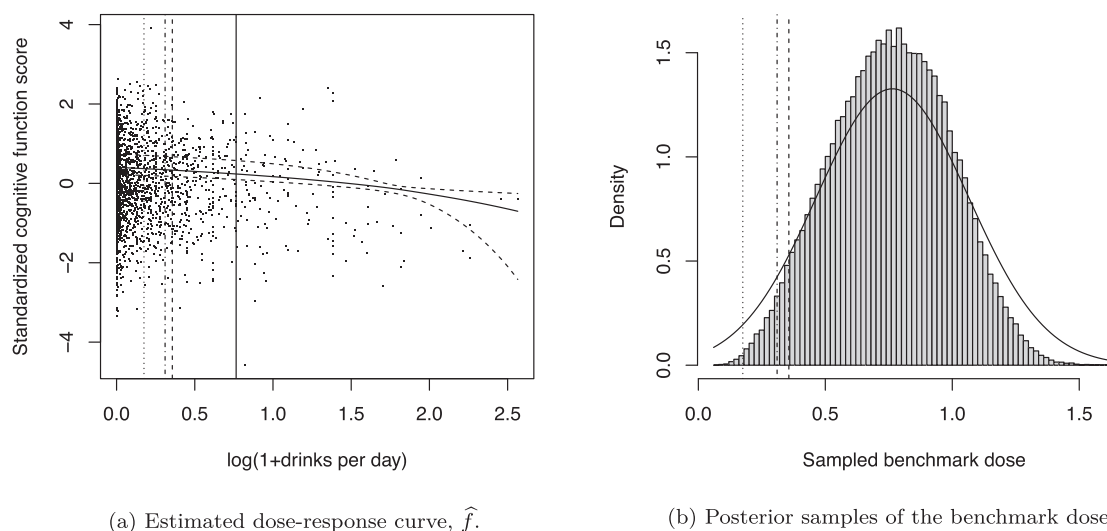(b) Posterior samples of the benchmark dose.

FIGURE 2 (a) Estimated dose-response curve (—) and uncertainty bands computed as pointwise 95% credible intervals from 100 000 samples from the posterior of the fitted curve (- - -). Vertical lines are the estimated BMD, $\widehat{x}_b$ (—) and the BMDLs: $\widehat{x}_1^{\Delta}$ (· · ·), $\widehat{x}_1^{\text{piv}}$ (- - -), and $\widehat{x}_1^{\text{boot}}$ (— · —). (b) 100 000 samples from the approximate posterior distribution of $\widehat{x}_b$ (■) and normal approximation to this distribution (—) upon which $\widehat{x}_1^{\Delta}$ is based.

$\widehat{x}_b$ is a highly non-linear function of the MLE, and should not be expected to have a sampling distribution that is close to Gaussian. An example of this is shown in Figure 2(b) for the PAE data analysis, in which the approximate posterior distribution of $\widehat{x}_b$ is extremely non-normal, leading to a substantial overestimation of uncertainty and a low value of $\widehat{x}_1^{\Delta}$. This recommendation echoes that of Moerbeek et al. (2004).

A substantial motivation for using semi-parametric approaches is to avoid specifying the form of the dose-response function. In designed experiments, there may be scientific reasons to favor a specific parametric model, but this is less often the case in epidemiological studies. Regarding the motivating study of the effects of PAE on child cognition, Jacobson et al. (2024) argue that the dose-response curve is plausibly non-linear, and its shape should be flat at low levels of exposure and then steeper beyond a certain level. Such situations in which the shape of the dose-response curve is partially specified are particularly amenable to analysis using the semi-parametric methods we develop in this paper.

The estimated dose-response curve shown in Figure 2(a) for the PAE data analysis is very flat at low exposure values. This appears to be due to a large amount of noise relative to signal in the PAE data themselves, suggesting that a single exposure measure may not be sufficiently informative (Jacobson et al., 2024). Akkaya Hocagil et al. (2024) perform a benchmark dose analysis for PAE using a bivariate exposure that incorporates both frequency and severity of drinking. In future work, we plan to develop an efficient computational method for the semi-parametric bivariate exposure case, making use of the fast computational methods developed in the present manuscript to improve the practical application of benchmark dose estimation with bivariate exposure.

Although the present paper presents methodology that is specifically tied to computations related to benchmark dose estimation and its associated lower confidence limit, the methods could be applied to more general inverse estimation problems. In such problems, the object of interest is a point on the *x*-axis in a semi-parametric regression model, and is hence obtained by solving a nonlinear equation. Inverse problems arise in a wide range of applied settings, for example, chemical calibration or medical imaging, and it would be interesting to apply similar approaches to those used in the present paper in other such problems.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices, Tables, and Figures referenced in Sections 3–5 are available with this paper at the Biometrics website on Oxford Academic. Code to reproduce the results in this paper is also available on the Biometrics website and at GitHub at https://github.com/awstringer1/bmd-paper-code.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

Data analyzed in Section 5 is available from the Harvard Data-Verse (Akkaya Hocagil, 2024) at https://doi.org/10.7910/DVN/SCLH9W.

REFERENCES

Aerts, M., Wheeler, M. W. and Abrahantes, J. C. (2020). An extended and unified modeling framework for benchmark dose estimation for both continuous and binary data. *Environmetrics*, 31, e2630.

Akkaya Hocagil, T. (2024). Data for: Benchmark dose profiles for bivariate exposures.

Akkaya Hocagil, T., Ryan, L. M., Cook, R. J., Dang, K.-D., Carter, R. C., Richardson, G. A. et al. (2024). Benchmark dose profiles for bivariate exposures. *Risk Analysis*, Online ahead of print, doi:10.1111/risa.14303.

Astley, S. J. and Clarren, S. K. (2000). Diagnosing the full spectrum of fetal alcohol-exposed individuals: introducing the 4- digit diagnostic code. *Alcohol and Alcoholism*, 35, 400–410.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A. et al. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378–400.

Brown, J. V., Bakeman, R., Coles, C. D., Sexson, W. R. and Demi, A. S. (1998). Maternal drug use during pregnancy: are preterm and full-term infants affected differently? *Developmental Psychology*, 34, 540–554.

Budtz-Jorgensen, E., Keiding, N. and Grandjean, P. (2001). Benchmark dose calculation from epidemiological data. *Biometrics*, 57, 698–706.

Chudley, A. E., Conry, J., Cook, J., Loock, C., Rosales, T. and LeBlanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *Canadian Medical Association Journal*, 172, S1–S21.

Coleman, T. F. and Li, Y. (1994). On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Mathematical Programming*, 67, 189–224.

Coles, C., Platzman, K., Raskind-Hood, C., Brown, R., Falek, A. and 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.

Cook, J. L., Green, C., Lilley, C., Anderson, S., Baldwin, M, Chudley, A. et al. (2016). Fetal alcohol spectrum disorder: a guideline for diagnosis across the lifespan. *Canadian Medical Association Journal*, 188, 191–197.

Crump, K. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology*, 4, 854–871.

Crump, K. S. (1995). Calculation of benchmark doses from continuous data. *Risk Analysis*, 15, 79–89.

Day, N., Sambamoorthi, U., Taylor, P., Richardson, G., Robles, N., Jhon, Y. et al. (1991). Prenatal marijuana use and neonatal outcome. *Neurotoxicology and Teratology*, 13, 329–334.

de Boor, C. (2001). *A Practical Guide to Splines (Applied Mathematical Sciences)*. revised edition. New York: Springer-Verlag.

Eddelbuettel, D. and Francois, R. (2011). Rcpp: seamless R and C++ integration. *Journal of Statistical Software*, 40, 1–18.

EFSA, S. C., More, S. J., Bampidis, V., Benford, D., Bragard, C., Halldorsson, T. I. et al. (2022). Guidance on the use of the benchmark dose approach in risk assessment. *EFSA Journal*, 20, e07584.

Haber, L. T., Dourson, M. L., Allen, B. C., Hertzberg, R. C., Parker, A., Vincent, M. J. et al. (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. *Critical Reviews in Toxicology*, 48, 387–415.

Hoyme, H., Kalberg, W., Elliot, A., Blankenship, J., Buckley, D., Marais, A. et al. (2016). Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. *Pediatrics*, 138, e20154256–e20154256.

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, 1787–1788.

Imbens, G. and Hirano, K. (2004). The propensity score with continuous treatments.

Jacobson, J., Akkaya Hacagil, T., Jacobson, S., Coles, C., Richardson, G., Olson, H. et al. (2024). Effects of prenatal alcohol exposure on cognitive development: a dose-response analysis. *Alcohol Clinical and Experimental Research*, 48, 623–639.

Jacobson, J. L., Jacobson, S., Sokol, R., Martier, S., Ager, J. and Kaplan-Estrin, M. (1993). Teratogenic effects of alcohol on infant development. *Alcoholism Clinical and Experimental Research*, 17, 174–183.

Kristensen, K., Nielson, A., Berg, C. W., Skaug, H. and Bell, B. M. (2016). TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70.

Lin, L., Piegorsch, W. and Bhattacharya, R. N. (2015). Nonparametric benchmark dose estimation with continuous dose-response data. *Scandinavian Journal of Statistics*, 42, 713–731.

Moerbeek, M., Piersma, A. H. and Slob, W. (2004). A comparison of three methods for calculating confidence intervals for the benchmark dose. *Risk Analysis: An International Journal*, 24, 31–40.

Piegorsch, W., Xiong, H., Bhattacharya, R. N. and Lin, L. (2012). Nonparametric estimation of benchmark doses in environmental risk assessment. *Environmetrics*, 23, 717–728.

Piegorsch, W. W., Xiong, H., Bhattacharya, R. N. and Lin, L. (2014). Benchmark dose analysis via nonparametric regression modeling. *Risk Analysis*, 34, 135–151.

Pya, N. and Wood, S. N. (2015). Shape constrained additive models. *Statistics and Computing*, 25, 543–559.

Richardson, G., Hamel, S., Goldschmidt, L. and Day, N. (1999). Maternal drug use during pregnancy: are preterm and full-term infants affected differently? *Pediatrics*, 104.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 63, 325–338.

Stratton, K., Howe, C. and Battaglia, F. C. (1996). *Fetal Alcohol Syndrome: Diagnosis, Epidemiology, Prevention, and Treatment*. Washington, DC: The National Academies Press.

Streissguth, A., Martin, D., Martin, J. and Barr, H. (1981). The Seattle Longitudinal Prospective Study on alcohol and pregnancy. *Neurobehav Toxicol Teratol*, 2, 223–233.

Stringer, A. (2024). Identifiability constraints in generalized additive models. *Canadian Journal of Statistics*, 52, 461–476.

United States (2012). *Benchmark dose technical guidance*.

Wheeler, M. and Bailer, A. J. (2012). Monotonic Bayesian semiparametric benchmark dose analysis. *Risk Analysis: An International Journal*, 32, 1207–1218.

Wheeler, M., Shao, K. and Bailer, A. J. (2015). Quantile benchmark dose estimation for continuous endpoints. *Environmetrics*, 26, 363–372.

Whitney, M. and Ryan, L. (2013). Uncertainty due to low-dose extrapolation: modified BMD methodology for epidemiological data. *Environmetrics*, 24, 289–297.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 73, 3–36.

Wood, S. (2017). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, 27, 985–989.

Wood, S., Pya, N. and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111, 1548–1575.