Computational Statistics and Data Analysis 158 (2021) 107189

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Robust designs for dose–response studies: Model and labelling robustness

Douglas P. Wiens[1]

Mathematical & Statistical Sciences, University of Alberta, Edmonton, Canada T6G 2G1

ARTICLE INFO

Article history:
Received 14 June 2020
Received in revised form 23 January 2021
Accepted 25 January 2021
Available online 5 February 2021

Keywords:
Adaptive
Logistic
Menarche
Minimax
Probit
Sequential

ABSTRACT

Methods for the construction of dose–response designs are presented that are robust against possible model misspecifications and mislabelled responses. The asymptotic properties are studied, leading to asymptotically minimax designs that minimize the maximum – over neighbourhoods of both types of model inadequacies – value of the mean squared error of the predictions. Both sequential and adaptive approaches are studied. Finite sample simulations and examples illustrate the gains to be made by adaptivity.

1. Introduction and summary

The literature on robustness of design in dose–response studies, or, more broadly, for generalized linear models, remains fairly meagre. Khuri et al. (2006) gave a comprehensive survey of design issues, considering robustness mostly against a poor choice of initial parameters, or of the prior in Bayesian design. For binary models see Abdelbasit and Plackett (1983). Dette et al. (2008) noted the sensitivity of locally optimal dose–response designs to perturbations of the local parameters, and to the chosen model for the continuous response. This avenue of investigation extended Biedermann et al. (2006) and was continued by Bretz et al. (2010). Adewale and Xu (2009) obtained generalized linear model designs that are robust against, among others, parametric departures from the fitted link. Li and Wiens (2011) designed for robustness against an infinite dimensional, nonparametric neighbourhood of such departures from the fitted link. Woods et al. (2006) proposed to optimize a loss function after averaging over a finite set of competing models, differing with respect to, for instance, the assumed parameter values, or the assumed link.

For some recent work in this area see as well Huang and Chen (2021), Holland-Letz and Kopp-Schneider (2015), Feller et al. (2017) and Xu and Sinha (2021).

Perhaps closest to the approach to robustness advocated in this article is Adewale and Wiens (2009), who adopted a logistic link, allowing for errors in the specification of the linear component. These model errors were allowed to range over a certain – rather sparse – neighbourhood, and the – locally optimal – design was chosen to minimize the average loss over this neighbourhood. In contrast, in the current work we allow an arbitrary link function, a much richer class of model alternatives, allow for labelling errors, and derive *sequential* and *adaptive*, *minimax* designs. A rough description of our sequential approach is that, given an n-point design, the experimenter will choose a further design point in order to minimize a first order approximation to the current maximum loss, maximized over the neighbourhood of departures

E-mail address: doug.wiens@ualberta.ca.

[1] Supplementary material is at www.ualberta.ca/~dwiens/.

https://doi.org/10.1016/j.csda.2021.107189
0167-9473/

from the fitted model and over the labelling errors. This process is then repeated, as necessary. In the adaptive approach the responses are learned, and the parameters which appear in the loss function are re-estimated, at each stage.

The mathematical framework, and asymptotic properties, of our methods are outlined in the next two sections of this article. Methods of design construction are given in Section 4, followed by a simulation study in Section 5.1. This study compares various estimation, weighting, and design schemes, and illustrates the gains to be made by our adaptive, robust designs over sequential but non-adaptive methods, and over random designs.

Our work in this article is partially motivated by the following study, to which we shall return in Section 5.2. Milicer and Szczotka (1966) investigated the age of menarche (onset of menstrual bleeding) in a large number of Warsaw girls. Here 'age' plays the role of dose and the response is $Y(x_0) = 1$ or 0 as the respondent reported that menarche had or had not occurred, when asked at age x_0. Lundblad and Jacobsen (2017) found that the age of menarche can be misreported, which we interpret as a mislabelled response, up to 5% of the time. Mohamad et al. (2013) found age of menarche to be highly, and negatively, correlated with Body Mass Index (BMI), which in turn (Jacobs, 2019) is negatively correlated with smoking patterns. Müller and Schmitt (1990) revisited the Warsaw study, and constructed designs to estimate the median age of menarche from a linear logistic fit relating $P(Y(x_0) = 1)$ to x_0, with no covariates. In Section 5.2 we fit a model relating age to BMI and smoking status, with interactions, allowing for model misspecification and response mislabelling.

2. Mathematical framework

To set the framework we consider a design space $\chi = \{\boldsymbol{x}_i\}_{i=1}^N \subset \mathbb{R}^{k+1}$ in which $\boldsymbol{x} = \left(x_0, \boldsymbol{x}'_{(1)}\right)'$, where x_0 is the 'dose' and $\boldsymbol{x}_{(1)}$ is a k-dimensional vector of covariates. At chosen dose/covariates $\boldsymbol{x}$ the experimenter observes a binary random variable Y. We envisage a framework in which this investigator assumes, and fits, the possibly incorrect model

$$P(Y = 1|\boldsymbol{x}) = G(\beta(\boldsymbol{x}; \boldsymbol{\theta}_0)), \tag{1}$$

for a strictly increasing, absolutely continuous distribution function $G(\cdot)$ with a density $g(\cdot)$ possessing two bounded derivatives. We take

$$\beta(\boldsymbol{x}; \boldsymbol{\theta}) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}, \tag{2}$$

for regressors $\boldsymbol{f}(\boldsymbol{x})$, and an unknown, d-dimensional parameter $\boldsymbol{\theta}$ ranging over a compact, convex space Θ.

Our focus is on efficient and robust estimation and prediction of $P(Y = 1|\boldsymbol{x})$, and of the conditional quantile function, i.e. the dose $x_0 = Q_p\left(\boldsymbol{x}_{(1)}\right)$ defined by

$$P\left(Y = 1|Q_p\left(\boldsymbol{x}_{(1)}\right), \boldsymbol{x}_{(1)}\right) = p,$$

for $p \in (0, 1)$. To quantify the robustness, first define

$$\beta_*\left(\boldsymbol{x}; \boldsymbol{\theta}_\gamma\right) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}_\gamma + \psi(\boldsymbol{x}), \tag{3}$$

where $\boldsymbol{\theta}_\gamma$ is the 'true' parameter under the mislabelling model described below, and $\psi(\boldsymbol{x})$ represents unknown model error constrained by

$$\sum_{i=1}^{N} \psi^2(\boldsymbol{x}_i) \leq \tau_1^2/n, \tag{4}$$

for a constant τ_1 controlling the magnitude of the error in the incorrectly fitted model (2). Define as well

$$H_\gamma(x) = (1-\gamma)G(x) + \gamma\bar{G}(x),$$

where $\bar{G} = 1 - G$ and γ is a mislabelling probability — see Copas (1988) and Carroll and Pederson (1993). We assume that

$$\gamma \leq \min\left(\tau_2/\sqrt{n}, .5\right), \tag{5}$$

for a user-specified τ_2. Our asymptotic statements will then always assume that $n > 4\tau_2^2$. Conditions (4) and (5) are necessary for a sensible asymptotic treatment; the dependence on n is moot if n is fixed. Knowledge of τ_1 and τ_2 is not needed for the construction of the designs — see Remark 6.

We suppose that in fact $P(Y = 1|\boldsymbol{x})$ is given by

$$P(1|\boldsymbol{x}) = H_\gamma\left(\beta_*\left(\boldsymbol{x}; \boldsymbol{\theta}_\gamma\right)\right), \tag{6}$$

so that in fitting (1) the experimenter, possibly incorrectly, takes $\gamma = 0$, $\psi \equiv 0$.

Remark 1. It is tempting to allow γ to depend on $\boldsymbol{x}$. See the remark following the proof of Theorem 2 to see the difficulties to which this leads.

Remark 2. Although $\left\{H_\gamma\right\}$ is a class of link functions, we use it only to model mislabelling, not errors in the link specification. For the latter form of robustness see Li and Wiens (2011) and references therein.

Neither $\boldsymbol{\theta}_\gamma$ nor ψ is completely defined by (3) — for instance we can replace $\boldsymbol{\theta}_\gamma$ by $\boldsymbol{\theta}_\gamma + \boldsymbol{\theta}_*$, and $\psi(\boldsymbol{x})$ by $\psi(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}_*$. We address this by *defining* $\boldsymbol{\theta}_\gamma$ by

$$\boldsymbol{\theta}_\gamma = \arg\min_{\boldsymbol{t}} \sum_{i=1}^{N} \left(H_\gamma^{-1}\{P(Y=1|\boldsymbol{x}_i)\} - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{t}\right)^2, \tag{7}$$

whence ψ in (3) may be *defined* by $\psi(\boldsymbol{x}) = H_\gamma^{-1}\{P(Y=1|\boldsymbol{x})\} - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}_\gamma$. In (7), $P(Y=1|\boldsymbol{x}_i)$ is the true, unknown probability, not necessarily of any given parametric form.

A consequence of the minimization is the condition

$$\sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i)\psi(\boldsymbol{x}_i) = \boldsymbol{0}. \tag{8}$$

Let $\boldsymbol{F}$ be the $N \times d$ matrix with rows $\{\boldsymbol{f}'(\boldsymbol{x}_i)\}_{i=1}^{N}$, assumed to have full column rank. Then the solution to the least squares problem (7) is

$$\boldsymbol{\theta}_\gamma = \left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\left(\ldots, H_\gamma^{-1}\{P(1|\boldsymbol{x}_i)\}, \ldots\right)'.$$

The parameter of interest assumes no mislabelling, and so is

$$\boldsymbol{\theta}_0 = \left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\left(\ldots, G^{-1}\{P(1|\boldsymbol{x}_i)\}, \ldots\right)', \tag{9}$$

whence

$$\boldsymbol{\theta}_\gamma - \boldsymbol{\theta}_0 = \left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\left(\ldots, \left(H_\gamma^{-1}\{P(1|\boldsymbol{x}_i)\} - G^{-1}\{P(1|\boldsymbol{x}_i)\}\right), \ldots\right)'.$$

From the definition of H_γ we obtain the expansion

$$H_\gamma^{-1}(t) = G^{-1}\left(\frac{t-\gamma}{1-2\gamma}\right) = G^{-1}(t) + \frac{\gamma(2t-1)}{g\left(G^{-1}(t)\right)} + O\left(\gamma^2\right),$$

and then substituting $t = P(1|\boldsymbol{x}) = H_\gamma\left(\beta_*\left(\boldsymbol{x};\boldsymbol{\theta}_\gamma\right)\right) = G\left(\beta\left(\boldsymbol{x};\boldsymbol{\theta}_\gamma\right)\right) + O\left(n^{-1/2}\right)$ gives

$$H_\gamma^{-1}\{P(1|\boldsymbol{x})\} = G^{-1}(P(1|\boldsymbol{x})) + \gamma\frac{2G\left(\beta\left(\boldsymbol{x};\boldsymbol{\theta}_\gamma\right)\right) - 1}{g\left(\beta\left(\boldsymbol{x};\boldsymbol{\theta}_\gamma\right)\right)} + O\left(n^{-1}\right). \tag{10}$$

With

$$\boldsymbol{h}(\boldsymbol{\theta}) \stackrel{def}{=} \left(\ldots, \frac{2G(\beta(\boldsymbol{x}_i;\boldsymbol{\theta})) - 1}{g(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}))}, \ldots\right)' : N \times 1,$$

the expansion (10) then gives $\boldsymbol{\theta}_\gamma - \boldsymbol{\theta}_0 = \left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\boldsymbol{h}\left(\boldsymbol{\theta}_\gamma\right)\gamma + O\left(n^{-1}\right)$. In particular $\boldsymbol{\theta}_\gamma - \boldsymbol{\theta}_0 = O\left(n^{-1/2}\right)$ and so

$$\boldsymbol{\theta}_\gamma - \boldsymbol{\theta}_0 = \left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\boldsymbol{h}(\boldsymbol{\theta}_0)\gamma + O\left(n^{-1}\right). \tag{11}$$

Our approach in this article is as follows. In Section 3 we study the asymptotic properties of solutions $\hat{\boldsymbol{\theta}}_n$ to a weighted likelihood equation. Weights are proposed in Section 3.2. This estimate is $\sqrt{n}$-consistent for $\boldsymbol{\theta}_\gamma$, hence for $\boldsymbol{\theta}_0$, and is asymptotically normally distributed. The asymptotic mean and covariance structures depend on $\boldsymbol{\theta}_0$, but in these structures $\boldsymbol{\theta}_0$ can be replaced by $\hat{\boldsymbol{\theta}}_n$ without altering the asymptotic properties.

For a design of size n let $\xi_n(\boldsymbol{x})$ be the design measure, i.e. the function on χ taking values $\xi_n(\boldsymbol{x}_i) = n_i/n$ if n_i observations are made at $\boldsymbol{x}_i$. We exhibit, in Theorem 1, the asymptotic value of the total MSE

$$\mathcal{L}_\beta(\xi_n;\psi,\gamma) \stackrel{def}{=} \sum_{i=1}^{N} E\left[\left\{\sqrt{n}\left(\beta\left(\boldsymbol{x}_i;\hat{\boldsymbol{\theta}}_n\right) - \beta(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\right)\right\}^2\right]. \tag{12}$$

In Section 3.1 we obtain (a surrogate of) the maximum value $\mathcal{L}^{\max}(\xi_n;\boldsymbol{\theta}_0)$ of $\mathcal{L}_\beta(\xi_n;\psi,\gamma)$, over model errors ψ satisfying (4) and (8), and mislabelling probabilities satisfying (5).

In Section 4 we consider the construction of *minimax* designs, i.e. designs minimizing $\mathcal{L}^{\max}(\xi_n;\boldsymbol{\theta}_0)$. We first assume that $\boldsymbol{\theta}_0$ is known, an assumption which is of course generally quite unrealistic, except in situations such as local optimality in Phase 2 clinical trials — see for instance Lange and Schmidli (2014). We present an algorithm for the sequential construction of designs. By *sequential* we understand that the choice of a subsequent design point does not depend on the past responses, but only on the properties of the current design; thus the entire design can be constructed before experimentation begins. The theory of this section parallels that in Wiens (2018). We then drop this assumption of known parameters and apply the sequential algorithm after taking an initial sample so as to obtain a working parameter. We also investigate *adaptive* designs, in which the responses are learned and the parameters re-estimated whenever a new design point is chosen.

In Section 5.1 we consider the particular model of dose/covariate interactions

$$\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta} = \theta_0 x_0 + \boldsymbol{\theta}_1'\boldsymbol{f}_1(\boldsymbol{x}_{(1)}) + \boldsymbol{\theta}_2' x_0 \boldsymbol{f}_2(\boldsymbol{x}_{(1)}), \tag{13}$$

with $\boldsymbol{f}_1(\boldsymbol{x}_{(1)}) = \boldsymbol{x}_{(1)} : 3\times 1$ containing the indicators of two competing methods (so that their sum serves as the intercept) and a quantitative covariate and $\boldsymbol{f}_2(\boldsymbol{x}_{(1)}) : 2\times 1$ consisting of one of these indicators and the covariate. For this model

$$Q_p(\boldsymbol{x}_{(1)}) = \frac{G^{-1}(p) - \boldsymbol{\theta}_1'\boldsymbol{f}_1(\boldsymbol{x}_{(1)})}{\theta_0 + \boldsymbol{\theta}_2'\boldsymbol{f}_2(\boldsymbol{x}_{(1)})}. \tag{14}$$

We compare various weights and links, and also construct and evaluate asymptotic confidence intervals on $Q_p(\boldsymbol{x}_{(1)})$.

Proofs are in Appendix A; the relevant computing code (in MATLAB) is available on the author's web site, as are some online appendices containing details not reported here.

3. Asymptotic theory

Define

$$r(t) = \frac{d}{dt}\log\frac{G(t)}{\bar{G}(t)} = \frac{g(t)}{G(t)\bar{G}(t)},$$

with derivative $\dot{r}(t)$. We note that if G is the logistic distribution $\left(1+e^{-t}\right)^{-1}$ then $r \equiv 1$. Define

$$\alpha(\boldsymbol{x};\boldsymbol{\theta}) = g(\beta(\boldsymbol{x};\boldsymbol{\theta}))\, r(\beta(\boldsymbol{x};\boldsymbol{\theta}))\, w(\boldsymbol{x};\boldsymbol{\theta}).$$

For a design of size n, with n_i observations made at $\boldsymbol{x}_i$, denote by y_i the proportion of times that $Y=1$, so that $n_i y_i$ is the number of such occurrences. Then under the experimenter's assumptions that $\psi \equiv 0$ and $\gamma = 0$, the weighted, with weights $w(\boldsymbol{x}_i;\boldsymbol{\theta})$, gradient of the log-likelihood is

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \sum_{i=1}^N \xi_n(\boldsymbol{x}_i)\, w(\boldsymbol{x}_i;\boldsymbol{\theta})\, r(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}))\,(y_i - G(\beta(\boldsymbol{x}_i;\boldsymbol{\theta})))\,\boldsymbol{f}(\boldsymbol{x}_i).$$

We define the estimate $\hat{\boldsymbol{\theta}}_n$ as a zero of $\boldsymbol{S}_n(\boldsymbol{\theta})$.

Define

$$\begin{aligned} v(\boldsymbol{x};\boldsymbol{\theta}) &= G(\beta(\boldsymbol{x};\boldsymbol{\theta}))\,\bar{G}(\beta(\boldsymbol{x};\boldsymbol{\theta})),\\ v_*(\boldsymbol{x};\boldsymbol{\theta}) &= H_\gamma(\beta_*(\boldsymbol{x};\boldsymbol{\theta}))\,\bar{H}_\gamma(\beta_*(\boldsymbol{x};\boldsymbol{\theta})). \end{aligned}$$

Note that $v(\boldsymbol{x};\boldsymbol{\theta})$ (resp. $v_*(\boldsymbol{x};\boldsymbol{\theta})$) is the variance of $Y_{|\boldsymbol{x}}$ under (1) (resp. (6)). An occasionally useful identity is

$$\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}) = v(\boldsymbol{x}_i;\boldsymbol{\theta})\, r^2(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}))\, w(\boldsymbol{x}_i;\boldsymbol{\theta}).$$

Let $\boldsymbol{D}_\alpha(\boldsymbol{\theta})$, $\boldsymbol{D}_{\xi_n}$ and $\boldsymbol{D}_w(\boldsymbol{\theta})$ be the diagonal matrices with diagonal elements $\{\alpha(\boldsymbol{x}_i;\boldsymbol{\theta})\}_{i=1}^N$, $\{\xi_n(\boldsymbol{x}_i)\}_{i=1}^N$ and $\{w(\boldsymbol{x}_i;\boldsymbol{\theta})\}_{i=1}^N$ respectively. Define $\boldsymbol{\psi} = (\ldots,\psi(\boldsymbol{x}_i),\ldots)'$ and

$$\begin{aligned} \boldsymbol{c}_{n,0}(\boldsymbol{\theta}_0) &= \sum_{i=1}^N \xi_n(\boldsymbol{x}_i)\,\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\left\{\left[\boldsymbol{f}'(\boldsymbol{x}_i)\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\boldsymbol{h}(\boldsymbol{\theta}_0) - h_i(\boldsymbol{\theta}_0)\right]\gamma + \psi(\boldsymbol{x}_i)\right\}\boldsymbol{f}(\boldsymbol{x}_i)\\ &= \boldsymbol{F}'\boldsymbol{D}_{\xi_n}\boldsymbol{D}_\alpha(\boldsymbol{\theta}_0)\left\{\boldsymbol{\psi} - \left[I_N - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right]\boldsymbol{h}(\boldsymbol{\theta}_0)\,\gamma\right\},\\ \boldsymbol{M}_{n,0}(\boldsymbol{\theta}_0) &= \sum_{i=1}^N \xi_n(\boldsymbol{x}_i)\,\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\,\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i) = \boldsymbol{F}'\boldsymbol{D}_{\xi_n}\boldsymbol{D}_\alpha(\boldsymbol{\theta}_0)\,\boldsymbol{F},\\ \boldsymbol{V}_{n,0}(\boldsymbol{\theta}_0) &= \sum_{i=1}^N \xi_n(\boldsymbol{x}_i)\, w(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\,\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\,\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i) = \boldsymbol{F}'\boldsymbol{D}_{\xi_n}\boldsymbol{D}_\alpha(\boldsymbol{\theta}_0)\,\boldsymbol{D}_w(\boldsymbol{\theta}_0)\,\boldsymbol{F}. \end{aligned}$$

The presence of $\boldsymbol{\psi}$ and γ in the vector $\boldsymbol{c}_{n,0}$ determines the bias of the estimate. The matrix $\boldsymbol{M}_{n,0}$ is the asymptotic weighted information matrix; it and $\boldsymbol{V}_{n,0}$ appears in the 'sandwich' covariance matrix. For constant weights $w \equiv 1$, the matrices $\boldsymbol{V}_{n,0}$ and $\boldsymbol{M}_{n,0}$ are identical.

We assume that for all sufficiently large n,

$$\sum_{i=1}^N \xi_n(\boldsymbol{x}_i)\,\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i) = \boldsymbol{F}'\boldsymbol{D}_{\xi_n}\boldsymbol{F} \succ \boldsymbol{0}, \tag{15}$$

i.e. is positive definite.

A version of (1) in the following result was proved in Adewale and Wiens (2009). Conditions for the existence of the zero $\hat{\boldsymbol{\theta}}_n$ are given in Fahrmeir (1990).

Theorem 1. *(1) Any $\sqrt{n}$-consistent zero $\hat{\boldsymbol{\theta}}_n$ of $\mathbf{S}_n(\boldsymbol{\theta})$ is asymptotically normally distributed:*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\sim AN\left(\sqrt{n}\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{c}_{n,0}(\boldsymbol{\theta}_0),\ \mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{V}_{n,0}(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\right). \tag{16}$$

(2) Define the estimate of $\beta(\boldsymbol{x};\boldsymbol{\theta})$ by $\hat{\beta}_n(\boldsymbol{x})=\beta(\boldsymbol{x};\hat{\boldsymbol{\theta}}_n)$, with target value $\beta_0(\boldsymbol{x})=\beta(\boldsymbol{x};\boldsymbol{\theta}_0)$. Then

$$\sqrt{n}\left(\hat{\beta}_n(\boldsymbol{x})-\beta_0(\boldsymbol{x})\right)\sim AN\begin{pmatrix}\sqrt{n}\left(\mathbf{f}'(\boldsymbol{x})\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{c}_{n,0}(\boldsymbol{\theta}_0)\right),\\ \mathbf{f}'(\boldsymbol{x})\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{V}_{n,0}(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{f}(\boldsymbol{x})\end{pmatrix}. \tag{17}$$

The total asymptotic mean squared error (12) is

$$\mathcal{L}_\beta(\xi_n;\psi,\gamma)=tr\mathbf{F}\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)^{-1}\mathbf{V}_{n,0}(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{F}'+n\left\{\mathbf{c}_{n,0}'(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{F}'\mathbf{F}\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{c}_{n,0}(\boldsymbol{\theta}_0)\right\}. \tag{18}$$

(3) If $\rho(\boldsymbol{\theta})$ is twice differentiable, with gradient $\dot{\boldsymbol{\rho}}(\boldsymbol{\theta})$, then

$$\sqrt{n}\left(\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho(\boldsymbol{\theta}_0)\right)\sim AN\begin{pmatrix}-\sqrt{n}\dot{\boldsymbol{\rho}}'(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{c}_{n,0}(\boldsymbol{\theta}_0),\\ \dot{\boldsymbol{\rho}}'(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\mathbf{V}_{n,0}(\boldsymbol{\theta}_0)\mathbf{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\dot{\boldsymbol{\rho}}(\boldsymbol{\theta}_0)\end{pmatrix}. \tag{19}$$

3.1. Maximum asymptotic loss

Here we maximize $\mathcal{L}_\beta(\xi_n;\psi,\gamma)$, at (12), subject to (4), (5), and (8). For this, first let

$$\mathbf{Q}_1=(\mathbf{q}_1\cdots\mathbf{q}_N)':N\times d$$

be such that its columns form an orthogonal basis for the column space of $\mathbf{F}$ – this is 'Q' in the qr-decomposition of $\mathbf{F}$ – and extend $\mathbf{Q}_1$ to an $N\times N$ orthogonal matrix $\left(\mathbf{Q}_1\vdots\mathbf{Q}_2\right)$. Define matrices

$$\begin{aligned}\mathbf{A}(\xi_n;\boldsymbol{\theta}_0)&=\sum_{i=1}^N\xi_n(\boldsymbol{x}_i)\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i'=\mathbf{Q}_1'\mathbf{D}_{\xi_n}\mathbf{D}_\alpha(\boldsymbol{\theta}_0)\mathbf{Q}_1,\\ \mathbf{B}(\xi_n;\boldsymbol{\theta}_0)&=\sum_{i=1}^N\xi_n(\boldsymbol{x}_i)w(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i'=\mathbf{Q}_1'\mathbf{D}_{\xi_n}\mathbf{D}_w(\boldsymbol{\theta}_0)\mathbf{D}_\alpha(\boldsymbol{\theta}_0)\mathbf{Q}_1,\\ \mathbf{R}(\xi_n;\boldsymbol{\theta}_0)&=\mathbf{A}(\xi_n;\boldsymbol{\theta}_0)\mathbf{B}^{-1}(\xi_n;\boldsymbol{\theta}_0)\mathbf{A}(\xi_n;\boldsymbol{\theta}_0),\\ \mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)&=\mathbf{Q}_2'\mathbf{D}_\alpha(\boldsymbol{\theta}_0)\mathbf{D}_{\xi_n}\mathbf{Q}_1\mathbf{A}^{-2}(\xi_n;\boldsymbol{\theta}_0)\mathbf{Q}_1'\mathbf{D}_{\xi_n}\mathbf{D}_\alpha(\boldsymbol{\theta}_0)\mathbf{Q}_2.\end{aligned}$$

Theorem 2. *If $\tau_1,\tau_2>0$ then the maximum of $\mathcal{L}_\beta(\xi_n;\psi,\gamma)$, over all ψ satisfying (4) and (8), and all γ satisfying (5), is*

$$\max_{\psi,\gamma}\mathcal{L}_\beta(\xi_n;\psi,\gamma)=tr\mathbf{R}^{-1}(\xi_n;\boldsymbol{\theta}_0)+\left\{2\lambda_0-\tau_1^2\mathbf{u}_0'\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{u}_0+\tau_2^2\mathbf{b}'(\boldsymbol{\theta}_0)\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{b}(\boldsymbol{\theta}_0)\right\}, \tag{20}$$

where

$$\mathbf{b}(\boldsymbol{\theta}_0)=\mathbf{Q}_2'\mathbf{h}(\boldsymbol{\theta}_0)=\mathbf{Q}_2'\left(\ldots,\frac{2G(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}_0))-1}{g(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}_0))},\ldots\right)',$$

and $(\mathbf{u}_0,\lambda_0)$ are determined by

$$\left(\tau_1^2\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)-\lambda_0\mathbf{I}\right)\mathbf{u}_0=\tau_1\tau_2\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{b}(\boldsymbol{\theta}_0),\ \text{with} \tag{21}$$

$$\lambda_0=\tau_1^2\mathbf{u}_0'\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{u}_0-\tau_1\tau_2\mathbf{u}_0'\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{b}(\boldsymbol{\theta}_0). \tag{22}$$

Necessarily, $\mathbf{u}_0'\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{b}(\boldsymbol{\theta}_0)\le 0$.
(i) If $\tau_2=0$ (no labelling errors) then (20) becomes

$$\max_\psi\mathcal{L}_\beta(\xi_n;\psi)_{|\tau_2=0}=tr\mathbf{R}^{-1}(\xi_n;\boldsymbol{\theta}_0)+\tau_1^2\lambda_*, \tag{23}$$

where, with $ch_{\max}$ denoting the maximum eigenvalue,

$$\begin{aligned}\lambda_*&=ch_{\max}\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)=ch_{\max}\mathbf{U}(\xi_n;\boldsymbol{\theta}_0)-1,\ \text{for}\\ \mathbf{U}(\xi_n;\boldsymbol{\theta}_0)&=\mathbf{A}^{-1}(\xi_n;\boldsymbol{\theta}_0)\mathbf{Q}_1'\mathbf{D}_{\xi_n}^2\mathbf{D}_\alpha^2(\boldsymbol{\theta}_0)\mathbf{Q}_1\mathbf{A}^{-1}(\xi_n;\boldsymbol{\theta}_0):d\times d.\end{aligned}$$

(ii) If $\tau_1=0$ (no model errors) then (20) becomes

$$\max_\gamma\mathcal{L}_\beta(\xi_n;\gamma)_{|\tau_1=0}=tr\mathbf{R}^{-1}(\xi_n;\boldsymbol{\theta}_0)+\tau_2^2\mathbf{b}'(\boldsymbol{\theta}_0)\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)\mathbf{b}(\boldsymbol{\theta}_0).$$

Remark 3. The maximum (20) is attained at $\gamma=\tau_2/\sqrt{n}$ and $\boldsymbol{\psi}_0=\left(\tau_1/\sqrt{n}\right)\mathbf{Q}_2\mathbf{u}_0$. In (23) it is attained at $\boldsymbol{\psi}_*=\left(\tau_1/\sqrt{n}\right)\mathbf{Q}_2\mathbf{u}_*$, where $\mathbf{u}_*$ is the unit eigenvector of $\mathbf{U}_*(\xi_n;\boldsymbol{\theta}_0)$ belonging to λ_*.

Remark 4. In each case the first term in max $\mathcal{L}_\beta$ arises solely from variation, the second solely from bias. It is interesting to note that the limiting (as $\nu_1+\nu_2 \to 1$), 'unbiased' design can be obtained explicitly. It is given by $\xi_n(\boldsymbol{x}_i) \propto 1/\alpha(\boldsymbol{x}_i;\boldsymbol{\theta}_0)$. For this design – which is not practical since it places mass on every point in χ – we have $\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)=\boldsymbol{0}$.

Remark 5. There is a natural monotonicity — the loss decreases if $\boldsymbol{U}_*$ becomes smaller with respect to the Loewner ordering by positive semidefiniteness.

If either $\tau_1 = 0$ or $\tau_2 = 0$ then, as seen from Theorem 2, the 'max' part of the minimax design problem is straightforward. If both are nonzero, then even the maximization must be done numerically — each time a possible design is assessed. Thus we shall instead aim to design so as to minimize a mixture of the two individual maxima, viz., we minimize $(1+\tau_1^2+\tau_2^2)$ times

$$\mathcal{L}_{\nu_1,\nu_2}(\xi_n;\boldsymbol{\theta}_0) = (1-\nu_1-\nu_2)\, tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0) + \nu_1\lambda_* + \nu_2\boldsymbol{b}'(\boldsymbol{\theta}_0)\,\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b}(\boldsymbol{\theta}_0), \tag{24}$$

where $\nu_1=\tau_1^2/(1+\tau_1^2+\tau_2^2)$ and $\nu_2=\tau_2^2/(1+\tau_1^2+\tau_2^2)$. Then the individual maxima can be recovered as

$$\max_{\psi}\mathcal{L}_\beta(\xi_n;\psi)_{|\tau_2=0} = \left(1+\tau_1^2\right)\mathcal{L}_{\nu_1,0}(\xi_n;\boldsymbol{\theta}_0),$$
$$\max_{\gamma}\mathcal{L}_\beta(\xi_n;\gamma)_{|\tau_1=0} = \left(1+\tau_2^2\right)\mathcal{L}_{0,\nu_2}(\xi_n;\boldsymbol{\theta}_0).$$

Remark 6. Note that $\mathcal{L}_{\nu_1,\nu_2}(\cdot;\cdot)$ does not depend on (τ_1,τ_2), which are thus not needed for the construction of the designs. In the examples below we evaluate $\mathcal{L}_{\nu_1,\nu_2}$ at various values of (ν_1,ν_2) for fixed values of (τ_1,τ_2). The values of (τ_1,τ_2) are used only in (4) and (5), with equality in each case. Then ν_1 and ν_2 represent the relative importance, to the experimenter, of the two types of bias and of the variance.

3.2. Weights

We have considered various choices of weights $\{w(\boldsymbol{x}_i;\boldsymbol{\theta})\}$. These include constant weights $w_0(\boldsymbol{x}) \equiv 1$, weights $w_1(\boldsymbol{x}_i)=h_{ii}^{-1/2}$, where $\{h_{ii}\}$ are the diagonal elements of the 'hat' matrix $\boldsymbol{H}=\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'$, and

$$w_2(\boldsymbol{x}_i;\boldsymbol{\theta}) = \frac{1}{r(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}))\sqrt{h_{ii}G(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}))\,\bar{G}(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}))}}.$$

In a study of sampling schemes for classification studies (López-Fidalgo and Wiens, 2020), weights w_2 were shown to minimize the variance component of the loss, in 'unbiased' sampling schemes.

We considered as well weights $w_3(\boldsymbol{x}) = 1/G(\beta(\boldsymbol{x};\boldsymbol{\theta}))\,\bar{G}(\beta(\boldsymbol{x};\boldsymbol{\theta}))$, inversely proportional to the variance of the response at $\boldsymbol{x}$.

Remark 7. In the simulations of Section 5.1 weights w_3 had the drawback that when using them the procedure used to find the zero of (25) (*fsolve* in MATLAB) very often failed to converge, even when run repeatedly from different starting points. Thus we compared constant weights w_0 with weights w_1 and w_2, and decided – see Section 5.1 – in favour of w_1, and the use of the probit link. In the example of Section 5.2 weights w_3 were more stable, and resulted in confidence intervals with close to nominal coverage, but were disappointingly inefficient for the design sizes considered there.

4. Design construction

In this section we consider the effect of augmenting an n-point design ξ_n by an additional point at $\boldsymbol{x}_i$, thus obtaining the design

$$\xi_{n+1}^{(i)} = \frac{n}{n+1}\xi_n + \frac{1}{n+1}\delta(\boldsymbol{x}_i).$$

(Here $\delta(\cdot)$ denotes a point mass.) Theorem 3 gives the points $\boldsymbol{x}_i$ which minimize the resulting maximum loss (24), up to terms which are $O\left(n^{-1}\right)$. The method is motivated by a similar application in Wiens (2018).

For any design ξ on χ, denote by $\lambda(\xi;\boldsymbol{\theta})$ and $\boldsymbol{z}(\xi;\boldsymbol{\theta})$ the maximum eigenvalue, and corresponding unit eigenvector, of $\boldsymbol{U}(\xi;\boldsymbol{\theta})$. Define $\boldsymbol{c}(\xi;\boldsymbol{\theta})=\mathbf{A}^{-1}(\xi;\boldsymbol{\theta})\,\boldsymbol{z}(\xi;\boldsymbol{\theta})$ and

$$\boldsymbol{S}(\xi;\boldsymbol{\theta}) = \boldsymbol{Q}_1'\boldsymbol{D}_\xi\boldsymbol{D}_\alpha(\boldsymbol{\theta})\,\boldsymbol{Q}_2 : d\times N-d. \tag{25}$$

Our results depend on the N diagonal elements $\{\boldsymbol{T}_{ii}\}_{i=1}^N$ of the matrix

$$\boldsymbol{T}(\boldsymbol{\xi};\boldsymbol{\theta}) = (1-\nu_1-\nu_2)\,\boldsymbol{T}^{(0)}(\boldsymbol{\xi};\boldsymbol{\theta}) + \nu_1\boldsymbol{T}^{(1)}(\boldsymbol{\xi};\boldsymbol{\theta}) + \nu_2\boldsymbol{T}^{(2)}(\boldsymbol{\xi};\boldsymbol{\theta}),$$

where

$$\boldsymbol{T}^{(0)}(\boldsymbol{\xi};\boldsymbol{\theta}) = \boldsymbol{D}_{\alpha}(\boldsymbol{\theta})\left[2\mathbf{Q}_1\boldsymbol{A}^{-1}(\xi;\boldsymbol{\theta})\boldsymbol{R}^{-1}(\xi;\boldsymbol{\theta}_0) - \boldsymbol{D}_w(\boldsymbol{\theta})\mathbf{Q}_1\boldsymbol{A}^{-2}(\xi;\boldsymbol{\theta})\right]\mathbf{Q}_1',$$
$$\boldsymbol{T}^{(1)}(\boldsymbol{\xi};\boldsymbol{\theta}) = \boldsymbol{D}_{\alpha}(\boldsymbol{\theta})\left[2\left\{\mathbf{Q}_1\boldsymbol{U}(\xi;\boldsymbol{\theta})\boldsymbol{A}(\xi;\boldsymbol{\theta}) - \boldsymbol{D}_{\alpha}(\boldsymbol{\theta})\boldsymbol{D}_{\xi}\mathbf{Q}_1\right\}\right]\boldsymbol{c}(\xi;\boldsymbol{\theta})\boldsymbol{c}'(\xi;\boldsymbol{\theta})\mathbf{Q}_1',$$
$$\boldsymbol{T}^{(2)}(\boldsymbol{\xi};\boldsymbol{\theta}) = \boldsymbol{D}_{\alpha}(\boldsymbol{\theta})\left[2\left\{\mathbf{Q}_1\boldsymbol{A}^{-1}(\xi;\boldsymbol{\theta})\boldsymbol{S}(\xi;\boldsymbol{\theta}_0) - \mathbf{Q}_2\right\}\boldsymbol{b}(\boldsymbol{\theta})\boldsymbol{b}'(\boldsymbol{\theta})\boldsymbol{S}'(\xi;\boldsymbol{\theta})\boldsymbol{A}^{-2}(\xi;\boldsymbol{\theta}_0)\right]\mathbf{Q}_1'.$$

Theorem 3. *(1) Let ξ_0 be a design on χ. When $\nu_1 > 0$ assume that the maximum eigenvalue $\lambda(\xi_0;\boldsymbol{\theta}_0)$ is simple. In order for ξ_0 to minimize $\mathcal{L}_{\nu_1,\nu_2}(\xi;\boldsymbol{\theta}_0)$, it is necessary to have*

$$\frac{\boldsymbol{T}_{ii}(\xi_0;\boldsymbol{\theta}_0)}{(1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\xi_0;\boldsymbol{\theta}_0)} - 1 \le 0 \text{ for all } i = 1,\ldots,N, \text{ with equality if } \xi_0(\boldsymbol{x}_i) > 0. \tag{26}$$

(2) Define $\pi(\xi) = \frac{\max_i \boldsymbol{T}_{ii}(\xi;\boldsymbol{\theta}_0)}{(1-\nu_1-\nu_2)tr\boldsymbol{R}^{-1}(\xi;\boldsymbol{\theta}_0)} - 1$. For any design, $\pi(\xi) \ge 0$. Condition (26) is equivalent to $\pi(\xi_0) = 0$.
(3) We have that

$$\mathcal{L}_{\nu_1,\nu_2}\left(\xi_{n+1}^{(i)};\boldsymbol{\theta}_0\right) = \mathcal{L}_{\nu_1,\nu_2}(\xi_n;\boldsymbol{\theta}_0) - \frac{1}{n+1}\left\{\boldsymbol{T}_{ii}(\xi_n;\boldsymbol{\theta}_0) - (1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0)\right\} + O\left(n^{-2}\right), \tag{27}$$

so that, including terms up to $O\left(n^{-1}\right)$, $\mathcal{L}_{\nu_1,\nu_2}\left(\xi_{n+1}^{(i)},\boldsymbol{\theta}_0\right)$ is minimized by

$$i_0 = \arg\max_i \boldsymbol{T}_{ii}(\xi_n;\boldsymbol{\theta}_0). \tag{28}$$

Remark 8. When $\tau_1 > 0$, that $\lambda(\xi_n)$ is simple is checked numerically.

Remark 9. Motivated by (2) of Theorem 3, in the computations we use the stopping rule $\pi(\xi_n) \le \pi_*$, for a suitably small $\pi_* > 0$. See Fig. 2, where we have plotted $\pi(\xi_n)$ and the limit $\pi_* = .02$.

Remark 10. In Wiens (2018) it is shown that, in a regression context and under suitable conditions, (26) is sufficient as well as necessary for asymptotic optimality. For brevity these arguments are not repeated here.

5. Simulations and examples

5.1. Example 1. A dose/covariate interaction model

Here we consider the dose/covariate interaction model (13), and present the results of a simulation study. For a specified p the required dose Q_p is given by (14). Define $\rho(\boldsymbol{\theta}_0) = x_0 = Q_p\left(\boldsymbol{x}_{(1)}\right)$, with $\hat{x}_0 = \rho\left(\hat{\boldsymbol{\theta}}_n\right)$. Then $\dot{\boldsymbol{\rho}}(\boldsymbol{\theta}_0) = -\rho(\boldsymbol{\theta}_0)\,\boldsymbol{t}(\boldsymbol{\theta}_0)$ for

$$\boldsymbol{t}(\boldsymbol{\theta}_0) = \begin{pmatrix} 1/\left(\theta_0 + \boldsymbol{\theta}_2'\boldsymbol{f}_2\left(\boldsymbol{x}_{(1)}\right)\right) \\ \boldsymbol{f}_1\left(\boldsymbol{x}_{(1)}\right)/\left(G^{-1}(p) - \boldsymbol{\theta}_1'\boldsymbol{f}_1\left(\boldsymbol{x}_{(1)}\right)\right) \\ \boldsymbol{f}_2\left(\boldsymbol{x}_{(1)}\right)/\left(\theta_0 + \boldsymbol{\theta}_2'\boldsymbol{f}_2\left(\boldsymbol{x}_{(1)}\right)\right) \end{pmatrix},$$

and (3) of Theorem 1 gives that

$$\sqrt{n}\left(\frac{\hat{x}_0}{x_0} - 1\right) \sim AN\left(\mu_n, \sigma_n^2\right), \text{ for} \tag{29}$$
$$\mu_n = \sqrt{n}\boldsymbol{t}'(\boldsymbol{\theta}_0)\boldsymbol{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{c}_{n,0}(\boldsymbol{\theta}_0), \text{ and}$$
$$\sigma_n^2 = \boldsymbol{t}'(\boldsymbol{\theta}_0)\boldsymbol{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}_{n,0}(\boldsymbol{\theta}_0)\boldsymbol{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{t}(\boldsymbol{\theta}_0).$$

The asymptotic normality (29) can be used to construct interval estimates, given an evaluation of the asymptotic mean μ_n, which depends on the unknown δ and $\boldsymbol{\psi}$. An asymptotic $100(1-\alpha)\%$ confidence interval on x_0 has endpoints $\hat{x}_0\left(1 - \frac{\hat{\mu}_n}{\sqrt{n}}\right) \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)|\hat{x}_0|\frac{\hat{\sigma}_n}{\sqrt{n}}$. Since $\mu_n/\sqrt{n} \to 0$ we settled on $\hat{\mu}_n/\sqrt{n} = 0$, giving endpoints

$$\hat{x}_0 \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)|\hat{x}_0|\sqrt{\frac{\boldsymbol{t}\left(\hat{\boldsymbol{\theta}}\right)'\boldsymbol{M}_{n,0}^{-1}\left(\hat{\boldsymbol{\theta}}\right)\boldsymbol{V}_{n,0}\left(\hat{\boldsymbol{\theta}}\right)\boldsymbol{M}_{n,0}^{-1}\left(\hat{\boldsymbol{\theta}}\right)\boldsymbol{t}\left(\hat{\boldsymbol{\theta}}\right)}{n}}, \tag{30}$$

with asymptotic coverage $1 - \alpha + o_P(1)$.

For the simulations, we took x_0 = dose, taking values $\{.1, .4, .7, 1\}$, $\boldsymbol{x}_{(1)} = \left(x_{(1),1}, x_{(1),2}, x_{(1),3}\right)'$ with $x_{(1),1} = I$(method 1), $x_{(1),2} = I$(method 2) $= 1 - x_{(1),1}$, $x_{(1),3}$ = quantitative covariate taking values $\{1, 2, 3, 4, 5\}$. We set

Table 1
Example 1. Means, over 50 runs, of loss components for converged adaptive designs. Weights $w_1(\boldsymbol{x})$; $\tau_1 = 15,\ \tau_2 = 1$.

Link↓	(ν_1, ν_2) :	(0, 0)	(.25, .25)	(.45, .45)	(.90, 0)	(0, .90)
Logistic	$\mathcal{L}_{\nu_1,\nu_2}$	388	196	41	41	41
	BIAS	.48	.40	.25	.32	.20
	$\sqrt{\text{MSE}}$	1.29	1.36	.92	.99	.78
Probit	$\mathcal{L}_{\nu_1,\nu_2}$	476	798	154	612	254
	BIAS	.41	.70	.32	.43	.45
	$\sqrt{\text{MSE}}$	1.19	1.50	1.10	1.11	1.02

Table 2
Example 1. Means, over 50 runs, of loss components for 200-point sequential and adaptive (in parentheses) designs. Weights $w_0(\boldsymbol{x})$; $\tau_1 = 15$, $\tau_2 = 1$.

Link↓	(ν_1, ν_2) :	(0, 0)	(.25, .25)	(.45, .45)	(.90, 0)	(0, .90)
Logistic	$\mathcal{L}_{\nu_1,\nu_2}$	1450 (388)	3292 (193)	1490 (42)	217 (41)	380 (43)
	BIAS	38 (.66)	36 (.73)	21 (.45)	31 (.73)	58 (.72)
	$\sqrt{\text{MSE}}$	84 (1.77)	60 (1.69)	40 (1.49)	58 (1.67)	94 (1.64)
Probit	$\mathcal{L}_{\nu_1,\nu_2}$	1246 (402)	2785 (201)	1547 (43)	243 (42)	568 (45)
	BIAS	36 (.27)	21 (.66)	23 (.79)	24 (.63)	35 (.78)
	$\sqrt{\text{MSE}}$	92 (1.85)	38 (1.69)	42 (1.62)	47 (1.55)	59 (1.58)

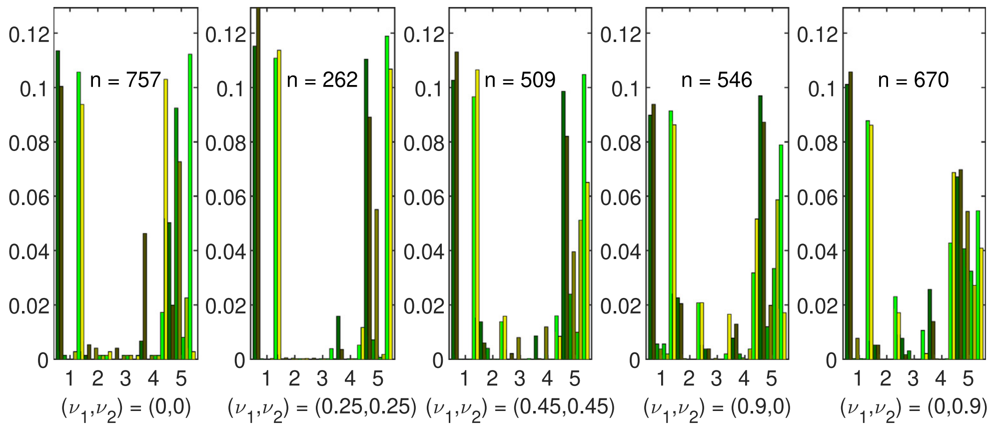

Fig. 1. Converged designs; probit link, weights $w_1(\cdot)$. Design sizes (n) at convergence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\boldsymbol{f}_1\left(\boldsymbol{x}_{(1)}\right) = \boldsymbol{x}_{(1)}, \boldsymbol{f}_2\left(\boldsymbol{x}_{(1)}\right) = \left(x_{(1),1}, x_{(1),3}\right)'$. Then $k = 3$, $N = 4 \cdot 2 \cdot 5 = 40$, $d = 6$. In order that the two types of error have about the same effect on $\left\|F\left(\boldsymbol{\theta}_\gamma - \boldsymbol{\theta}_0\right) + \boldsymbol{\psi}\right\|$ we took $\tau_1 = 15,\ \tau_2 = 1$, and chose

$$\boldsymbol{\theta}_\gamma = (.05, -.25, -.25, -.5, 0, 1)',$$

to compute $\beta_*\left(\boldsymbol{x}; \boldsymbol{\theta}_\gamma\right)$ as at (3). We took an initial design of size $n_{init} = 20$. This design was used to simulate ψ, for which we iterated between the construction of $\psi_*(\boldsymbol{\theta}_0)$ as in Remark 3 (with constant weights), and the computation of $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0(\psi)$ from (9) and (6). This process generally converged very quickly.

When proceeding adaptively we simulated responses from (3) and re-estimated the parameters each time before choosing a new design point as at (28). In the sequential cases observations were simulated only for the initial sample, and again after the designs were constructed.

This procedure was carried out 50 times, with the mean outputs reported in the tables of the online Appendix 2 (from which Tables 1 and 2, presented here, were selected). These tables give the values of the minimized maximum loss $\mathcal{L}_{\nu_1,\nu_2}(\xi_n; \boldsymbol{\theta}_0)$ for various choices of ν_1, ν_2, together with the BIAS $= \left\| aver \boldsymbol{F}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\|$ and root-MSE $= \left(aver \left\|\boldsymbol{F}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\|^2\right)^{1/2}$, for weights w_0, w_1, w_2 respectively. We considered the logistic link ($G_0(t) = 1/(1+e^{-t})$) and the probit ($G_0(t) = \Phi(t)$) link. To facilitate comparisons between the links, they were normalized to have unit variance: $G(t) = G_0(\sigma t)$.

A summary of the material in the aforementioned Appendix 2 is that the robust designs outperformed the random designs, and the adaptive designs outperformed the sequential designs, almost uniformly in terms of MSE and BIAS. Thus

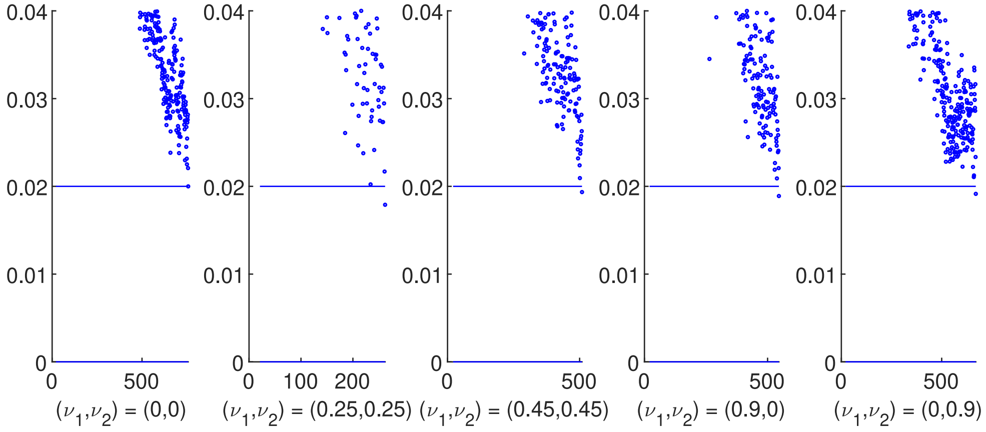

Fig. 2. Convergence measures $\pi(\xi_n)$ for the designs of Fig. 1.

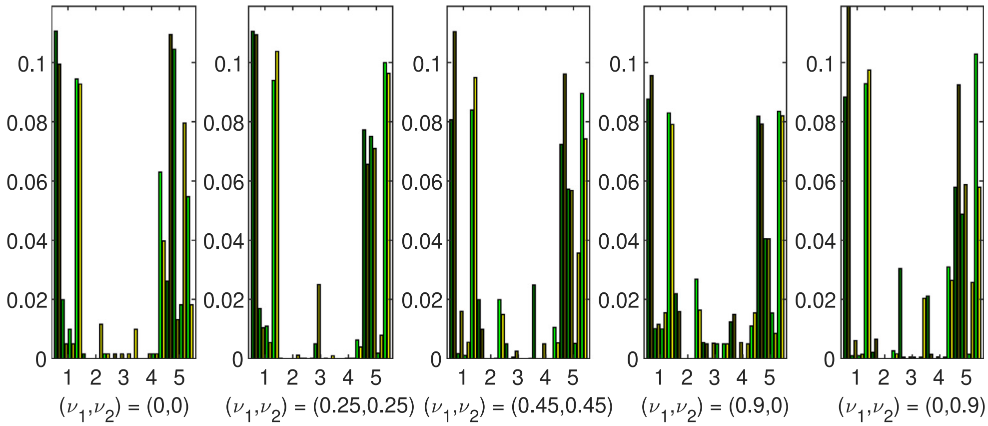

Fig. 3. Adaptively constructed designs of size 200; probit link, weights $w_0(\cdot)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we discuss only the robust, adaptive designs. We ran the procedure described in Section 4 to convergence, defined by $\pi_* = .02$ (recall Remark 9); we found – recall Remark 7 – that the use of the weights w_3 very often led to severe numerical instabilities. As seen in Table 1 the combination of weights w_1 together with the probit link enjoyed an advantage in terms of MSE and BIAS. The designs tend to become more diffuse with increasing ν_1.

See Figs. 1 and 2 for the designsand convergence measures in a typical (probit; weights w_1) run, and Table 1 for comparative values of the loss. In the design plots the labels 1 to 5 indicate the covariate values, with 8 bars per group. The bars are shades of green for method 1, yellow for method 2, with the intensity of the colour increasing with the dose.

To compare designs of equal size, we also obtained sequential and adaptive designs with a final study size of $n_{\max} = 200$. See Fig. 3 and Table 2. This led to similar conclusions as above, moderated in light of the smaller design sizes. As expected the adaptive designs were uniformly better that their sequential counterparts, often by a huge margin. When used with the – generally superior – probit link, weights w_1 again were slightly better than the others, followed closely by constant weights w_0.

Both the sequential and adaptive designs achieve a fair degree of balance across method groups, even though this was not imposed as a requirement. For instance for the designs of Fig. 3, the proportions of allocations to method 2 were .49, .50, .50, .50, .50.

5.2. Example 2: Menarche study

5.2.1. Simulation of the data

The data for the original Warsaw study of Milicer and Szczotka (1966) as described in §1 are summarized in Table 2 of Müller and Schmitt (1990), where the 2869 observations on $x_0 = age$ are grouped into 24 bins, and $\sum Y(x_0)$ reported for each bin. For this study we use these same age classes, with $E[Y(x_0)]$ obtained from the Warsaw study. We simulated 2869 data points $\boldsymbol{x}_{(1)}$ following, approximately, models postulated by Edvardsson et al. (2009) and Mohamad et al. (2013)

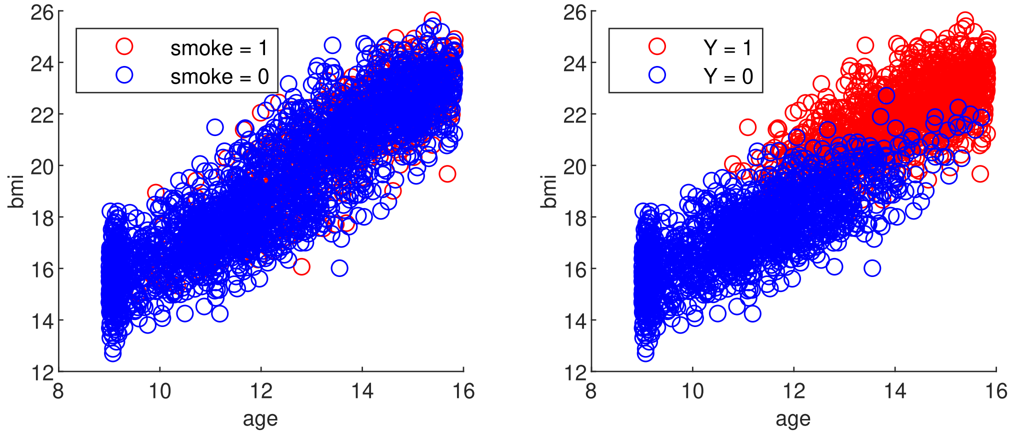

Fig. 4. Representative, simulated menarche data.

Table 3
Example 2. Means, over 50 runs, of loss components for 200-point adaptive designs; $\tau_1 = 15,\ \tau_2 = 1$.

Link↓	Weights:	w_0		w_1		w_2	
	(ν_1, ν_2):	(.45, .45)	(0, .90)	(.45, .45)	(0, .90)	(.45, .45)	(0, .90)
Logistic	BIAS	5.7	6.4	5.3	5.9	6.3	5.6
	$\sqrt{\text{MSE}}$	8.0	8.1	8.8	8.1	8.6	7.2
	Coverage (%)	69.7	66.0	84.0	79.3	72.0	68.7
Probit	BIAS	6.2	6.1	4.6	5.2	4.7	5.4
	$\sqrt{\text{MSE}}$	11.8	8.2	6.4	7.0	7.2	7.0
	Coverage (%)	66.3	63.3	74.0	75.3	69.0	65.7

relating age of menarche, BMI and smoking. These were then grouped into the 24 quantitative age classes, 3 quantitative bmi classes – all members of the age and BMI classes are identified with the class means – and 2 smoking classes, yielding a design space of size $N = 24 \cdot 3 \cdot 2 = 144$.

We fit (13), with

$$\boldsymbol{f}_1\left(\boldsymbol{x}_{(1)}\right) = \boldsymbol{f}_2\left(\boldsymbol{x}_{(1)}\right) = \boldsymbol{x}_{(1)} = \begin{pmatrix} x_{(1),1} \\ x_{(1),2} \end{pmatrix}, \text{ for } x_{(1),1} = bmi,\ x_{(1),2} = I\,(smoker)\,.$$

For each of the N members of the design space we estimate $P\,(Y = 1)$ by the observed proportions among the 2869 observations; when our algorithm calls for a response from one of these classes to be simulated we generate a Bernoulli r.v. with 'true' probability of success equal to this estimate.

A regression using all N design points and these 'true' probabilities as the responses results in our parameter vector $\boldsymbol{\theta}_\gamma$, from which $\boldsymbol{\theta}_0$ is computed and used in evaluating our designs. See Fig. 4 for a typical case of the simulated population. The 'coverages' in Table 3 are the realized coverages of 90% confidence intervals on $x_0 = age$, as at (30), averaged over 50 runs and all 6 smoking/bmi combinations. See Figs. 5–7 for all 50×6 such intervals for the particular choices of inputs given in the captions, and the corresponding designs weights, averaged over runs. From these one sees that when there is little or no weight placed on response mislabelling (small ν_2, as in Fig. 5), then confidence interval coverage is poor. Another feature shown by this figure is that when ν_1 is large the design weights tend to be more uniformly spread over the ages in each bmi/smoking class. When ν_2 is larger, as in Fig. 6, coverage is much improved and the design weights are more concentrated near a few extreme values.

Acknowledgements

The research of Douglas P. Wiens is supported by the Natural Sciences and Engineering Council of Canada. This article has benefited from the insightful comments of two anonymous referees.

Appendix A. Derivations

Proof of Theorem 1. (1) First define

$$\boldsymbol{c}_n\left(\boldsymbol{\theta}_0\right) = E\left[\boldsymbol{U}_n\left(\boldsymbol{\theta}_0\right)\right] = \sum_{i=1}^{N} \xi_n\left(\boldsymbol{x}_i\right) \boldsymbol{f}\left(\boldsymbol{x}_i\right) w\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right) \left\{\left(H_\gamma\left(\beta_*\left(\boldsymbol{x}_i; \boldsymbol{\theta}_\gamma\right)\right) - G\left(\beta\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right)\right) r\left(\beta\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right)\right\}. \tag{A.1}$$

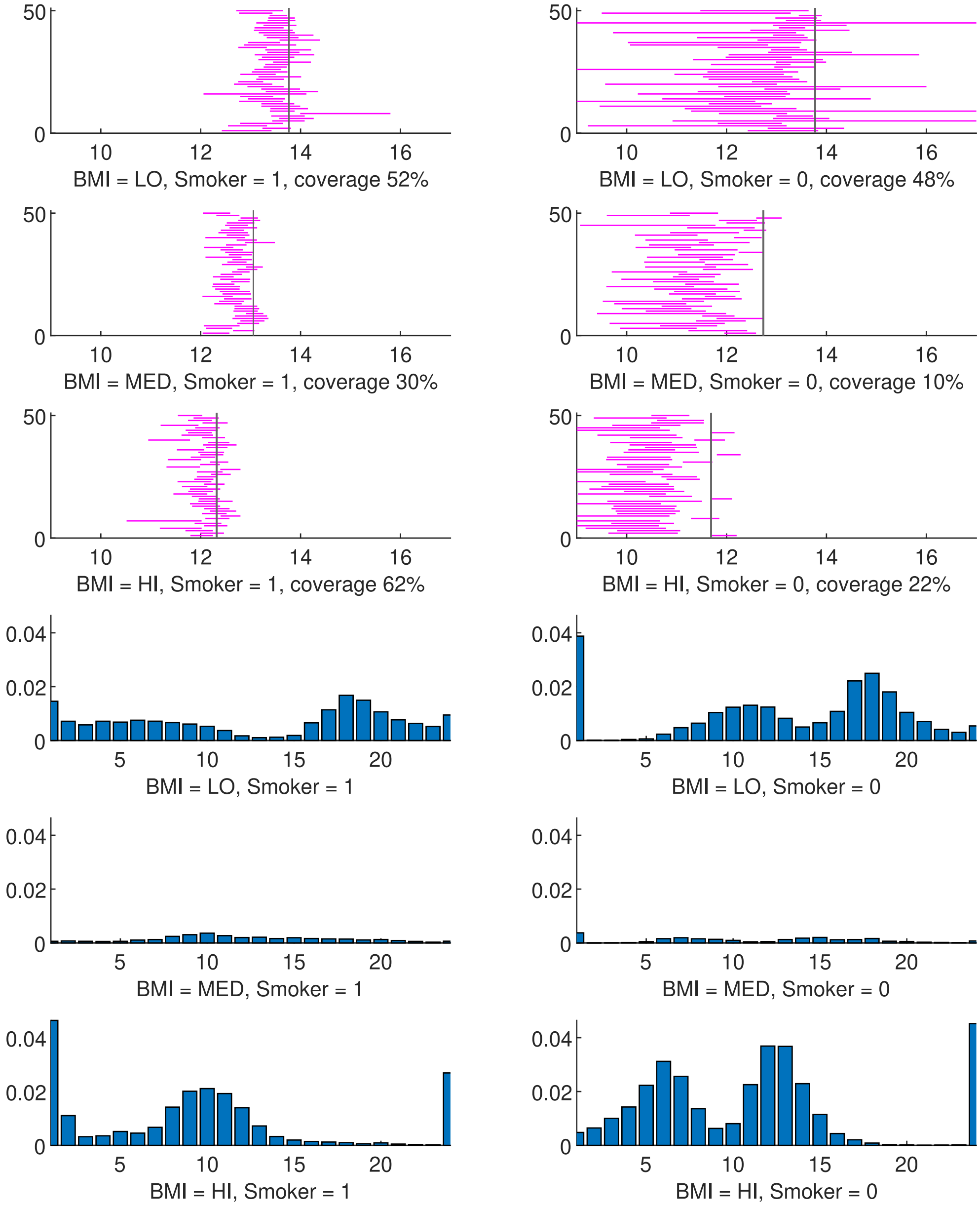

Fig. 5. Example 2. Confidence intervals and mean design weights; $(\nu_1, \nu_2) = (.9, 0)$, estimation weights w_1, logistic regression.

Consider the expansion

$$\begin{aligned} & H_\gamma\left(\beta_*\left(\boldsymbol{x}_i;\boldsymbol{\theta}_\gamma\right)\right) \\ =\ & G\left\{\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)+\left[\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\left(\boldsymbol{\theta}_\gamma-\boldsymbol{\theta}_0\right)+\psi\left(\boldsymbol{x}_i\right)\right]\right\} \\ & +\gamma\left(1-2G\left\{\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)+\left[\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\left(\boldsymbol{\theta}_\gamma-\boldsymbol{\theta}_0\right)+\psi\left(\boldsymbol{x}_i\right)\right]\right\}\right) \\ =\ & G\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)+g\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\left[\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\left(\boldsymbol{\theta}_\gamma-\boldsymbol{\theta}_0\right)+\psi\left(\boldsymbol{x}_i\right)\right] \\ & +\gamma\left(1-2G\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\right)+O\left(n^{-1}\right) \\ =\ & G\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)+g\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\left\{\left[\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\left(\boldsymbol{\theta}_\gamma-\boldsymbol{\theta}_0\right)+\psi\left(\boldsymbol{x}_i\right)\right]-\left[\boldsymbol{h}_i\left(\boldsymbol{\theta}_0\right)\gamma\right]\right\}+O\left(n^{-1}\right). \end{aligned} \tag{A.2}$$

Together with (11), and defining $\boldsymbol{e}_i$ to be the ith column of $\boldsymbol{I}_N$, this gives

$$H_\gamma\left(\beta_*\left(\boldsymbol{x}_i;\boldsymbol{\theta}_\gamma\right)\right)-G\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)=g\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\left\{\left[\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'-\boldsymbol{e}_i'\right]\boldsymbol{h}\left(\boldsymbol{\theta}_0\right)\gamma+\psi\left(\boldsymbol{x}_i\right)\right\}+O\left(n^{-1}\right),$$

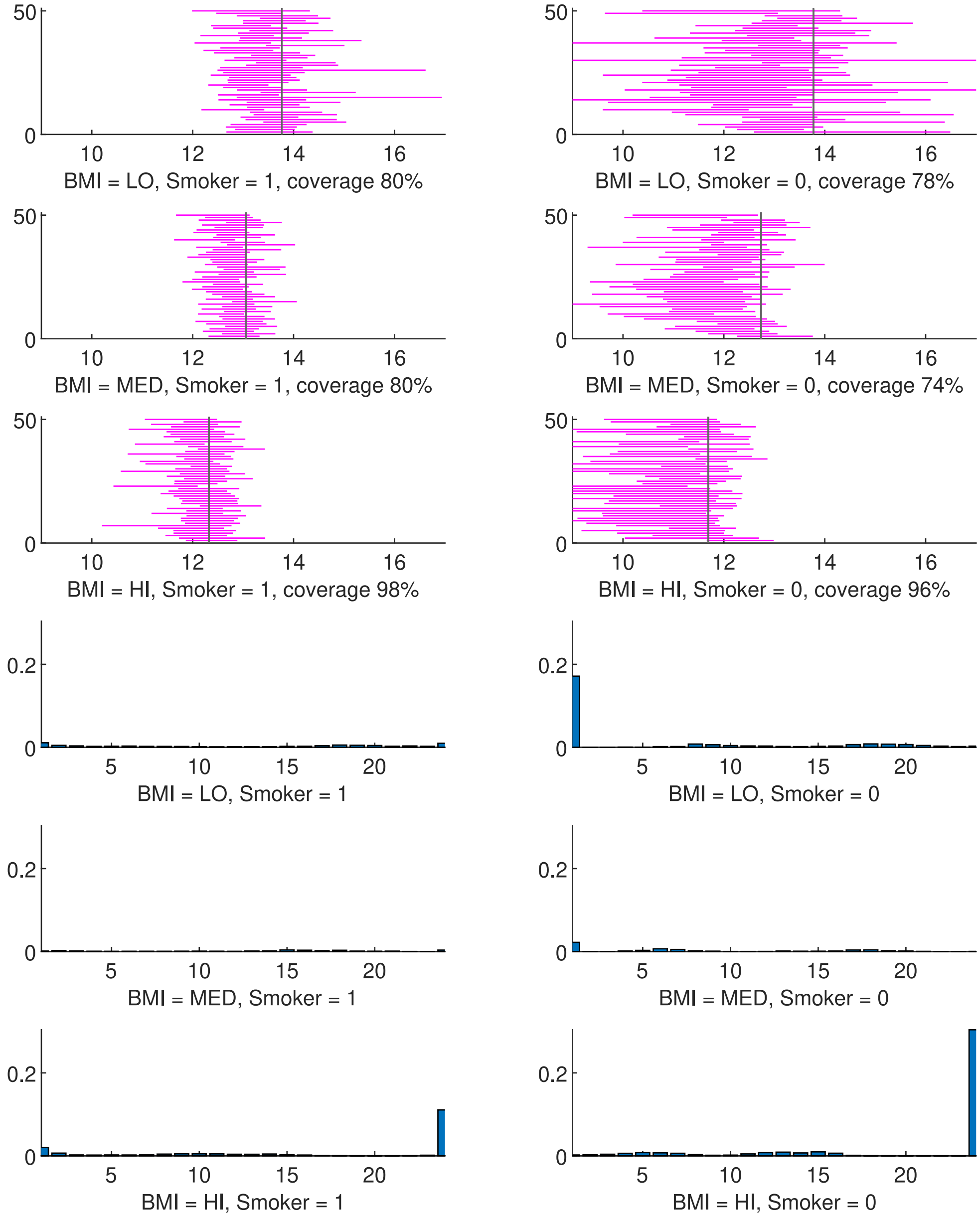

Fig. 6. Example 2. Confidence intervals and mean design weights; $(\nu_1, \nu_2) = (0, .9)$, estimation weights w_1, logistic regression.

which in (A.1) gives

$$\boldsymbol{c}_n(\boldsymbol{\theta}_0) = \boldsymbol{c}_{n,0}(\boldsymbol{\theta}_0) + O\left(n^{-1}\right). \tag{A.3}$$

Note that we have also shown that $H_\gamma\left(\beta_*\left(\boldsymbol{x}_i; \boldsymbol{\theta}_\gamma\right)\right) - G\left(\beta\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right) = O\left(n^{-1/2}\right)$. Then

$$\begin{aligned}
&\boldsymbol{M}_n(\boldsymbol{\theta}_0) \overset{def}{=} E\left[-\dot{\boldsymbol{S}}_n(\boldsymbol{\theta}_0)\right] \\
&= \sum_{i=1}^{N} \xi_n(\boldsymbol{x}_i)\, \alpha(\boldsymbol{x}_i; \boldsymbol{\theta}_0)\, \boldsymbol{f}(\boldsymbol{x}_i)\, \boldsymbol{f}'(\boldsymbol{x}_i) \\
&\quad - \sum_{i=1}^{N} \xi_n(\boldsymbol{x}_i)\left(H_\gamma\left(\beta_*\left(\boldsymbol{x}_i; \boldsymbol{\theta}_\gamma\right)\right) - G\left(\beta\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right)\right)\left\{\begin{array}{c} w(\boldsymbol{x}_i; \boldsymbol{\theta}_0)\, \dot{r}\left(\beta\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right) \\ +r\left(\beta\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right) \dot{w}(\boldsymbol{x}_i; \boldsymbol{\theta}_0) \end{array}\right\} \boldsymbol{f}'(\boldsymbol{x}_i)
\end{aligned}$$

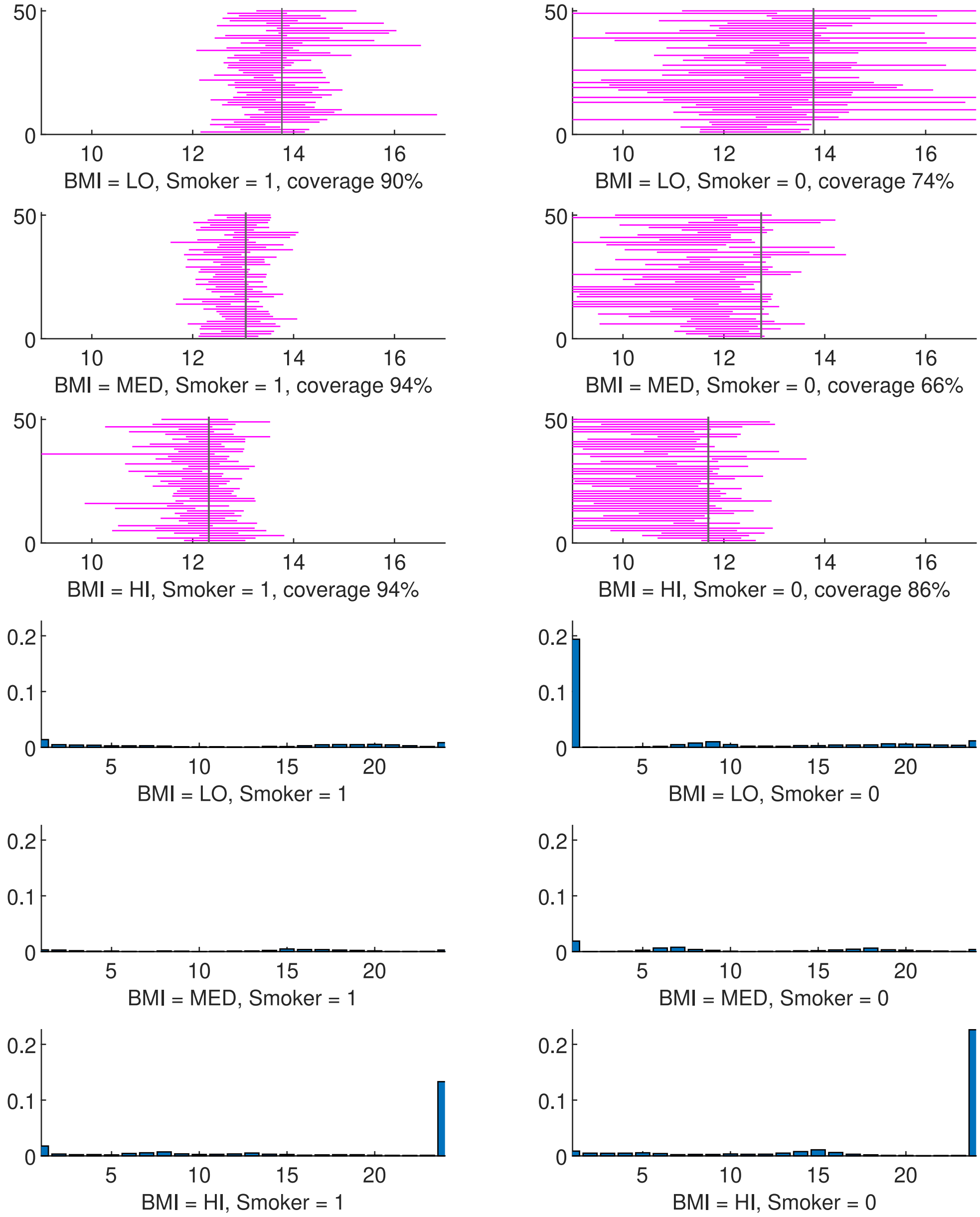

Fig. 7. Example 2. Confidence intervals and mean design weights; $(\nu_1, \nu_2) = (.45, .45)$, estimation weights w_1, logistic regression.

$$= \boldsymbol{M}_{n,0}(\boldsymbol{\theta}_0) + O\left(n^{-1/2}\right), \tag{A.4}$$

and $\boldsymbol{V}_n(\boldsymbol{\theta}_0) \stackrel{def}{=} \mathrm{cov}\left[\sqrt{n}\boldsymbol{S}_n(\boldsymbol{\theta}_0)\right]$

$$= \sum_{i=1}^{N} \xi_n(\boldsymbol{x}_i)\, w^2(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\, v_*\left(\boldsymbol{x}_i;\boldsymbol{\theta}_\gamma\right) r^2\left(\beta(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\right)\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$$

$$= \boldsymbol{V}_{n,0}(\boldsymbol{\theta}_0) + O\left(n^{-1/2}\right). \tag{A.5}$$

Note that, as well, $\boldsymbol{M}_n(\boldsymbol{\theta}) = -\partial\boldsymbol{c}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$.

Under conditions as in Fahrmeir (1990) there is a zero $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{S}_n(\boldsymbol{\theta})$ which is $\sqrt{n}$-consistent for a zero $\boldsymbol{\theta}_n^*$ of $\boldsymbol{c}_n(\boldsymbol{\theta})$. We are interested in the asymptotic distribution of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$, for which we assume that Θ is large enough to include both $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_n^*$. By the Mean Value Theorem,

$$\boldsymbol{c}_n\left(\boldsymbol{\theta}_n^*\right)-\boldsymbol{c}_n\left(\boldsymbol{\theta}_0\right)=\left(\frac{\partial \boldsymbol{c}_n\left(\tilde{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}_n^*-\boldsymbol{\theta}_0\right)=-\boldsymbol{M}_n\left(\tilde{\boldsymbol{\theta}}_n\right)\left(\boldsymbol{\theta}_n^*-\boldsymbol{\theta}_0\right),$$

for $\tilde{\boldsymbol{\theta}}_n$ on the line joining $\boldsymbol{\theta}_n^*$ to $\boldsymbol{\theta}_0$. By the compactness of Θ, $\alpha\left(\boldsymbol{x}_i;\boldsymbol{\theta}\right)$ is bounded away from 0 on this line and so, by (15) and (A.4), we may assume that $\boldsymbol{M}_n\left(\tilde{\boldsymbol{\theta}}_n\right)$ is positive definite. Thus

$$\left(\boldsymbol{\theta}_n^*-\boldsymbol{\theta}_0\right)=-\boldsymbol{M}_n^{-1}\left(\tilde{\boldsymbol{\theta}}_n\right)\left(\boldsymbol{c}_n\left(\boldsymbol{\theta}_n^*\right)-\boldsymbol{c}_n\left(\boldsymbol{\theta}_0\right)\right)=O\left(n^{-1/2}\right). \tag{A.6}$$

Now let $\boldsymbol{\delta}_i\left(\boldsymbol{\theta}\right)$ be the $d^2\times 1$ vector of second partial derivatives of the ith component of $\dot{\boldsymbol{l}}_n$, arranged appropriately. Then an expansion of the likelihood equation around $\boldsymbol{\theta}_0$ gives

$$\mathbf{0}=\boldsymbol{S}_n\left(\hat{\boldsymbol{\theta}}_n\right)=\boldsymbol{S}_n\left(\boldsymbol{\theta}_0\right)+\dot{\boldsymbol{S}}_n\left(\boldsymbol{\theta}_0\right)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)+\begin{pmatrix}\boldsymbol{\delta}_1'\left(\tilde{\boldsymbol{\theta}}_{n,1}\right)\\ \vdots \\ \boldsymbol{\delta}_d'\left(\tilde{\boldsymbol{\theta}}_{n,d}\right)\end{pmatrix}\left(\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\otimes\mathbf{I}_d\right)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right),$$

for $\tilde{\boldsymbol{\theta}}_{n,j}$ on the line segment between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, for each j. Thus

$$\sqrt{n}\boldsymbol{S}_n\left(\boldsymbol{\theta}_0\right)=-\left[\dot{\boldsymbol{S}}_n\left(\boldsymbol{\theta}_0\right)+\begin{pmatrix}\boldsymbol{\delta}_1'\left(\tilde{\boldsymbol{\theta}}_{n,1}\right)\\ \vdots \\ \boldsymbol{\delta}_d'\left(\tilde{\boldsymbol{\theta}}_{n,d}\right)\end{pmatrix}\left(\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\otimes\mathbf{I}_d\right)\right]\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right).$$

From (A.6) and the fact that $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n^*\right)$ is $O_p(1)$ we conclude that $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)$ is $O_p(1)$; this together with the fact that the elements of the $\delta_i\left(\boldsymbol{\theta}\right)$ are uniformly bounded – this follows from our assumption that g possesses two bounded derivatives – yields that

$$\sqrt{n}\boldsymbol{S}_n\left(\boldsymbol{\theta}_0\right)=-\left[\dot{\boldsymbol{S}}_n\left(\boldsymbol{\theta}_0\right)+\boldsymbol{R}_n\right]\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)+o_p(1),$$

where $\boldsymbol{R}_n$ is $O_P\left(n^{-1/2}\right)$.

By the Central Limit Theorem, $\sqrt{n}\boldsymbol{S}_n\left(\boldsymbol{\theta}_0\right)$ is asymptotically normal, with asymptotic mean $\sqrt{n}\boldsymbol{c}_n\left(\boldsymbol{\theta}_0\right)$ and asymptotic covariance $\boldsymbol{V}_n\left(\boldsymbol{\theta}_0\right)$. By (A.4) and Chebyshev's Inequality, we may conclude that

$$-\dot{\boldsymbol{S}}_n\left(\boldsymbol{\theta}_0\right)=\boldsymbol{M}_{n,0}\left(\boldsymbol{\theta}_0\right)+o_p\left(1\right), \tag{A.7}$$

as long as $\mathrm{VAR}\left[\boldsymbol{s}'\left(-\dot{\boldsymbol{S}}_n\left(\boldsymbol{\theta}_0\right)\right)\boldsymbol{t}\right]\rightarrow 0$ for all $\boldsymbol{s},\boldsymbol{t}$. This is

$$\begin{aligned}&\mathrm{VAR}\sum_{i=1}^N\xi_n\left(\boldsymbol{x}_i\right)\left\{\frac{\partial}{\partial\boldsymbol{\theta}_0}w\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)r\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\right\}y_i\boldsymbol{s}'\boldsymbol{f}\left(\boldsymbol{x}_i\right)\boldsymbol{t}'\boldsymbol{f}\left(\boldsymbol{x}_i\right)\\ &=\frac{1}{n}\sum_{i=1}^N\xi_n\left(\boldsymbol{x}_i\right)\left\{\frac{\partial}{\partial\boldsymbol{\theta}_0}w\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)r\left(\beta\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\right\}^2 v_*\left(\boldsymbol{x}_i;\boldsymbol{\theta}_\gamma\right)\left(\boldsymbol{s}'\boldsymbol{f}\left(\boldsymbol{x}_i\right)\boldsymbol{t}'\boldsymbol{f}\left(\boldsymbol{x}_i\right)\right)^2,\end{aligned}$$

which is $O\left(n^{-1}\right)$. Now (16) follows from (A.3)–(A.5) and (A.7).

(2) Now (17) and (18) follow from straightforward calculations; the latter uses (8).

(3) By the delta method applied to (16), with $\boldsymbol{\mu}_n\overset{def}{=}\boldsymbol{\theta}_0+\boldsymbol{M}_{n,0}^{-1}\left(\boldsymbol{\theta}_0\right)\boldsymbol{c}_{n,0}\left(\boldsymbol{\theta}_0\right)$,

$$\sqrt{n}\left(\rho\left(\hat{\boldsymbol{\theta}}_n\right)-\rho\left(\boldsymbol{\mu}_n\right)\right)\sim AN\left(\mathbf{0},\dot{\boldsymbol{\rho}}'\left(\boldsymbol{\mu}_n\right)\boldsymbol{M}_{n,0}^{-1}\left(\boldsymbol{\theta}_0\right)\boldsymbol{V}_{n,0}\left(\boldsymbol{\theta}_0\right)\boldsymbol{M}_{n,0}^{-1}\left(\boldsymbol{\theta}_0\right)\dot{\boldsymbol{\rho}}\left(\boldsymbol{\mu}_n\right)\right).$$

But $\rho\left(\boldsymbol{\mu}_n\right)=\rho\left(\boldsymbol{\theta}_0\right)+\dot{\boldsymbol{\rho}}'\left(\boldsymbol{\theta}_0\right)\boldsymbol{M}_{n,0}^{-1}\left(\boldsymbol{\theta}_0\right)\boldsymbol{c}_{n,0}\left(\boldsymbol{\theta}_0\right)+o\left(n^{-1/2}\right)$ and $\dot{\boldsymbol{\rho}}\left(\boldsymbol{\mu}_n\right)=\dot{\boldsymbol{\rho}}\left(\boldsymbol{\theta}_0\right)+o(1)$, whence (19) follows. □

Proof of Theorem 2. Recall (25), and note that

$$\boldsymbol{U}_*\left(\xi_n;\boldsymbol{\theta}_0\right)=\boldsymbol{S}'\left(\xi_n;\boldsymbol{\theta}_0\right)\boldsymbol{A}^{-2}\left(\xi_n;\boldsymbol{\theta}_0\right)\boldsymbol{S}\left(\xi_n;\boldsymbol{\theta}_0\right).$$

Conditions (4) and (8) force $\boldsymbol{\psi} = \left(\tau_1/\sqrt{n}\right)\mathbf{Q}_2\boldsymbol{u}$, for some $\boldsymbol{u}$ with $\|\boldsymbol{u}\| \le 1$. Recall that $\mathbf{Q}_2'\boldsymbol{h}(\boldsymbol{\theta}_0) \stackrel{def}{=} \boldsymbol{b}(\boldsymbol{\theta}_0)$. Then in (18),

$$\begin{aligned}\sqrt{n}\boldsymbol{F}\boldsymbol{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\,\boldsymbol{c}_{n,0}(\boldsymbol{\theta}_0) &= \mathbf{Q}_1\boldsymbol{A}^{-1}(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{S}(\xi_n;\boldsymbol{\theta}_0)\left(\tau_1\boldsymbol{u} - \sqrt{n}\boldsymbol{b}(\boldsymbol{\theta}_0)\,\gamma\right),\\ \boldsymbol{F}\boldsymbol{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\,\boldsymbol{V}_{n,0}(\boldsymbol{\theta}_0)\,\boldsymbol{M}_{n,0}^{-1}(\boldsymbol{\theta}_0)\,\boldsymbol{F}' &= \mathbf{Q}_1\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0)\,\mathbf{Q}_1',\end{aligned}$$

and so

$$\mathcal{L}_\beta(\xi_n;\psi,\gamma) = tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0) + \left(\tau_1\boldsymbol{u} - \sqrt{n}\boldsymbol{b}(\boldsymbol{\theta}_0)\,\gamma\right)'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\left(\tau_1\boldsymbol{u} - \sqrt{n}\boldsymbol{b}(\boldsymbol{\theta}_0)\,\gamma\right).$$

We maximize over γ first, subject to (5). The quadratic form is

$$\tau_1^2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{u} - 2\tau_1\sqrt{n}\gamma\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b} + \left(\sqrt{n}\gamma\right)^2\boldsymbol{b}'(\boldsymbol{\theta}_0)\,\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b}(\boldsymbol{\theta}_0).$$

This becomes larger if the sign of $\boldsymbol{u}$ is chosen such that

$$\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b}(\boldsymbol{\theta}_0) \le \mathbf{0}, \tag{A.8}$$

which is now assumed. Then the maximum is attained at $\gamma = \tau_2/\sqrt{n}$, with

$$\max_\gamma \mathcal{L}_\beta(\xi_n;\psi,\gamma) = tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0) + \left\{\begin{array}{c}\tau_1^2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{u} - 2\tau_1\tau_2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b}\\ +\tau_2^2\boldsymbol{b}'(\boldsymbol{\theta}_0)\,\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b}(\boldsymbol{\theta}_0)\end{array}\right\}. \tag{A.9}$$

By virtue of (A.8), a maximizing $\boldsymbol{u}$ will have maximum norm. We introduce a Lagrange multiplier λ and maximize

$$\tau_1^2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{u} - 2\tau_1\tau_2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{b}(\boldsymbol{\theta}_0) - \lambda\left(\boldsymbol{u}'\boldsymbol{u} - 1\right).$$

This leads to (21) and – upon imposing the side condition – to (22), and then to (20).
(i) If $\tau_2 = 0$ then

$$\mathcal{L}_\beta(\xi_n;\psi) = tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0) + \tau_1^2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{u},$$

and $\boldsymbol{u}_0 = \arg\max_{\|\boldsymbol{u}\|=1}\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{u}$ is the eigenvector, of unit norm, belonging to the maximum characteristic root of $\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)$, with $\max_\psi \mathcal{L}_\beta(\xi_n;\psi) = tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0) + \tau_1^2 ch_{\max}\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)$. By virtue of the fact that products $\boldsymbol{P}\boldsymbol{P}'$ and $\boldsymbol{P}'\boldsymbol{P}$ have the same non-zero characteristic roots, we have

$$ch_{\max}\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0) = ch_{\max}\boldsymbol{A}^{-1}(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{S}(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{S}'(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{A}^{-1}(\xi_n;\boldsymbol{\theta}_0).$$

The identity $\mathbf{Q}_2\mathbf{Q}_2' = \mathbf{I}_N - \mathbf{Q}_1\mathbf{Q}_1'$ gives that

$$\boldsymbol{S}(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{S}'(\xi_n;\boldsymbol{\theta}_0) = \mathbf{Q}_1'\left[\boldsymbol{D}_\xi\boldsymbol{D}_\alpha(\boldsymbol{\theta}_0)\right]^2\mathbf{Q}_1 - \boldsymbol{A}^2(\xi_n;\boldsymbol{\theta}_0),$$

whence

$$ch_{\max}\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0) = ch_{\max}\boldsymbol{U}(\xi_n;\boldsymbol{\theta}_0) - 1,$$

yielding (23).
(ii) is immediate. □

Remark 11. If the mislabelling probabilities are allowed to be $\boldsymbol{x}$-dependent, with $\gamma(\boldsymbol{x}_i) = \left(\tau_2/\sqrt{n}\right)v_i$ for $\boldsymbol{v} \stackrel{def}{=} (v_1,\ldots,v_N)' \in [0,1]^N$, then the analogue of (A.9) is

$$\max_{\gamma,\psi}\mathcal{L}_\beta(\xi_n;\psi,\gamma) = tr\boldsymbol{R}^{-1}(\xi_n;\boldsymbol{\theta}_0) + \max_{\boldsymbol{u},\boldsymbol{v}}\left\{\begin{array}{c}\tau_1^2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\boldsymbol{u} - 2\tau_1\tau_2\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\mathbf{Q}_2'\boldsymbol{D}_h(\boldsymbol{\theta}_0)\,\boldsymbol{v}\\ +\tau_2^2\boldsymbol{v}'\boldsymbol{D}_h(\boldsymbol{\theta}_0)\,\mathbf{Q}_2\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\mathbf{Q}_2'\boldsymbol{D}_h(\boldsymbol{\theta}_0)\,\boldsymbol{v}\end{array}\right\},$$

where $\boldsymbol{D}_h(\boldsymbol{\theta}_0) \stackrel{def}{=} diag(\ldots,h_i(\boldsymbol{\theta}_0),\ldots)$ and the sign of $\boldsymbol{u}$ will necessarily be such that $\boldsymbol{u}'\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\,\mathbf{Q}_2'\boldsymbol{D}_h(\boldsymbol{\theta}_0)\,\boldsymbol{v} \le 0$. The maximization over $\boldsymbol{v}$ requires a quadratic programme to be solved, repeatedly. The problem can also be written as

$$\max_{\|\boldsymbol{u}\|\le 1,\boldsymbol{v}\in[0,1]^N}(\boldsymbol{u},\boldsymbol{v})'\begin{pmatrix}\tau_1\boldsymbol{I}_{N-d}\\ -\tau_2\boldsymbol{D}_h(\boldsymbol{\theta}_0)\,\mathbf{Q}_2\end{pmatrix}\boldsymbol{U}_*(\xi_n;\boldsymbol{\theta}_0)\begin{pmatrix}\tau_1\boldsymbol{I}_{N-d} & -\tau_2\mathbf{Q}_2'\boldsymbol{D}_h(\boldsymbol{\theta}_0)\end{pmatrix}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix}.$$

Proof of Theorem 3. For brevity we write ξ_i for $\xi(\boldsymbol{x}_i)$, and suppress the dependence on $\boldsymbol{\theta}_0$ when there is no risk of confusion. First take $k = 1$.
(1), (2): A necessary condition for a minimum of $\mathcal{L}_{\nu_1,\nu_2}(\xi;\boldsymbol{\theta}_0)$ at ξ_0 is that, for any design ξ_1 and with $\xi_t = (1-t)\,\xi_0 + t\xi_1$,

$$0 \le \frac{d}{dt}\mathcal{L}_{\nu_1,\nu_2}(\xi_t;\boldsymbol{\theta}_0)_{|t=0} = (1-\nu_1-\nu_2)\frac{d}{dt}tr\boldsymbol{R}^{-1}(\xi_t)_{|t=0} + \nu_1\frac{d}{dt}\lambda(\xi_t)_{|t=0} + \nu_2\frac{d}{dt}\boldsymbol{b}'\boldsymbol{U}_*(\xi_t)\,\boldsymbol{b}_{|t=0}. \tag{A.10}$$

We calculate – details in the online Appendix 1 – that

$$\frac{d}{dt}tr\boldsymbol{R}^{-1}(\xi_t)_{|t=0} = tr\boldsymbol{R}^{-1}(\xi_0) - tr\left\{2\boldsymbol{R}^{-1}(\xi_0)\,\boldsymbol{A}(\xi_1)\,\boldsymbol{A}^{-1}(\xi_0) - \boldsymbol{A}^{-1}(\xi_0)\,\boldsymbol{A}(\xi_1)\,\boldsymbol{A}^{-1}(\xi_0)\right\},$$

and, with $\boldsymbol{P}(\xi) \overset{def}{=} \boldsymbol{A}^{-1}(\xi)\boldsymbol{S}(\xi) : d \times N - d$, that

$$\frac{d}{dt}\boldsymbol{b}'\boldsymbol{U}_*(\xi_t)\boldsymbol{b}_{|_{t=0}} = 2\boldsymbol{b}'\boldsymbol{P}'(\xi_0)\boldsymbol{A}^{-1}(\xi_0)\{\boldsymbol{S}(\xi_1) - \boldsymbol{A}(\xi_1)\boldsymbol{P}(\xi_0)\}\boldsymbol{b}.$$

By Theorem 1 of Magnus (1985), $\lambda(\xi_t)$ is differentiable, with

$$\begin{aligned}\frac{d}{dt}\lambda(\xi_t)_{|_{t=0}} &= \boldsymbol{z}'(\xi_0)\left[\frac{d}{dt}\boldsymbol{U}(\xi_t)\right]_{|_{t=0}}\boldsymbol{z}(\xi_0)\\ &= -2\boldsymbol{c}'(\xi,\boldsymbol{\theta}_0)\left\{\boldsymbol{A}(\xi_1)\boldsymbol{A}^{-1}(\xi_0)\mathbf{Q}_1'\boldsymbol{D}_{\xi_0}^2\boldsymbol{D}_\alpha^2(\boldsymbol{\theta}_0)\mathbf{Q}_1 - \mathbf{Q}_1'\boldsymbol{D}_{\xi_1}\boldsymbol{D}_\alpha^2(\boldsymbol{\theta}_0)\boldsymbol{D}_{\xi_0}\mathbf{Q}_1\right\}\boldsymbol{c}(\xi,\boldsymbol{\theta}_0).\end{aligned}$$

Recall that $\{\boldsymbol{q}_i'\}_{i=1}^N$ are the rows of $\mathbf{Q}_1$; let $\{\boldsymbol{e}_i'\}_{i=1}^N$ be the rows of $\mathbf{I}_N$, and substitute

$$\boldsymbol{A}(\xi_1) = \sum_{i=1}^N \xi_{1,i}\alpha_i\boldsymbol{q}_i\boldsymbol{q}_i',\ \boldsymbol{B}(\xi_1) = \sum_{i=1}^N \xi_{1,i}\alpha_i w_i\boldsymbol{q}_i\boldsymbol{q}_i',\ \boldsymbol{D}_{\xi_1} = \sum_{i=1}^N \xi_{1,i}\boldsymbol{e}_i\boldsymbol{e}_i',\ \boldsymbol{S}(\xi_1) = \sum_{i=1}^N \xi_{1,i}\boldsymbol{q}_i\boldsymbol{e}_i'\boldsymbol{D}_\alpha\mathbf{Q}_2,$$

into the expressions above. After suitable rearrangements we obtain

$$\frac{d}{dt}\mathcal{L}_{\nu_1,\nu_2}(\xi_t;\boldsymbol{\theta}_0)_{|_{t=0}} = (1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\boldsymbol{\xi}_0) - \sum_{i=1}^N \xi_{1,i}\boldsymbol{T}_{ii}(\boldsymbol{\xi}_0). \tag{A.11}$$

Thus (A.10) is equivalent to

$$\sum_{i=1}^N \xi_{1,i}\left\{\frac{\boldsymbol{T}_{ii}(\boldsymbol{\xi}_0)}{(1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\boldsymbol{\xi}_0)} - 1\right\} \le 0,\ \text{for all designs } \boldsymbol{\xi}_1. \tag{A.12}$$

Some algebra confirms that $tr\boldsymbol{D}_{\boldsymbol{\xi}}\boldsymbol{T}(\boldsymbol{\xi}) = (1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\boldsymbol{\xi})$ for any design $\boldsymbol{\xi}$, so that

$$\pi(\boldsymbol{\xi}) \ge \sum_{i=1}^N \xi_i\left\{\frac{\boldsymbol{T}_{ii}(\boldsymbol{\xi})}{(1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\boldsymbol{\xi})} - 1\right\} = \frac{tr\boldsymbol{D}_{\boldsymbol{\xi}}\boldsymbol{T}(\boldsymbol{\xi})}{(1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\boldsymbol{\xi})} - 1 = 0. \tag{A.13}$$

Condition (A.12) implies that

$$\boldsymbol{T}_{ii}(\boldsymbol{\xi}_0)/\left[(1-\nu_1-\nu_2)\,tr\boldsymbol{R}^{-1}(\boldsymbol{\xi}_0)\right] - 1 \le 0 \text{ for all } i;$$

this together with (A.13) applied to $\boldsymbol{\xi}_0$ implies that equality must hold at the points at which $\xi_{0,i} > 0$ and that $\pi(\boldsymbol{\xi}_0) = 0$. Conversely, if $\pi(\boldsymbol{\xi}_0) = 0$ then the term in braces in (A.13) must vanish on the support of $\boldsymbol{\xi}_0$.
(3): An expansion of $\mathcal{L}_{\nu_1,\nu_2}(\xi_t;\boldsymbol{\theta}_0)$ gives

$$\mathcal{L}_{\nu_1,\nu_2}(\xi_t;\boldsymbol{\theta}_0) = \mathcal{L}_{\nu_1,\nu_2}(\xi_0;\boldsymbol{\theta}_0) + t\frac{d}{dt}\mathcal{L}_{\nu_1,\nu_2}(\xi_t;\boldsymbol{\theta}_0)_{|_{t=0}} + O\left(t^2\right).$$

With $t = 1/(n+1)$, $\xi_0 = \xi_n$, $\xi_1 = \delta(\boldsymbol{x}_i)$ we have $\xi_t = \xi_{n+1}^{(i)}$, and then (A.11) gives (27), from which (28) is immediate. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2021.107189.

References

Abdelbasit, K.M., Plackett, R.L., 1983. Experimental design for binary data. J. Amer. Statist. Assoc. 78, 90–98.
Adewale, A., Wiens, D.P., 2009. Robust designs for misspecified logistic models. J. Stat. Plan. Inference 139, 3–15, Special Issue on "Metaheuristics, Combinatorial Optimization and Design of Experiments".
Adewale, A., Xu, X., 2009. Robust designs for generalized linear models with possible overdispersion and misspecified link functions. Comput. Statist. Data Anal. 54, 875–890.
Biedermann, S., Dette, H., Pepelyshev, A., 2006. Some robust design strategies for percentile estimation in binary response models. Canad. J. Statist. 4, 535–561.
Bretz, F., Dette, H., Pinheiro, J., 2010. Practical considerations for optimal designs in clinical dose finding studies. Stat. Med. 29, 731–742.
Carroll, R., Pederson, S., 1993. On robustness in the logistic regression model. J. R. Stat. Soc. Ser. B Stat. Methodol. 55, 693–706.
Copas, J.P., 1988. Binary regression models for contaminated data. J. R. Stat. Soc. Ser. B Stat. Methodol. 50, 225–265.
Dette, H., Bretz, F., Pepelyshev, A., Pinheiro, J., 2008. Optimal designs for dose-finding studies. J. Amer. Statist. Assoc. 103, 1225–1237.
Edvardsson, E., Lendahls, L., Håkannson, A., 2009. When do adolescents become smokers? Scand. J. Primary Health Care 27, 41–46.
Fahrmeir, L., 1990. Maximum likelihood estimation in misspecified generalized linear models. Statistics 21, 487–502.
Feller, C., Schorning, K., Dette, H., Behrmann, G., Bornkamp, B., 2017. Optimal designs for dose response curves with common parameters. Ann. Statist. 45, 2102–2132.
Holland-Letz, T., Kopp-Schneider, A., 2015. Optimal experimental designs for dose–response studies with continuous endpoints. Arch. Toxicol. 89, 2059–2068.

Huang, H., Chen, X., 2021. Compromise design for combination experiment of two drugs. Comput. Statist. Data Anal. 157, 107–150.
Jacobs, M., 2019. Adolescent smoking: The relationship between cigarette consumption and bmi. Addict. Behav. Rep. 9 (10015), 3.
Khuri, A.I., Mukherjee, B., Sinha, B.K., Ghosh, M., 2006. Design issues for generalized linear models: a review. Statist. Sci. 21, 376–399.
Lange, M.R., Schmidli, H., 2014. Optimal design of clinical trials with biologics using dose-time-response models. Stat. Med. 33, 5249–5264.
Li, P., Wiens, D.P., 2011. Robustness of design in dose-response studies. J. R. Stat. Soc. Ser. B Stat. Methodol. 17, 215–238.
López-Fidalgo, J., Wiens, D.P., 2020. Robust Active Learning with Binary Responses. University of Alberta Technical Report, URL www.ualberta.ca/~dwiens/.
Lundblad, M.W., Jacobsen, B.K., 2017. The reproducibility of self-reported age at menarche: the Trømsø study. BMC Women's Health 17 (62), 179–191.
Magnus, J.R., 1985. On differentiating eigenvalues and eigenvectors. Econometric Theory 1, 179–191.
Milicer, H., Szczotka, F., 1966. Age at menarche in warsaw girls in 1965. Human Biol. 40, 249–259.
Mohamad, K., Jamshidi, L., Nouri Jelyani, K., 2013. Is age of menarche related with body mass index? Iran. J. Public Health 42, 1043–1048.
Müller, H.-G., Schmitt, T., 1990. Choice of number of doses for maximum likelihood estimation of the ED50 for quantal dose-response data. Biometrics 46, 117–129.
Wiens, D.P., 2018. I-robust and D-robust designs on a finite design space. Stat. Comput. 28, 241–258.
Woods, D.C., Lewis, S.M., Eccleston, J.A., Russell, K.G., 2006. Designs for generalized linear models with several variables and model uncertainty. Technometrics 48, 84–292.
Xu, X., Sinha, S.K., 2021. Robust designs for generalized linear mixed models with possible model misspecification. J. Statist. Plann. Inference 210, 20–41.