Journal of Multivariate Analysis 182 (2021) 104712

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Optimal designs for mixed continuous and binary responses with quantitative and qualitative factors

Ming-Hung Kao *, Hazar Khogeer

School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, 85287-1804, USA

ARTICLE INFO

Article history:
Received 21 March 2020
Received in revised form 2 December 2020
Accepted 2 December 2020
Available online 9 December 2020

AMS 2010 subject classifications:
62H99
62K05

Keywords:
Approximate design
Chebyshev system
Complete class
Optimality criteria
Sequential quadratic programming

ABSTRACT

This work is concerned with optimal designs for multivariate regression of responses of mixed variable types (continuous and binary) on quantitative and qualitative factors. New complete class results with respect to the Loewner ordering, and relevant Chebyshev systems are derived to identify a small class of designs, within which locally optimal designs can be found for a group of models and optimality criteria. The complete class results facilitate the search of optimal designs via some general-purpose optimization techniques. Extensions of some previous results for characterizing optimal designs are also provided.

1. Introduction

Experiments involving two or more response variables are ubiquitous. For cases with only continuous responses, optimal designs for multivariate regression models are developed in several previous works; see, e.g., [19,24,35], and Chapter 5 of [11]. These designs allow experimenters to collect informative data for making precise and valid statistical inferences; they also can serve as benchmarks for evaluating the quality of the designs selected by the experimenter. However, there exist many cases where some of the multivariate responses are continuous, whereas the others are, say, categorical. Examples of this sort, and some multivariate data analysis methods to jointly analyze responses of mixed variable types can be found in, e.g., [32] and [6]. For such a situation, the optimal designs developed in the previously mentioned studies can be inappropriate, and the identification of new optimal designs is needed.

This work is concerned with optimal designs for multivariate regression analysis of responses of mixed variable types. For convenience, we refer to such multivariate responses as mixed responses, and a corresponding model is termed as a mixed responses model. Two common likelihood-based approaches for building mixed responses models include the latent variable methods and factorization methods [5,32]. The former approach assumes that the categorical responses are induced from some continuous, unobserved latent variables. With this approach, Fedorov et al. [12] obtained optimal designs for models where the binary response (z) is assumed to be a dichotomization of a continuous latent variable whose joint distribution with the continuous response (y) is a bivariate normal distribution. For the factorization method, the joint distribution of y and z is expressed as $f(y, z) = f_z(z)f_{y|z}(y|z)$ for specific marginal distribution $f_z(z)$ and conditional distribution $f_{y|z}(y|z)$ [5,7,14]. Recently, Kim and Kao [23] proposed some optimal designs for a mixed responses regression

* Corresponding author.
E-mail address: mkao3@asu.edu (M.-H. Kao).

https://doi.org/10.1016/j.jmva.2020.104712

model that is built based on the factorization method. The obtained optimal designs determine the optimal set of distinct values of a continuous covariate (x) (e.g., dose levels in a dose–response study), and the frequency of occurrences of each x-value in the experiment.

Following this research direction, we develop some optimal design results for mixed responses regression models under the factorization approach. For clarity, we put our focus on cases with one continuous response and one binary response. But in contrast to the previous study, we allow our models to include both quantitative and qualitative factors; the qualitative factors divide experimental subjects into homogeneous subject groups. In addition, our results can be applied to a larger collection of models than that of [23]. These models, which include some popularly used ones, may or may not assume common parameters across submodels (to be defined in Section 2), and/or across subject groups. The main tool that we adapt to facilitate the identification of optimal designs is the complete class approach by Yang and Stufken [34]. Roughly speaking, this approach gives a small class of designs within which an optimal design can be found. It often greatly reduces the number of decision variables in the optimization problem to allow or facilitate the use of some well-developed optimization techniques such as those considered in [16] and [25].

The previously mentioned complete class approach of [34] is with respect to the Loewner ordering, and can thus be considered for most commonly used optimality criteria. In addition to obtaining a core of the information matrix (see Definition 1), a key ingredient of this approach is to identify vectors of functions that form a Chebyshev system (Definition 2). Numerous previous works provide Chebyshev systems useful for deriving optimal designs; e.g., [10,11,17,18,29,33,34,36]. Here, we add to this literature by presenting some additional Chebyshev systems that can be used to derive complete classes for mixed responses models. The result can also be applied in other settings where the elements of the information matrix or its core lie within the space spanned by the Chebyshev systems (Theorem 3). We also provide useful extensions of previous results to give insights into some properties of optimal designs. Optimal designs are then identified by using a well-developed computational approach. We note that optimal designs in the current setting depend on some model parameters, possibly including error variances. To address this issue, we work on locally optimal designs [4] that are optimal for given parameter values; see, e.g., [15] for the usefulness of locally optimal designs. We also follow previous design works to consider the approximate design approach in the sense of Kiefer [22]; i.e., the relative frequency of appearances of a design point can be any real value between 0 and 1. A rounding of the obtained approximate design may then be considered to obtain an exact design for a given sample size; e.g., [28].

In Section 2, we introduce our mixed responses models. Our developed Chebyshev systems, complete classes, and related results can be found in Section 3. Some optimal designs are provided in Section 4, and a conclusion is in Section 5. Proofs for some results are deferred to the Appendix.

2. Model, information matrix, and optimality criterion

Let $Y(\ell, x)$ and $Z(\ell, x)$ be respectively the continuous and binary response variables of an experimental subject in the ℓ-th subject group with a continuous covariate $X = x \in [A_\ell, B_\ell] \subset \mathbb{R}$; $\ell \in \{1, \dots, L\}$. Here, the subject groups are formed by two or more qualitative factors (and/or categorizations of some continuous variables). We consider mixed responses models of the following form:

$$\begin{aligned}
&Y(\ell, x) \mid Z(\ell, x) = z \sim \mathcal{N}(\mu_{z,\ell}(x), \sigma_z^2),\ \ \Pr\{Z(\ell, x) = z\} = P\{c_\ell(x)\}^z[1 - P\{c_\ell(x)\}]^{1-z},\ z \in \{0, 1\}, \qquad (1)\\
&\mu_{0,\ell}(x) = \mathbf{f}_*^\top(x)\boldsymbol{\beta}_* + \mathbf{f}_0^\top(x)\boldsymbol{\beta}_0 + \mathbf{h}_{*,\ell}^\top(x)\boldsymbol{\gamma}_{*,\ell} + \mathbf{h}_{0,\ell}^\top(x)\boldsymbol{\gamma}_{0,\ell},\ \ \mu_{1,\ell}(x) = \mathbf{f}_*^\top(x)\boldsymbol{\beta}_* + \mathbf{f}_1^\top(x)\boldsymbol{\beta}_1 + \mathbf{h}_{*,\ell}^\top(x)\boldsymbol{\gamma}_{*,\ell} + \mathbf{h}_{1,\ell}^\top(x)\boldsymbol{\gamma}_{1,\ell},\\
&\quad c_\ell(x) = \mathbf{f}_2^\top(x)\boldsymbol{\beta}_2 + \mathbf{h}_{2,\ell}^\top(x)\boldsymbol{\gamma}_{2,\ell}.
\end{aligned}$$

Model (1) has three submodels, including the two conditional models for $Y(\ell, x) \mid Z(\ell, x) = 0$ and $Y(\ell, x) \mid Z(\ell, x) = 1$, respectively, and the marginal model for $Z(\ell, x)$. The vectors $\mathbf{f}_*(x)$, $\mathbf{h}_{*,\ell}(x)$, $\mathbf{f}_s(x)$, and $\mathbf{h}_{s,\ell}(x)$ for $s \in \{0, 1, 2\}$ consist of given functions of x; $\boldsymbol{\beta}_s \in \mathbb{R}^{m_s}$ is the vector of unknown coefficients of $\mathbf{f}_s(x)$ for the common terms across the L subject groups; $\boldsymbol{\gamma}_{s,\ell} \in \mathbb{R}^{m_{s,\ell}}$ is the unknown parameter vector for the effect of the subject group and/or its interaction with x; $\boldsymbol{\beta}_* \in \mathbb{R}^{m_*}$ plays a similar role as $\boldsymbol{\beta}_s$, and $\boldsymbol{\gamma}_{*,\ell} \in \mathbb{R}^{m_{*,\ell}}$ has a similar definition as $\boldsymbol{\gamma}_{s,\ell}$, but they allow common terms across the first two submodels (i.e., the two conditional models). For these two submodels, σ_0^2 and σ_1^2 represent the error variances. $P(\cdot)$ in the third submodel of (1) represents some differentiable cumulative distribution function. For convenience, we write $\mathbf{f}_*$ as the set of all elements in the vector $\mathbf{f}_*(x)$. Similarly, $\mathbf{f}_s$, $\mathbf{h}_{*,\ell}$, and $\mathbf{h}_{s,\ell}$ represent the sets of elements of the corresponding vectors in (1). For model (1), we also assume that $\mathbf{f}_*$, $\mathbf{f}_s$, $\mathbf{h}_{*,\ell}$, and $\mathbf{h}_{s,\ell}$ are mutually disjoint sets for each given (s, ℓ); $s \in \{0, 1, 2\}$, $\ell \in \{1, \dots, L\}$. It is not uncommon that $\mathbf{h}_{*,\ell}$ and $\mathbf{h}_{s,\ell}$ do not vary with ℓ and/or s, but our model formulation also allow situations when they do.

Example 1. For a breast cancer study, the following mixed responses model is considered in [1]:

$$Y \mid Z = z \sim \begin{cases} \mathcal{N}(\beta_{0,0} + \beta_{*,1}x + \beta_{*,2}x^2, \sigma^2), & z = 0,\\ \mathcal{N}(\beta_{1,0} + \beta_{*,1}x + \beta_{*,2}x^2, \sigma^2), & z = 1,\end{cases}$$

$$\Pr(Z = 1) = \frac{1}{1 + \exp(\beta_{2,0} + \beta_{2,1}x)}.$$

Here, x is the dose of a treatment, Y is the continuous efficacy response, and Z is the presence of an important adverse effect. For this model, we set $L=1$, $\mathbf{f}_*=\{x,x^2\}$, $\boldsymbol{\beta}_*=(\beta_{*,1},\beta_{*,2})^\top$, $\mathbf{f}_0=\mathbf{f}_1=\{1\}$, $c_1(x)=\mathbf{f}_2(x)^\top\boldsymbol{\beta}_2=\beta_{2,0}+\beta_{2,1}x$, $\sigma_0^2=\sigma_1^2=\sigma^2$, and the remaining terms for E(Z) and E($Y|Z$) in (1) are zero, i.e., $\mathbf{h}_{*,\ell}=\mathbf{h}_{s,\ell}=\{\emptyset\}$. Model (1) thus includes this example as a special case, and it also accommodates cases having qualitative factors such as gender, race, treatment groups, etc.

With (1), the joint probability function $f(y,z)$ for the mixed responses can be obtained by the factorization $f(y,z)=f_z(z)f_{y|z}(y|z)$. For a single subject, we then have the following elemental Fisher information matrix of the parameter vector $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\boldsymbol{\gamma}_1^\top,\dots,\boldsymbol{\gamma}_L^\top)^\top$; see also [14,23]. Here, $\boldsymbol{\beta}=(\boldsymbol{\beta}_*^\top,\boldsymbol{\beta}_0^\top,\boldsymbol{\beta}_1^\top,\boldsymbol{\beta}_2^\top)^\top$, $\boldsymbol{\gamma}_\ell^\top=(\boldsymbol{\gamma}_{*,\ell}^\top,\boldsymbol{\gamma}_{0,\ell}^\top,\boldsymbol{\gamma}_{1,\ell}^\top,\boldsymbol{\gamma}_{2,\ell}^\top)^\top\in\mathbb{R}^{m_{\cdot,\ell}}$, $m_{\cdot,\ell}=m_{*,\ell}+\sum_{s=0}^2 m_{s,\ell}$, and to save space, we write $m(a:b)=m_{\cdot,a}+\cdots+m_{\cdot,b}$.

$$\mathbb{M}(\boldsymbol{\theta};\ell,x)=\sum_{s=0}^{2}\Gamma_s\{c_\ell(x)\}\mathbf{g}_{s,\ell}(x)\mathbf{g}_{s,\ell}^\top(x),\;\;\Gamma_0(c)=\frac{1-P(c)}{\sigma_0^2},\;\;\Gamma_1(c)=\frac{P(c)}{\sigma_1^2},\;\;\Gamma_2(c)=\frac{\{P'(c)\}^2}{P(c)-P^2(c)},\tag{2}$$

$$\mathbf{g}_{0,\ell}(x)=(\mathbf{f}_*^\top(x),\mathbf{f}_0^\top(x),\mathbf{0}_{m_1+m_2+m(1:\ell-1)}^\top,\mathbf{h}_{*,\ell}^\top(x),\mathbf{h}_{0,\ell}^\top(x),\mathbf{0}_{m_{2,\ell}+m(\ell+1:L)}^\top)^\top,$$

$$\mathbf{g}_{1,\ell}(x)=(\mathbf{f}_*^\top(x),\mathbf{0}_{m_0}^\top,\mathbf{f}_1^\top(x),\mathbf{0}_{m_2+m(1:\ell-1)}^\top,\mathbf{h}_{*,\ell}^\top(x),\mathbf{0}_{m_{0,\ell}}^\top,\mathbf{h}_{1,\ell}^\top(x),\mathbf{0}_{m_{2,\ell}+m(\ell+1:L)}^\top)^\top,$$

$$\mathbf{g}_{2,\ell}(x)=(\mathbf{0}_{m_*+m_0+m_1}^\top,\mathbf{f}_2^\top(x),\mathbf{0}_{m(1:\ell-1)+m_{*,\ell}+m_{0,\ell}+m_{1,\ell}}^\top,\mathbf{h}_{2,\ell}^\top(x),\mathbf{0}_{m(\ell+1:L)}^\top)^\top.$$

$P'(c)$ is the first derivative of $P(c)$ with respect to c, $\mathbf{0}_a$ is the vector of a zeros; and all the remaining terms are as in (1). We note that this information matrix depends on the unknown parameters in the third submodel of (1) through $c_\ell(x)$. It also involves the error variances σ_0^2, and σ_1^2, but is free of the parameters in $\mu_{0,\ell}(x)$ and $\mu_{1,\ell}(x)$. Our aim is at an optimal design allowing the most precise inference of $\boldsymbol{\theta}$. We consider approximate designs of the following form:

$$\xi=\left\{\begin{array}{ccccccc}(1,x_{1,1}) & \cdots & (1,x_{1,n_1}) & (2,x_{2,1}) & \cdots & (L,x_{L,n_L})\\ w_{1,1} & \cdots & w_{1,n_1} & w_{2,1} & \cdots & w_{L,n_L}\end{array}\right\}.\tag{3}$$

For each ℓ, the $(\ell,x_{\ell,j})$'s in (3) are the n_ℓ distinct x-values to appear in the ℓ-th subject group; $w_{\ell,j}$ represents the proportion of appearances of $(\ell,x_{\ell,j})$; i.e., the proportion of ℓth-group subjects having $X=x_{\ell,j}$, $j\in\{1,\dots,n_\ell\}$, $\ell\in\{1,\dots,L\}$. Note that the Carathéodory theorem allows us to assume that n_ℓ is finite; see also Section 3. Following Kiefer [22], a design ξ is viewed as a probability measure, $(\ell,x_{\ell,j})$ is a support point, and $\xi(\ell,x_{\ell,j})=w_{\ell,j}$ is the corresponding weight; $w_{\ell,j}$ can be any real number between 0 and 1, but $\sum_{\ell,j}w_{\ell,j}=1$. It is useful to factorize a design as $\xi(\ell,x_{\ell,j})=\eta(\ell)\tau_\ell(x_{\ell,j})$, and write $w_{\ell,j}=w_\ell w_{j|\ell}$. We refer to η as the marginal design determining the marginal weight, $\eta(\ell)=w_\ell=\sum_{j=1}^{n_\ell}w_{\ell,j}$, for each subject group. For given ℓ, τ_ℓ is referred to as the conditional design for X within the ℓ-th group; and $\tau_\ell(x_{\ell,j})=w_{j|\ell}=w_{\ell,j}/w_\ell$. For convenience, we also write the (joint) design as $\xi=\eta\times\{\tau_\ell\}$, and note that, in contrast to [35], the conditional design τ_ℓ is allowed to vary across subject groups.

With a design ξ of (3), the information matrix for $\boldsymbol{\theta}$ is:

$$\mathbf{M}(\boldsymbol{\theta};\xi)=\sum_{\ell=1}^{L}\eta(\ell)\int_{A_\ell}^{B_\ell}\mathbb{M}(\boldsymbol{\theta};\ell,x)\,\mathrm{d}\tau_\ell(x).\tag{4}$$

We say that a design ξ_1 is at least as informative (about $\boldsymbol{\theta}$) as another design ξ_2 if $\mathbf{M}(\boldsymbol{\theta};\xi_1)-\mathbf{M}(\boldsymbol{\theta};\xi_2)$ is nonnegative definite, which is also written as $\mathbf{M}(\boldsymbol{\theta};\xi_1)\geq_{\mathcal{L}}\mathbf{M}(\boldsymbol{\theta};\xi_2)$, or simply $\xi_1\geq_{\mathcal{L}}\xi_2$, with $\geq_{\mathcal{L}}$ denoting the Loewner ordering. This implies that ξ_1 is no worse than ξ_2 under commonly used real-valued, Loewner-isotonic optimality criteria ϕ; i.e., $\mathbf{M}(\boldsymbol{\theta};\xi_1)\geq_{\mathcal{L}}\mathbf{M}(\boldsymbol{\theta};\xi_2)\Rightarrow\phi\{\mathbf{M}(\boldsymbol{\theta};\xi_1)\}\geq\phi\{\mathbf{M}(\boldsymbol{\theta};\xi_2)\}$ [2]. Such optimality criteria include the popularly considered family of ϕ_p-criteria with $p\in[-\infty,1]$ [22,26]:

$$\phi_p(\mathbf{M}_\zeta)=\left\{\frac{1}{m}\mathrm{trace}(\mathbf{M}_\zeta^p)\right\}^{1/p},\tag{5}$$

where ζ denotes the design being evaluated and $\mathbf{M}_\zeta$ is the corresponding m-by-m information matrix of interest. The ϕ_{-1}-criterion is also known as the A-optimality criterion that aims at minimizing the average (asymptotic) variance of parameter estimates. With $p=0$ (or more precisely, $p\to 0$; see [22]), we have the D-criterion, $\phi_0=\det(\mathbf{M}_\zeta)^{1/m}$; the D-criterion is directly linked to the volume of the (asymptotic) confidence ellipsoid of the parameters of interest. The other limiting case is $p=-\infty$; $\phi_{-\infty}$ corresponds to the minimum eigenvalue of $\mathbf{M}_\zeta$, and is also known as the E-criterion. The ϕ_1-criterion is sometimes called a T- or a trace criterion. As an exception of the ϕ_p family, a feasible ϕ_1-optimal design that allows estimable parameters might not exist; see a discussion in Sec. 6.5 of [27]. In this work, we assume situations where feasible optimal designs exist; the existence of such designs is established in Theorem 7.13 of [27] for ϕ_p with $p\in[-\infty,1)$. In addition, we focus only on designs allowing nonsingular information matrices, and refer to such designs as nonsingular designs.

Kim and Kao [23] provided some A- and D-optimal designs for a mixed responses model having no group effects. Their model and optimal design results can be viewed as a special case of ours (e.g., by setting $L=1$). With the more general model formulation in (1), we discuss two situations where (i) there is no common parameter across subject groups; and (ii) there are some common parameters across groups. For both situations, there might or might not be common parameters ($\boldsymbol{\beta}_*$ and/or $\boldsymbol{\gamma}_{*,\ell}$) between $\mu_{0,\ell}(x)$ and $\mu_{1,\ell}(x)$.

3. Optimal design and complete class results

3.1. Models with no common parameter across groups

With model (1), a rather flexible setting is by allowing each group to have its own mean parameters without assuming a common term across groups. This corresponds to the case where $\mathbf{f}_* = \mathbf{f}_s = \{\emptyset\}$ for $s \in \{0, 1, 2\}$. The parameters of interest are reduced to $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_L^\top)^\top$ whose information matrix under the design $\xi = \eta \times \{\tau_\ell\}$ becomes:

$$\mathbf{M}(\boldsymbol{\gamma}; \xi) = \oplus_{\ell=1}^{L} \eta(\ell) \mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell), \;\; \mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell) = \int_{A_\ell}^{B_\ell} \sum_{s=0}^{2} \Gamma_s\{c_\ell(x)\} \bar{\mathbf{h}}_{s,\ell}(x) \bar{\mathbf{h}}_{s,\ell}^\top(x) \,\mathrm{d}\tau_\ell(x), \tag{6}$$

$$\bar{\mathbf{h}}_{0,\ell}(x) = (\mathbf{h}_{*,\ell}^\top(x), \mathbf{h}_{0,\ell}^\top(x), \mathbf{0}_{m_{1,\ell}+m_{2,\ell}}^\top)^\top, \;\; \bar{\mathbf{h}}_{1,\ell}(x) = (\mathbf{h}_{*,\ell}^\top(x), \mathbf{0}_{m_{0,\ell}}^\top, \mathbf{h}_{1,\ell}^\top(x), \mathbf{0}_{m_{2,\ell}}^\top)^\top, \bar{\mathbf{h}}_{2,\ell}(x) = (\mathbf{0}_{m_{*,\ell}+m_{0,\ell}+m_{1,\ell}}^\top, \mathbf{h}_{2,\ell}^\top(x))^\top.$$

Here, $\oplus$ is the matrix direct sum, and we refer to $\mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)$ as the information matrix for $\boldsymbol{\gamma}_\ell$ of group ℓ under the conditional design τ_ℓ; the remaining terms in (6) are as in (2). It then follows that $\mathbf{M}(\boldsymbol{\gamma}; \xi)$ is nonsingular if and only if $\eta(\ell)\mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)$ is nonsingular for all ℓ. In addition, the ϕ_p-optimal design for $\boldsymbol{\gamma}$ can be obtained from the ϕ_p-optimal conditional design for each $\boldsymbol{\gamma}_\ell$ as stated below; see also [8,30].

Theorem 1. *For given marginal design η with $\eta(\ell) > 0 \; \forall \ell$, a design $\xi_\eta = \eta \times \{\tau_\ell^*\}$ is ϕ_p-optimal if the conditional design τ_ℓ^* maximizes $\phi_p\{\mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)\}$ for all $\ell \in \{1, \ldots, L\}$.*

Theorem 1 follows from the block-diagonal structure of the information matrix, and its proof is provided in the Appendix. To identify optimal τ_ℓ, it often is useful to find an upper bound for the number of support points n_ℓ; see [16,25]. The well-known Carathéodory theorem implies that any given τ_ℓ can be represented by a discrete measure having $n_\ell \le m_{\cdot,\ell}(m_{\cdot,\ell}+1)/2+1$ support points [25]. With the special structure of $\mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)$, the same theory gives the following sharper bound for n_ℓ by eliminating from the information matrix the replicated elements, and those that do not vary with designs. The proof of this result is omitted.

Theorem 2. *For any information matrix $\mathbf{M}$ of $\boldsymbol{\gamma}_\ell$, there exists a design τ_ℓ having n_ℓ support points such that $\mathbf{M} = \mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)$, where*

$$n_\ell \le m_{*,\ell}\left(\frac{m_{*,\ell}+1}{2} + m_{0,\ell} + m_{1,\ell}\right) + \sum_{s=0}^{2} \frac{m_{s,\ell}(m_{s,\ell}+1)}{2} + 1.$$

An even sharper bound for n_ℓ may sometimes be achieved. To give such a sharper bound, we adapt here the complete class approach of Yang and Stufken [34] to identify a class of designs, called a complete class, $\mathcal{T}_\ell$, so that for any given τ_ℓ, we can find at least one corresponding $\tilde{\tau}_\ell \in \mathcal{T}_\ell$ with $\tilde{\tau}_\ell \ge_{\mathcal{L}} \tau_\ell$ (see also [21]). An optimal design can thus be found within $\mathcal{T}_\ell$. To introduce this approach, we first define a 'core' of the information matrix, and the Chebyshev system [20]. Here, we use a slightly more general notation to let $\mathbf{M}_\zeta$ be the m-by-m information matrix of interest, and the competing designs ζ have $\int_C^D \mathrm{d}\zeta(c) = 1$.

Definition 1. Let $\mathbf{M}_\zeta$ be the information matrix under design ζ. We call $\mathbf{C}(c)$ a core of $\mathbf{M}_\zeta$ if $\exists$ a nonsingular $\mathbf{P}$ that does not depend on designs such that

$$\mathbf{M}_\zeta = \mathbf{P}\left\{\int_C^D \mathbf{C}(c)\,\mathrm{d}\zeta(c)\right\}\mathbf{P}^\top.$$

Definition 2. A vector of continuous real functions $\boldsymbol{\psi}(c) = (\psi_1(c), \ldots, \psi_K(c))^\top$ on $[C, D]$ is a Chebyshev system if $\det\{((\psi_k(c_j)))_{k,j\in\{1,\ldots,K\}}\} > 0$ for all $C \le c_1 < \cdots < c_K \le D$. Here, $((\psi_k(c_j)))_{k,j\in\{1,\ldots,K\}}$ represents the K-by-K matrix whose (k, j)th element is $\psi_k(c_j)$.

The complete class approach of [34] was developed by considering models with univariate response. Following the same idea, we describe in the next theorem an approach that can also be applied to other situations including the present design problem. A proof that is built upon the results of [34] can be found in the Appendix.

Theorem 3. *Suppose $\boldsymbol{\psi}(c) = (\psi_0(c) \equiv 1, \psi_1(c), \ldots, \psi_{K-1}(c))^\top$ is a Chebyshev system on $[C, D]$ such that, for a positive integer $\nu < m$, every element in the first ν rows of a core $\mathbf{C}(c)$ for $\mathbf{M}_\zeta$ is a linear combination of $\psi_k(c)$'s. Suppose in addition that, for all $\mathbf{a} \ne \mathbf{0}$, (A) $(\boldsymbol{\psi}^\top(c), \psi_K^a(c) \equiv \mathbf{a}^\top\mathbf{C}_{22}(c)\mathbf{a})^\top$ is a Chebyshev system, or (B) $(\boldsymbol{\psi}^\top(c), -\psi_K^a(c))^\top$ is a Chebyshev system, where $\mathbf{C}_{22}(c)$ is the $(m-\nu)$-by-$(m-\nu)$, lower-right submatrix of $\mathbf{C}(c)$. We have the following:*

(i) *If K is odd and (A) holds, designs having at most $(K+1)/2$ support points, including D, form a complete class;*
(ii) *If K is odd and (B) holds, then designs having at most $(K+1)/2$ support points, including C, form a complete class;*
(iii) *If K is even and (A) holds, then designs having at most $K/2+1$ support points, including C and D, form a complete class;*
(iv) *If K is even and (B) holds, then designs having at most $K/2$ support points form a complete class.*

It is noteworthy that the condition in Theorem 3 is slightly weaker than Theorem 1 of [34]. In particular, the elements of $\boldsymbol{\psi}$ do not need to be in $\mathbf{C}$; and even when they do, $\boldsymbol{\psi}\backslash\{\psi_0\}$ might not be a maximal set of linearly independent nonconstant functions in the first ν rows of $\mathbf{C}$ (see, e.g., [23,29]). It also is easily seen that the theorem still holds by replacing $\mathbf{C}_{22}(c)$ in $\psi_K^a(c)$ with a matrix that (after some simultaneous row and column permutations) has $\mathbf{C}_{22}(c)$ as a leading principal submatrix. Moreover, since $\mathbf{M}_1 \geq_{\mathcal{L}} \mathbf{M}_2 \Rightarrow \tilde{\mathbf{P}}\mathbf{M}_1\tilde{\mathbf{P}}^\top \geq_{\mathcal{L}} \tilde{\mathbf{P}}\mathbf{M}_2\tilde{\mathbf{P}}^\top$, we have the following corollary, which also helps in identifying some complete classes of our current setting.

Corollary 1. *Let the information matrix of interest be $\tilde{\mathbf{M}}_\zeta = \tilde{\mathbf{P}}\mathbf{M}_\zeta\tilde{\mathbf{P}}^\top$, where $\tilde{\mathbf{P}}$ is some (not necessary square) matrix that does not depend on the design. Suppose that the Chebyshev systems described in Theorem 3 exist for $\mathbf{M}_\zeta$. Then, Theorem 3 (i)–(iv) still hold for $\tilde{\mathbf{M}}_\zeta$.*

A key ingredient of this complete class approach is to find relevant Chebyshev systems. Here, we provide some Chebyshev systems useful for finding $\tau_\ell^* = \arg\max_{\tau_\ell} \phi\{\mathbf{M}(\gamma_\ell; \tau_\ell)\}$.

Theorem 4. *Let $\alpha(c) = e^c/(e^c+1)$ for $c \in [C, D]$. The following vectors of functions form Chebyshev systems:*

(i) $(1, \alpha(c)^2, \alpha(c)^2 c, \alpha(c), \alpha(c)c, c)^\top$;
(ii) $(1, \alpha(c)^2, \alpha(c)^2 c, \alpha(c), \alpha(c)c, c, \mathbf{a}^\top \boldsymbol{\Lambda}(c)\mathbf{a})^\top$ *with* $\boldsymbol{\Lambda}(c) = \mathrm{diag}\{\alpha(c)c^2, c^2-\alpha(c)c^2, \alpha(c)c^2-\alpha(c)^2c^2\}$ *for any* $\mathbf{a} = (a_1, a_2, a_3)^\top \neq \mathbf{0}$;
(iii) $(1, \alpha(c)^2, \alpha(c)^2 c, \alpha(c)^2 c^2, \alpha(c)^2 c^3, \alpha(c), \alpha(c)c, \alpha(c)c^2, \alpha(c)c^3, c, c^2, c^3)^\top$;
(iv) $(1, \alpha(c)^2, \alpha(c)^2 c, \alpha(c)^2 c^2, \alpha(c)^2 c^3, \alpha(c), \alpha(c)c, \alpha(c)c^2, \alpha(c)c^3, c, c^2, c^3, \mathbf{a}^\top\boldsymbol{\Lambda}(c)\mathbf{a})^\top$ *with* $\boldsymbol{\Lambda}(c) = \mathrm{diag}\{\alpha(c)c^4, c^4 - \alpha(c)c^4\}$ *for any* $\mathbf{a} = (a_1, a_2)^\top \neq \mathbf{0}$;
(v) $(1, \alpha(c)^2, \alpha(c)^2 c, \alpha(c)^2 c^2, -\alpha(c), -\alpha(c)c, -\alpha(c)c^2, c, c^2, c^3)^\top$;
(vi) $(1, \alpha(c)^2, \alpha(c)^2 c, \alpha(c)^2 c^2, -\alpha(c), -\alpha(c)c, -\alpha(c)c^2, c, c^2, c^3, c^4)^\top$.

A proof of Theorem 4 can be found in the Appendix. We note that other Chebyshev systems can also be formed by the next result.

Theorem 5. *Suppose $\boldsymbol{\psi}_A(c) = \mathbf{A}\boldsymbol{\psi}(c)$ with $\det(\mathbf{A}) > 0$. Then, $\boldsymbol{\psi}_A(c)$ is a Chebyshev system if and only if $\boldsymbol{\psi}(c)$ is a Chebyshev system.*

Proof. Let ψ_k^A and ψ_k denote the elements of $\boldsymbol{\psi}_A$ and $\boldsymbol{\psi}$, respectively; $k \in \{1, \ldots, K\}$. We have $\psi_k^A(c) = \sum_j a_{k,j}\psi_j(c)$ with $a_{k,j}$'s being the elements of $\mathbf{A}$, and for $C \leq c_1 < \cdots < c_K \leq D$,

$$\det\{((\psi_k^A(c_j)))_{k,j\in\{1,\ldots,K\}}\} = \det(\mathbf{A})det\{((\psi_k(c_j)))_{k,j\in\{1,\ldots,K\}}\}.$$

With Definition 2, we have that if $\boldsymbol{\psi}(c)$ is a Chebyshev system then so is $\boldsymbol{\psi}_A(c)$. The converse follows by observing that $\boldsymbol{\psi}(c) = \mathbf{A}^{-1}\boldsymbol{\psi}_A(c)$. □

We are ready to give complete class results for finding optimal τ_ℓ under some mixed responses models. Here, we follow previous works, e.g., [1], to assume a commonly used logistic regression for the binary response Z. We note that our complete class results not only contain that of [23] as a special case, but also cover other cases, including the widely used quadratic regression for $\mathrm{E}(Y|Z=z)$ (e.g., [1,29]). Recall that $\mathbf{f}_* = \mathbf{f}_s = \{\emptyset\}$ for $s \in \{0, 1, 2\}$ in this subsection.

Theorem 6. *For given ℓ, consider the information matrix for $\boldsymbol{\gamma}_\ell$ in (6), where $P(\cdot)$ is either $\alpha(\cdot)$ defined in Theorem 5 or $1-\alpha(\cdot)$, and $c_\ell(x) = \gamma_{0,2,\ell} + \gamma_{1,2,\ell}x$; $x \in [A_\ell, B_\ell]$. We have the following results:*

(i) *Suppose $\mathbf{h}_{*,\ell} \cup \mathbf{h}_{z,\ell} = \{1, x\}$ for $z \in \{0, 1\}$. Then, designs with at most 4 support points, including A_ℓ and B_ℓ form a complete class;*
(ii) *Suppose $\mathbf{h}_{*,\ell} \cup \mathbf{h}_{z,\ell} = \{1, x, x^2\}$ for $z \in \{0, 1\}$. Then, designs with at most 7 support points, including A_ℓ and B_ℓ form a complete class.*

Proof. Let $c = \gamma_{0,2,\ell} + \gamma_{1,2,\ell}x \in [C_\ell, D_\ell]$. For appropriate matrices $\tilde{\mathbf{P}}$, and $\mathbf{P} = (\oplus_{s=0}^2 \mathbf{P}_s/\sigma_s)\mathbf{Q}^\top$ with $\sigma_2 = 1$ and $\mathbf{Q}$ being some permutation matrix, the information matrix for $\boldsymbol{\gamma}_\ell$ can be written as (with a slight abuse of notation)

$$\mathbf{M}(\gamma_\ell; \tau_\ell) = \tilde{\mathbf{P}}\mathbf{P}\left\{\int_{C_\ell}^{D_\ell} \mathbf{C}(c)\,\mathrm{d}\tau_\ell(c)\right\}\mathbf{P}^\top\tilde{\mathbf{P}}^\top,$$

where $\mathbf{C}(c) = \mathbf{Q}[\oplus_{s=0}^{2}\tilde{\Gamma}_s(c)\tilde{\mathbf{h}}_{s,\ell}(c)\tilde{\mathbf{h}}_{s,\ell}^{\top}(c)]\mathbf{Q}^{\top}$, $\tilde{\Gamma}_0(c) = 1 - P(c)$, $\tilde{\Gamma}_1(c) = P(c)$, and $\tilde{\Gamma}_2(c) = P(c)\{1 - P(c)\}$, $\tilde{\mathbf{h}}_{0,\ell}(c)$ and $\tilde{\mathbf{h}}_{1,\ell}(c)$ are $(1, c)^{\top}$ for (i), and are $(1, c, c^2)^{\top}$ for (ii), and $\tilde{\mathbf{h}}_{2,\ell}(c) = (1, c)^{\top}$. For (ii), we can set

$$\mathbf{P}_0 = \mathbf{P}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{\gamma_{0,2,\ell}}{\gamma_{1,2,\ell}} & \frac{1}{\gamma_{1,2,\ell}} & 0 \\ \frac{\gamma_{0,2,\ell}^2}{\gamma_{1,2,\ell}^2} & -\frac{2\gamma_{0,2,\ell}}{\gamma_{1,2,\ell}^2} & \frac{1}{\gamma_{1,2,\ell}^2} \end{pmatrix}, \ \mathbf{P}_2 = \begin{pmatrix} 1 & 0 \\ -\frac{\gamma_{0,2,\ell}}{\gamma_{1,2,\ell}} & \frac{1}{\gamma_{1,2,\ell}} \end{pmatrix}. \tag{7}$$

As for (i), the three $\mathbf{P}_s$'s can be set to $\mathbf{P}_2$ in (7). We then have $\tilde{\mathbf{P}} = 1$ when $\mathbf{h}_{*,\ell} = \{\emptyset\}$; the corresponding $\tilde{\mathbf{P}}$ for other cases can also be easily determined. For example, when $\mathbf{h}_{*,\ell} = \{x^2\}$, and with $\mathbf{I}_a$ being the identity matrix of size $a \times a$,

$$\tilde{\mathbf{P}} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & \mathbf{0}_2^{\top} \\ 1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0}_2^{\top} \\ 0 & 1 & 0 & 0 & 0 & 0 & \mathbf{0}_2^{\top} \\ 0 & 0 & 0 & 1 & 0 & 0 & \mathbf{0}_2^{\top} \\ 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{0}_2^{\top} \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{I}_2 \end{pmatrix}.$$

We then have (i) from Theorems 3, 4(i) and 4(ii) with $\nu = 3$, and (ii) from Theorems 3, 4(iii) and 4(iv) with $\nu = 6$. In particular, the permutation matrix $\mathbf{Q}$ is selected so that the $\mathbf{C}_{22}$ in Theorem 3 is the corresponding $\boldsymbol{\Lambda}$ matrix in Theorem 4. □

Theorem 6 includes complete class results for a univariate approach where separate models are fitted to $Y(\ell, x)$ and $Z(\ell, x)$ by ignoring their possible correlation; note that under our mixed responses model, $\mathrm{corr}\{Y(\ell, x), Z(\ell, x)\}$ is:

$$\frac{d(x)}{\sqrt{\sigma_0^2/P\{c_\ell(x)\} + \sigma_1^2/[1 - P\{c_\ell(x)\}] + d(x)^2}}, \ d(x) = \mu_{1,\ell}(x) - \mu_{0,\ell}(x).$$

In particular, the univariate approach corresponds to the case with $\mathbf{h}_{0,\ell} = \mathbf{h}_{1,\ell} = \{\emptyset\}$, and $\sigma_0^2 = \sigma_1^2$. We refer readers to [32] for situations where the univariate approach gives the same inference results as multivariate approaches that allow correlated Y and Z. Complete class results for some other mixed responses models having $\mu_{0,\ell}(x) \neq \mu_{1,\ell}(x)$ are also provided in Theorem 6. In the next theorem, we further find a smaller complete class than Theorem 6(ii) by focusing on (i) the univariate approach, and (ii) a model having the same form as in Example 1 with $\sigma_0^2 = \sigma_1^2$. It is noteworthy that the information matrix for these two cases does not involve $P\{c_\ell(x)\}x^3$.

Theorem 7. *As in Theorem 6(ii), but suppose also that $\sigma_0^2 = \sigma_1^2 = \sigma^2$, and*

(i) $\mathbf{h}_{*,\ell} = \{1, x, x^2\}$ *(i.e., $\mathbf{h}_{0,\ell} = \mathbf{h}_{1,\ell} = \{\emptyset\}$);*
(ii) $\mathbf{h}_{*,\ell} = \{x, x^2\}$ *(i.e., $\mathbf{h}_{0,\ell} = \mathbf{h}_{1,\ell} = \{1\}$).*

Then, designs with at most 6 support points, including A_ℓ and B_ℓ form a complete class.

Proof. We provide a proof with (ii) here, and note that the proof with (i) is similar. With some algebra, we can rewrite the information matrix as:

$$\mathbf{M}(\gamma_\ell; \tau_\ell) = \mathbf{P}\left\{\int_{C_\ell}^{D_\ell} \mathbf{C}(c)\, \mathrm{d}\tau_\ell(c)\right\}\mathbf{P}^{\top}, \ \mathbf{P} = \begin{pmatrix} \mathbf{0}_4\mathbf{0}_2^{\top} & \mathbf{P}_* \\ \mathbf{P}_2 & \mathbf{0}_2\mathbf{0}_4^{\top} \end{pmatrix},$$

$$\mathbf{P}_* = \frac{1}{\sigma}\begin{pmatrix} -\frac{\gamma_{0,2,\ell}}{\gamma_{1,2,\ell}} & -\frac{\gamma_{0,2,\ell}}{\gamma_{1,2,\ell}} & \frac{1}{\gamma_{1,2,\ell}} & 0 \\ \frac{\gamma_{0,2,\ell}^2}{\gamma_{1,2,\ell}^2} & \frac{\gamma_{0,2,\ell}^2}{\gamma_{1,2,\ell}^2} & -\frac{2\gamma_{0,2,\ell}}{\gamma_{1,2,\ell}^2} & \frac{1}{\gamma_{1,2,\ell}^2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

$\mathbf{P}_2$ is as in (7), $\mathbf{C}(c) = [P(c)\{1 - P(c)\}(1, c)^{\top}(1, c)] \oplus \mathbf{C}_0(c)$, and

$$\mathbf{C}_0(c) = \begin{pmatrix} 1 - P(c) & 0 & \{1 - P(c)\}c & \{1 - P(c)\}c^2 \\ 0 & P(c) & P(c)c & P(c)c^2 \\ \{1 - P(c)\}c & P(c)c & c^2 & c^3 \\ \{1 - P(c)\}c^2 & P(c)c^2 & c^3 & c^4 \end{pmatrix}.$$

Our claim then follows from Theorems 3, 4(v), and 4(vi) with $\nu = 5$. □

For given optimality criterion ϕ, we can then search for a ϕ-optimal design within the derived complete classes. In their Theorem 2.5, Hu et al. [17] provided some sufficient conditions for the uniqueness of the optimal design. They require the size of the support of designs to be at least the size of the information matrix of interest. This result thus cannot be directly applied to our study. Note that the optimal designs in our setting can have a support size smaller than the number of parameters of interest. Nevertheless, with a slight modification of their proof, the result essentially gives the next theorem with wider applications, including some of our cases. We defer our modified proof to the Appendix.

Theorem 8. *Suppose the condition in Theorem 3 holds, and the set $\{1, \psi_{1,1}, \ldots, \psi_{v,m}\}$ has exactly K linearly independent elements, where $\psi_{i,j}$ is the (i, j)th element of the core $\mathbf{C}$. Then for $p \in (-\infty, 1)$, the nonsingular ϕ_p-optimal design is unique.*

Clearly, our complete class results for the inference of $\boldsymbol{\gamma}_\ell$, $\ell \in \{1, \ldots, L\}$, can directly be applied to cases with no qualitative factor (or with $L = 1$ subject group). When $L > 1$ and the experimenter has control on the marginal weights w_ℓ's of the subject groups, the optimal marginal design η^* under the ϕ_p-optimality criteria can be found using the next theorem. The result can be proved following the same arguments as for Theorem 7.1 of [30]; see also Theorem 1 of [8]. The proof is omitted. We note that, to maximize the ϕ_1-criterion, one may set $\eta^*(\ell) = 1$ for the ℓth group that has $\max \operatorname{trace}\{\mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)\}$. But this design gives a singular information matrix; see also Section 2 for a discussion on this criterion. It also is noteworthy that when both η and τ_ℓ are controllable, $\xi^* = \eta^* \times \{\tau_\ell^*\}$ is ϕ_p-optimal for $p \in [-\infty, 1)$ if and only if η^* satisfies (8) and τ_ℓ^* is ϕ_p-optimal [30].

Theorem 9. *For given τ_ℓ allowing nonsingular $\mathbf{M}(\boldsymbol{\gamma}_\ell; \tau_\ell)$, $\ell \in \{1, \ldots, L\}$, the design $\xi = \eta^* \times \{\tau_\ell\}$ is ϕ_p-optimal for $p \in [-\infty, 1)$ if and only if the marginal design η^* satisfies the following:*

$$\eta^*(\ell) = \begin{cases} \frac{[\operatorname{trace}\{\mathbf{M}^p(\boldsymbol{\gamma}_\ell;\tau_\ell)\}]^{1/(1-p)}}{\sum_{j=1}^{L}[\operatorname{trace}\{\mathbf{M}^p(\boldsymbol{\gamma}_j;\tau_j)\}]^{1/(1-p)}}, & p \neq -\infty, 0, \\ \frac{m_{.,\ell}}{\sum_{j=1}^{L} m_{.,j}}, & p = 0, \\ \frac{\lambda_1^{-1}\{\mathbf{M}(\boldsymbol{\gamma}_\ell;\tau_\ell)\}}{\sum_{j=1}^{L}\lambda_1^{-1}\{\mathbf{M}(\boldsymbol{\gamma}_j;\tau_j)\}}, & p = -\infty, \end{cases} \tag{8}$$

where $\lambda_1(\mathbf{M})$ is the smallest eigenvalue of $\mathbf{M}$.

3.2. Models with common parameters across groups

We now include situations where some common parameters can be assumed across groups; i.e., $\cup_{s=0}^{2}\mathbf{f}_s(x) \cup \mathbf{f}_* \neq \{\emptyset\}$ in (1); but, we exclude cases with $\cup_{s,\ell}\mathbf{h}_{s,\ell}(x) \cup \mathbf{h}_* = \{\emptyset\}$ since they reduce to models with $L = 1$, and the results in the previous subsection can be applied. In line with [13,31], we extend Theorem 6 to obtain complete class for the joint design ξ.

Theorem 10. *Consider model (1) with $P(\cdot)$ being either $\alpha(\cdot)$ or $1 - \alpha(\cdot)$ defined in Theorem 5, and $\mathbf{f}_2 \cup \mathbf{h}_{2,\ell} = \{1, x\}$ for $\ell \in \{1, \ldots, L\}$. With the information matrix of $\boldsymbol{\theta}$ in (4), we have the following results:*

(i) *Suppose $\mathbf{f}_* \cup \mathbf{f}_z \cup \mathbf{h}_{*,\ell} \cup \mathbf{h}_{z,\ell} = \{1, x\}$ for $z \in \{0, 1\}$. Then, designs with at most $4L$ support points, including $A_1, \ldots, A_L$, and $B_1, \ldots, B_L$ form a complete class;*
(ii) *Suppose $\mathbf{f}_* \cup \mathbf{f}_z \cup \mathbf{h}_{*,\ell} \cup \mathbf{h}_{z,\ell} = \{1, x, x^2\}$ for $z \in \{0, 1\}$. Then, designs with at most $7L$ support points, including $A_1, \ldots, A_L$, and $B_1, \ldots, B_L$ form a complete class.*

Proof. With appropriate matrices $\tilde{\mathbf{P}}_\ell$ and $\mathbf{P}_\ell$ that do not depend on the joint design ξ, the information matrix of $\boldsymbol{\theta}$ can be written as

$$\mathbf{M}(\boldsymbol{\theta}; \xi) = \sum_{\ell=1}^{L} \eta(\ell)\mathbf{M}(\tau_\ell); \;\; \mathbf{M}(\tau_\ell) = \tilde{\mathbf{P}}_\ell \mathbf{P}_\ell \left\{ \int_{C_\ell}^{D_\ell} \mathbf{C}(c)\, \mathrm{d}\tau_\ell(c) \right\} \mathbf{P}_\ell^\top \tilde{\mathbf{P}}_\ell^\top,$$

where $[C_\ell, D_\ell]$ is the range of $c_\ell(x)$, and $\mathbf{C}(c)$ is as in the proof of Theorem 6. With the same $\mathbf{C}(c)$, let $\mathcal{T}_\ell$ be the complete class identified in Theorem 6 (for group ℓ). Corollary 1 then implies that for any given design $\xi = \eta \times \{\tau_\ell\}$, we can find $\tilde{\tau}_\ell \in \mathcal{T}_\ell$ such that $\mathbf{M}(\tilde{\tau}_\ell) \geq_{\mathcal{L}} \mathbf{M}(\tau_\ell)$ for $\ell \in \{1, \ldots, L\}$, and thus with $\tilde{\xi} = \eta \times \{\tilde{\tau}_\ell\}$,

$$\mathbf{M}(\boldsymbol{\theta}; \tilde{\xi}) = \sum_{\ell=1}^{L} \eta(\ell)\mathbf{M}(\tilde{\tau}_\ell) \geq_{\mathcal{L}} \sum_{\ell=1}^{L} \eta(\ell)\mathbf{M}(\tau_\ell) = \mathbf{M}(\boldsymbol{\theta}; \xi).$$

Our claim then follows. □

We note that Theorem 10 still holds even when $\cup_{s=0}^{2}\mathbf{f}_s(x) \cup \mathbf{f}_* = \{\emptyset\}$. But for this situation, results in the previous subsection allow us to separately obtain the optimal conditional design τ_ℓ within the identified complete class $\mathcal{T}_\ell$ for each subject group. The obtained τ_ℓ's can then be combined with the given marginal design η or the optimal one in Theorem 9

Table 1
Optimal conditional designs of treatment doses and weights for the two subject groups of different tumor sizes.

Criterion	Group1	Group2
D	$\tau_1 = \begin{Bmatrix} 0 & 29.39 & 39.08 & 50 \\ .17 & .40 & .14 & .29 \end{Bmatrix}$	$\tau_2 = \begin{Bmatrix} 0 & 32.21 & 50 \\ .17 & .44 & .39 \end{Bmatrix}$
A	$\tau_1 = \begin{Bmatrix} 0 & 23.43 & 48.05 \\ .09 & .59 & .32 \end{Bmatrix}$	$\tau_2 = \begin{Bmatrix} 0 & 25.93 & 50 \\ .09 & .62 & .29 \end{Bmatrix}$

to form the optimal joint design $\xi = \eta \times \{\tau_\ell\}$. However, when there exist common parameters across groups, we will need to simultaneously identify the best support points and their weights for all the L groups.

4. Example

We consider a similar setting as in Example 1. But as in [1], we assume two subject groups, e.g., formed by the tumor size (small/large). We work on a rather flexible model without assuming common parameters across groups. But for comparison purposes, we still follow [1] to set $\mathbf{h}_{*,\ell} = \{x, x^2\}$. Consequently, $\mu_{0,\ell}(x)$ and $\mu_{1,\ell}(x)$ have the same quadratic curve with possibly different intercepts. Our model is:

$$Y(\ell, x) \mid Z(\ell, x) = z \sim \begin{cases} \mathcal{N}(\gamma_{0,0,\ell} + \gamma_{*,1,\ell}x + \gamma_{*,2,\ell}x^2, \sigma^2), & z = 0, \\ \mathcal{N}(\gamma_{1,0,\ell} + \gamma_{*,1,\ell}x + \gamma_{*,2,\ell}x^2, \sigma^2), & z = 1, \end{cases}$$

$$\Pr\{Z(\ell, x) = 1\} = \frac{1}{1 + \exp(\gamma_{2,0,\ell} + \gamma_{2,1,\ell}x)}, \quad \ell \in \{1, 2\}.$$

The range of x is $[0, 50]$ mg, $(\gamma_{2,0,1}, \gamma_{2,1,1}) = (7, -0.2)$ for $\ell = 1$, $(\gamma_{2,0,2}, \gamma_{2,1,2}) = (7, -0.18)$ for $\ell = 2$, and $\sigma^2 = 0.05$. We apply Theorem 7(ii) to obtain both D- and A-optimal conditional designs. With our complete class results, many optimization techniques, such as those in [16,25], can be considered. We choose to use the sequential quadratic programming implemented with the fmincon() function of MATLAB (2016b) by using random initial designs.

Table 1 presents the A- and D-optimal conditional designs for the two groups; [1] also obtained the same D-optimal design for group 1. For both groups, $m_{\cdot,\ell} = 6$ parameters, and with (8), the D-optimal marginal design has $\eta(\ell) = 0.5$ for $\ell \in \{1, 2\}$. As for A-criterion, the optimal marginal design for group 1 is

$$\eta(1) = \sqrt{\text{trace}\{\mathbf{M}^{-1}(\boldsymbol{\gamma}_1; \tau_1)\}} \Big/ \sum_{\ell=1}^{2} \sqrt{\text{trace}\{\mathbf{M}^{-1}(\boldsymbol{\gamma}_\ell; \tau_\ell)\}} = 0.4962.$$

5. Conclusion

In this work, we derive some Chebyshev systems for identifying complete classes to facilitate the search of locally optimal designs under mixed responses models with quantitative and qualitative factors. We discuss cases where there are common parameters across subject groups and/or across submodels, and cases without assuming common parameters. Our complete class results give small upper bounds for the number of support points that needs to be considered when obtaining optimal designs. Some existing optimization techniques can then be directly applied to search over the identified complete class for optimal designs.

CRediT authorship contribution statement

Ming-Hung Kao: Conceptualization, Methodology, Writing - review & editing. **Hazar Khogeer:** Methodology, Software, Writing - review & editing.

Acknowledgments

The research of Ming-Hung Kao was in part supported by National Science Foundation, USA grants DMS-13-52213 and CMMI-17-26445. The authors are grateful to the Editor, Associate Editor and referees for their comments and suggestions, which resulted in an improvement of the presentation.

Appendix

Proof of Theorem 1. We first consider a large class of criteria of the form $\phi_f(\mathbf{M}_\zeta) = \sum_{i=1}^{m} f\{\lambda_i(\mathbf{M}_\zeta)\}$, where ζ is the nonsingular design being evaluated, $\mathbf{M}_\zeta$ is the m-by-m information matrix, $\lambda_i(\mathbf{M}_\zeta)$ is the ith smallest eigenvalue of $\mathbf{M}_\zeta$, and $f(\cdot)$ is finite, concave, and nondecreasing on $(0, \infty)$. We further assume that, for any positive a and λ, the function

$f(\cdot)$ has either (A) $f(a\lambda) = g(a)f(\lambda)$ for some $g(\cdot) > 0$, or (B) $f(a\lambda) = g(a) + f(\lambda)$; and $\mathbf{M}_\zeta = \oplus_{\ell=1}^{L} a_\ell \mathbf{M}_{\ell,\zeta}$ for some $a_\ell > 0$ and m_ℓ-by-m_ℓ matrix $\mathbf{M}_{\ell,\zeta}$ with $\sum_\ell m_\ell = m$. Clearly, the eigenvalues of $\mathbf{M}_\zeta$ are $a_\ell \lambda_i(\mathbf{M}_{\ell,\zeta})$'s. In addition, $\mathbf{M}_\zeta$ is positive definite if and only if $\mathbf{M}_{\ell,\zeta}$'s are positive definite for all $\ell \in \{1, \ldots, L\}$. We also have

$$\phi_f(\mathbf{M}_\zeta) = \sum_{\ell=1}^{L}\sum_{i=1}^{m_\ell} f\{a_\ell \lambda_i(\mathbf{M}_{\ell,\zeta})\} = \begin{cases} \sum_{\ell=1}^{L}\sum_{i=1}^{m_\ell} g(a_\ell) f\{\lambda_i(\mathbf{M}_{\ell,\zeta})\} = \sum_{\ell=1}^{L} g(a_\ell)\phi_f(\mathbf{M}_{\ell,\zeta}), & \text{if (A)}, \\ \sum_{\ell=1}^{L}\sum_{i=1}^{m_\ell} g(a_\ell) + f\{\lambda_i(\mathbf{M}_{\ell,\zeta})\} = \sum_{\ell=1}^{L}\{m_\ell g(a_\ell) + \phi_f(\mathbf{M}_{\ell,\zeta})\}, & \text{if (B)}. \end{cases}$$

By setting a_ℓ to the given $\eta(\ell)$, $\mathbf{M}_{\ell,\zeta} = \mathbf{M}(\gamma_\ell; \tau_\ell)$, and $\mathbf{M}_\zeta = \mathbf{M}(\gamma; \xi)$, we have that $\phi_f\{\mathbf{M}(\gamma; \xi)\}$ is maximized if and only if $\phi_f\{\mathbf{M}(\gamma_\ell; \tau_\ell)\}$ is maximized for all $\ell \in \{1, \ldots, L\}$. It can also be seen that for $p \in (-\infty, 1]$, the ϕ_p-optimal design can be obtained by maximizing a corresponding ϕ_f with (see also [3]):

$$f(a\lambda) = \begin{cases} (a\lambda)^p = a^p\lambda^p, & p \in (0, 1], \\ \ln a\lambda = \ln a + \ln \lambda, & p = 0, \\ -(a\lambda)^p = a^p(-\lambda^p), & p \in (-\infty, 0). \end{cases}$$

Moreover, maximizing $\phi_{-\infty}(\mathbf{M}_\zeta) = \lambda_1(\mathbf{M}_\zeta)$ is to maximize the minimum $a_\ell \lambda_1(\mathbf{M}_{\ell,\zeta})$. For given $a_\ell' s$, a sufficient condition for this is that the design maximizes all $\lambda_1(\mathbf{M}_{\ell,\zeta})$'s. This completes the proof. □

The next Lemma A.1 is essentially Lemmas 1 and 2 of [34], which is useful for proving Theorem 3. In the lemma, the constant function $\psi_0(c)$ is included in $\boldsymbol{\psi}(c)$ for its applications to design measures; see also [9]. In addition, for given functions $\boldsymbol{\psi}$, and ψ_K^a such as those in Theorem 3, we follow [34] to say that a set $S_1 = \{(c_i, w_i) : w_i > 0, c_i \in [C, D], i \in \{1, \ldots, n_1\}\}$ of size n_1 dominates another set $S_2 = \{(\tilde{c}_i, \tilde{w}_i) : \tilde{w}_i > 0, \tilde{c}_i \in [C, D], i \in \{1, \ldots, n_2\}\}$ of size n_2, or simply $S_1 >_\psi S_2$, if

$$\sum_{i=1}^{n_1} w_i \psi_k(c_i) = \sum_{j=1}^{n_2} \tilde{w}_j \psi_k(\tilde{c}_j), \ \ k \in \{0, \ldots, K-1\}, \tag{9}$$
$$\sum_{i=1}^{n_1} w_i \psi_K^a(c_i) > \sum_{j=1}^{n_2} \tilde{w}_j \psi_K^a(\tilde{c}_j), \ \ \forall \mathbf{a} \neq \mathbf{0}.$$

Lemma A.1. *With the same notation as in Theorem 3, suppose for each $\mathbf{a} \neq \mathbf{0}$ (i) $\boldsymbol{\psi}(c)$ and $(\boldsymbol{\psi}^\top(c), \psi_K^a(c))^\top$ are Chebyshev systems on $[C, D]$ or (ii) $\boldsymbol{\psi}(c)$ and $(\boldsymbol{\psi}^\top(c), -\psi_K^a(c))^\top$ are Chebyshev systems on $[C, D]$. In addition, let S_1, S_2, and $>_\psi$ be defined as above. We have the following:*

(i) *If K is odd and (i) holds, then for any set S_2 that either has $n_2 = (K+1)/2$ but does not contain the endpoint D, or has $n_2 > (K+1)/2$, $\exists$ a set S_1 of size $n_1 = (K+1)/2$ and $(D, w_D) \in S_1$ for some $w_D > 0$ such that $S_1 >_\psi S_2$;*
(ii) *If K is odd and (ii) holds, then for any set S_2 that either has $n_2 = (K+1)/2$ but does not contain the endpoint C, or has $n_2 > (K+1)/2$, $\exists$ a set S_1 of size $n_1 = (K+1)/2$ and $(C, w_C) \in S_1$ for some $w_C > 0$ such that $S_1 >_\psi S_2$;*
(iii) *If K is even and (i) holds, then for any set S_2 that has $n_2 = K/2$ but does not contain both endpoints C and D, or has $n_2 = K/2 + 1$ but does not contain at least one of the endpoints, or has $n_2 > K/2 + 1$, $\exists$ a set S_1 of size $n_1 = K/2 + 1$ and $\{(C, w_C), (D, w_D)\} \subset S_1$ for some $w_C, w_D > 0$ such that $S_1 >_\psi S_2$;*
(iv) *If K is even and (ii) holds, then for any set S_2 of size $n_2 \geq K/2 + 1$, $\exists$ a set S_1 of size $n_1 = K/2$ such that $S_1 >_\psi S_2$.*

With the previous lemma, we now provide a proof for Theorem 3.

Proof of Theorem 3. We first partition the core as $\mathbf{C} = ((\mathbf{C}_{ij}))_{i,j \in \{1,2\}}$, where $\mathbf{C}_{22}$ gives ψ_K^a, and $\mathbf{C}_{21} = \mathbf{C}_{12}^\top$. Suppose ζ_1 and ζ_2 are two design measures having a finite support, and $\zeta_1 >_\psi \zeta_2$. Then (9) implies that:

$$\int_C^D \mathbf{C}_{ij}(c)\, \mathrm{d}\zeta_1(c) = \int_C^D \mathbf{C}_{ij}(c)\, \mathrm{d}\zeta_2(c), \ \ \forall (i, j) \neq (2, 2),$$
$$\int_C^D \mathbf{a}^\top \mathbf{C}_{22}(c) \mathbf{a}\, \mathrm{d}\zeta_1(c) > \int_C^D \mathbf{a}^\top \mathbf{C}_{22}(c) \mathbf{a}\, \mathrm{d}\zeta_2(c), \ \ \forall \mathbf{a} \neq \mathbf{0}.$$

The equality holds because each element in $\mathbf{C}_{ij}(c)$ for $(i, j) \neq (2, 2)$ can be expressed as $\sum_{k=0}^{K-1} b_k \psi_k(c)$ for some constant b_k. Since $\mathbf{M}_\zeta = \mathbf{P} \int \mathbf{C}(c)\, \mathrm{d}\zeta(c) \mathbf{P}^\top$ with a nonsingular $\mathbf{P}$, we then have $\mathbf{M}_{\zeta_1} \geq_{\mathcal{L}} \mathbf{M}_{\zeta_2}$ but $\mathbf{M}_{\zeta_1} \neq \mathbf{M}_{\zeta_2}$; for convenience, we denote this as $\zeta_1 >_{\mathcal{L}} \zeta_2$. Thus, we have $\zeta_1 >_\psi \zeta_2 \Rightarrow \zeta_1 >_{\mathcal{L}} \zeta_2$. To complete the proof, we replace the set S_i in Lemma A.1 by design ζ_i; $i \in \{1, 2\}$. It can then be seen that, for any ζ_2 not in the corresponding complete class, we can find a design ζ_1 inside the complete class such that $\zeta_1 >_\psi \zeta_2 \Rightarrow \zeta_1 >_{\mathcal{L}} \zeta_2$. Our claim then follows by observing that the identified complete class also contains all the remaining designs ζ for which a dominating design is not guaranteed by Lemma A.1, but clearly, $\zeta \geq_{\mathcal{L}} \zeta$. □

Lemma A.2 is due to [20], and is applied to verify that a vector of smooth functions is a Chebyshev system; see also [9].

Lemma A.2. *Let $\psi_0(c), \ldots, \psi_K(c)$ be K times differentiable functions defined on $[C, D]$, $\kappa_0(c) = \psi_0(c)$, and $\kappa_{j+1}(c) = (D_jD_{j-1}\ldots D_0\psi_{j+1})(c)$ for $j \in \{0, \ldots, K-1\}$ with*

$$(D_j\psi)(c) = \frac{d}{dc}\left\{\frac{\psi(c)}{\kappa_j(c)}\right\}.$$

If $\kappa_j(c) > 0$ on $[C, D]$ for all $j \in \{0, \ldots, K\}$, then $(\psi_0(c), \ldots, \psi_K(c))^\top$ is a Chebyshev system.

We present below our proof for Theorem 4, which is built upon Lemma A.2.

Proof of Theorem 4. With the same notation as in Lemma A.2, we have $\kappa_0 = \psi_0 = 1$ for (i)–(vi); and for (ii) and (iv), $\kappa_K = D_{K-1}\ldots D_0\mathbf{a}^\top\boldsymbol{\Lambda}\mathbf{a}$. It can be seen that such a $\kappa_K > 0$ for all $\mathbf{a} \neq \mathbf{0}$ if and only if the corresponding (element-wise) derivatives of $\boldsymbol{\Lambda}$ give a positive definite matrix; see also [31]. We now verify that all the κ_j's in (i)–(vi) are positive. For (i) and (ii), we have $\kappa_1(c) = 2e^{2c}/(e^c+1)^3$, $\kappa_2(c) = \kappa_4(c) = 1 + e^c/2$, $\kappa_3(c) = 2(1+e^{-c})/(e^c+2)^2$, $\kappa_5(c) = 4e^{-c}$, $\kappa_6(c) = \mathbf{a}^\top\mathrm{diag}(2e^{2c} + e^c/2,\ 2 + e^c/2,\ e^c/2)\mathbf{a}$, where $\mathrm{diag}(d_1, \ldots, d_r)$ is the diagonal matrix whose diagonal elements are $d_1, \ldots, d_r$. Clearly, $\kappa_k(c) > 0$ for $k \in \{1, \ldots, 5\}$, and $\kappa_6(c) > 0$ for all $\mathbf{a} \neq \mathbf{0}$. For (iii) and (iv), $\kappa_1(c) = 2e^{2c}/(e^c+1)^3$, $\kappa_2(c) = 1 + e^c/2$, $\kappa_3(c) = 2 + 2e^c/(e^c+2)^2$, $\kappa_4(c) = 3(e^{2c} + 10e^c + 16)/(e^c+4)^2$, $\kappa_5(c) = 4(4e^{-c}+1)/\{3(e^c+8)^2\}$, $\kappa_6(c) = 1 + e^c/8$, $\kappa_7(c) = \kappa_{10}(c) = 2$, $\kappa_8(c) = \kappa_{11}(c) = 3$, $\kappa_9(c) = 8e^{-c}/3$, and $\kappa_{12}(c) = \mathbf{a}^\top\mathrm{diag}\{e^c(16e^{2c}+1)/4, 4 + e^c/4\}\mathbf{a}$. All these $\kappa_k(c)$'s are positive. For (v) and (vi), $\kappa_1(c) = 2e^{2c}/(e^c+1)^3$, $\kappa_2(c) = 1 + e^c/2$, $\kappa_3(c) = 2 + 2e^c/(e^c+2)^2$, $\kappa_4(c) = (4e^{-c}+2)/(e^c+4)^2$, $\kappa_5(c) = 1 + e^c/4$, $\kappa_6(c) = \kappa_8(c) = 2$, $\kappa_7(c) = 4e^{-c}$, $\kappa_9(c) = 21e^c/4 + 3(e^c-1)^2$, and $\kappa_{10}(c) = 4\{16(e^c-1)^4 + 38e^{3c} + 225e^{2c} + 38e^c\}/(4e^{2c} - e^c + 4)^2$ are all positive. □

A detailed proof of Theorem 8 is provided below.

Proof of Theorem 8. We first consider an optimality criterion ϕ that is (I) strictly isotonic, and (II) strictly concave on positive definite matrices; i.e., (I) $\phi(\mathbf{M}_{\zeta_1}) > \phi(\mathbf{M}_{\zeta_2})$ for nonsingular designs ζ_1 and ζ_2 with $\zeta_1 >_{\mathcal{L}} \zeta_2$, and (II) $\phi\{(1-\alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2\} > (1-\alpha)\phi(\mathbf{M}_1) + \alpha\phi(\mathbf{M}_2)$ for any $\alpha \in (0, 1)$, positive definite matrix $\mathbf{M}_1$, and nonnegative definite, nonzero matrix $\mathbf{M}_2$ with $\mathbf{M}_2 \not\propto \mathbf{M}_1$. See also Section 5.2 of [27]. Here, the notation $>_{\mathcal{L}}$ is defined as in the proof of Theorem 3. Let ζ^* be a nonsingular design maximizing $\phi(\mathbf{M}_\zeta)$. With (I) and from the proof of Theorem 3, we see that ζ^* is in the complete class; otherwise, we have another ζ_1 with $\zeta_1 >_\psi \zeta^* \Rightarrow \zeta_1 >_{\mathcal{L}} \zeta^* \Rightarrow \phi(\mathbf{M}_{\zeta_1}) > \phi(\mathbf{M}^*_\zeta)$. Suppose now that there exists another ϕ-optimal design ζ_2. Condition (II) then implies that $\mathbf{M}_{\zeta_2} \propto \mathbf{M}_{\zeta^*}$; otherwise, $(1-\alpha)\mathbf{M}_{\zeta^*} + \alpha\mathbf{M}_{\zeta_2}$ is ϕ-better for any $\alpha \in (0, 1)$. But, $\mathbf{M}_{\zeta_2} = a\mathbf{M}_{\zeta^*}$ for some $a > 0$ further implies the positive definiteness of $\mathbf{M}_{\zeta_2}$. With $\phi(\mathbf{M}_{\zeta_2}) = \phi(\mathbf{M}_{\zeta^*})$ and (I), we then have $a = 1$, and consequently, $\mathbf{C}_{\zeta^*} = \mathbf{C}_{\zeta_2}$, where $\mathbf{C}_\zeta = \int_C^D \mathbf{C}(c)\,\mathrm{d}\zeta(c)$ for the core $\mathbf{C}$. Along with $\int_C^D \mathrm{d}\zeta^*(c) = \int_C^D \mathrm{d}\zeta_2(c) (= 1)$, and the condition of Theorem 3, this implies the existence of a Q-by-K constant matrix $\mathbf{B}$ of rank Q such that $\int_C^D \mathbf{B}\boldsymbol{\psi}(c)\,\mathrm{d}\zeta^*(c) = \int_C^D \mathbf{B}\boldsymbol{\psi}(c)\,\mathrm{d}\zeta_2(c)$; and $\mathbf{B}\boldsymbol{\psi}(c)$ includes $\psi_0 = 1$. Note in addition that $\int_C^D \psi_K^a(c)\,\mathrm{d}\zeta^*(c) = \int_C^D \psi_K^a(c)\,\mathrm{d}\zeta_2(c)$ for any $\mathbf{a} \neq \mathbf{0}$. But the condition in the current theorem further implies that $Q = K$, and $\mathbf{B}$ can be selected (e.g., by row permutations) to have $\det(\mathbf{B}) > 0$; thus, $\mathbf{B}\boldsymbol{\psi}(c)$ is a Chebyshev system by Theorem 5. The same as $\boldsymbol{\psi}(c)$, $\mathbf{B}\boldsymbol{\psi}(c)$ and $e\psi_K^a(c)$ also form a Chebyshev system for an $e = 1$ or -1 (e.g., by applying Theorem 5 again). For given $\mathbf{a} \neq \mathbf{0}$, we let $\boldsymbol{\psi}^B(c) = ((\mathbf{B}\boldsymbol{\psi}(c))^\top, \psi_K^B(c) \equiv e\psi_K^a(c))^\top$, and define the following moment space for finite measures ρ:

$$\mathcal{M}_{K+1}^B = \left\{\mathbf{d}(\rho) = (d_0(\rho), \ldots, d_K(\rho))^\top :\ d_k(\rho) = \int_C^D \psi_k^B(c)\,\mathrm{d}\rho(c), k \in \{0, \ldots, K\}\right\}.$$

Each $\mathbf{d}(\rho) \in \mathcal{M}_{K+1}^B$ has a representation $\mathbf{d}(\rho) = \sum_{j=1}^{K+2} t_j\boldsymbol{\psi}^B(c_j)$; $t_j \geq 0$, and $c_j \in [C, D]$; see Chapter II of [20]. For the design ζ^*, $\mathbf{C}_{\zeta^*}$ corresponds to a $\mathbf{d}(\zeta^*) \in \mathcal{M}_{K+1}^B$. As ζ^* has a finite support size n, we have $\mathbf{d}(\zeta^*) = \sum_{j=1}^n w_j^*\boldsymbol{\psi}^B(c_j^*)$, where c_j^* is a support point of ζ^* with weight $w_j^* > 0$. This gives a representation of $\mathbf{d}(\zeta^*)$. In addition, with ζ^* being in the complete class specified in Theorem 3, the index $I\{\mathbf{d}(\zeta^*)\} \leq K/2$. Here, the index is the minimal number of $\boldsymbol{\psi}^B(c_j)$'s needed to represent $\mathbf{d}(\zeta^*)$, but $\boldsymbol{\psi}^B(C)$ and $\boldsymbol{\psi}^B(D)$ are counted as half points. Following Theorem II.2.1 of [20], $\mathbf{d}(\zeta^*)$ is a boundary point of $\mathcal{M}_{K+1}^B$, and it admits a unique representation having positive coefficients t_j. Consequently, $\zeta^* = \zeta_2$. Our claim then follows from the fact that ϕ_p-criteria satisfy (I) and (II) for $p \in (-\infty, 1)$ by Theorem 6.13 of [27]. □

References

[1] A. Biswas, J. López-Fidalgo, Compound designs for dose-finding in the presence of nondesignable covariates, Pharm. Statist. 12 (2013) 92–101.
[2] C.-S. Cheng, Complete class results for the moment matrices of designs over permutation-invariant sets, Ann. Statist. 23 (1995) 41–54.
[3] C.-S. Cheng, Optimal design: Exact theory, in: S. Ghosh, C.R. Rao (Eds.), Handbook of Statistics, Vol. 13, North-Holland, Amsterdam, 1996, pp. 977–1006.
[4] H. Chernoff, Locally optimal designs for estimating parameters, Ann. Math. Stat. 24 (1953) 586–602.
[5] D.R. Cox, N. Wermuth, Response models for mixed binary and quantitative variables, Biometrika 79 (1992) 441–461.
[6] A.R. De Leon, K.C. Chough, Analysis of Mixed Data: Methods & Applications, Chapman and Hall/CRC, Boca Raton, 2013.
[7] X. Deng, R. Jin, QQ models: Joint modeling for quantitative and qualitative quality responses in manufacturing systems, Technometrics 57 (2015) 320–331.

[8] H. Dette, K. Kettelhake, F. Bretz, Designing dose-finding studies with an active control for exponential families, Biometrika 102 (2015) 937–950.
[9] H. Dette, V.B. Melas, A note on the de la garza phenomenon for locally optimal designs, Ann. Statist. 39 (2011) 1266–1281.
[10] H. Dette, V.B. Melas, A. Pepelyshev, Standardized maximin E-optimal designs for the Michaelis–Menten model, Statist. Sinica 13 (2003) 1147–1163.
[11] V.V. Fedorov, Theory of Optimal Experiments, Academic Press, New York, 1972.
[12] V. Fedorov, W. Yuehui, Z. Rongmei, Optimal dose-finding designs with correlated continuous and discrete responses, Stat. Med. 31 (2012) 217–234.
[13] C. Feller, K. Schorning, H. Dette, G. Bermann, B. Bornkamp, Optimal designs for dose response curves with common parameters, Ann. Statist. 45 (2017) 2102–2132.
[14] G.M. Fitzmaurice, N.M. Laird, Regression-models for a bivariate discrete and continuous outcome with clustering, J. Amer. Statist. Assoc. 90 (1995) 845–852.
[15] I. Ford, B. Torsney, C.F.J. Wu, The use of a canonical form in the construction of locally optimal designs for nonlinear problems, J. R. Stat. Soc. Ser. B Stat. Methodol. 54 (1992) 569–583.
[16] R. García-Ródenas, J. García-García, J. López-Fidalgo, J. Martín-Baos, W. Wong, A comparison of general-purpose optimization algorithms for finding optimal approximate experimental designs, Comput. Statist. Data Anal. 144 (2020) 106844.
[17] L.W. Hu, M. Yang, J. Stufken, Saturated locally optimal designs under differentiable optimality criteria, Ann. Statist. 43 (2015) 30–56.
[18] S.-H. Huang, M.-N.L. Huang, K. Shedden, W.K. Wong, Optimal group testing designs for estimating prevalence with uncertain testing errors, J. R. Stat. Soc. Ser. B Stat. Methodol. 79 (2017) 1547–1563.
[19] L. Imhof, Optimum designs for a multiresponse regression model, J. Multivariate Anal. 72 (2000) 120–131.
[20] S. Karlin, W.J. Studden, Tchebycheff Systems with Applications in Analysis and Statistics, John Wiley & Sons, New York, 1966.
[21] J. Kiefer, Optimum experimental designs, J. R. Stat. Soc. Ser. B Stat. Methodol. 21 (1959) 272–319.
[22] J. Kiefer, General equivalence theory for optimum designs (approximate theory), Ann. Statist. 2 (1974) 849–879.
[23] S. Kim, M.-H. Kao, Locally optimal designs for mixed binary and continuous responses, Statist. Probab. Lett. 148 (2019) 112–117.
[24] O. Krafft, M. Schaefer, D-optimal designs for a multivariate regression model, J. Multivariate Anal. 42 (1992) 130–140.
[25] L. Pronzato, A.A. Zhigljavsky, Algorithmic construction of optimal designs on compact sets for concave and differentiable criteria, J. Statist. Plann. Inference 154 (2014) 141–155.
[26] F. Pukelsheim, On linear regression designs which maximize information, J. Statist. Plann. Inference 4 (1980) 339–364.
[27] F. Pukelsheim, Optimal Design of Experiments, John Wiley & Sons, New York, 1993.
[28] F. Pukelsheim, S. Rieder, Efficient rounding of approximate designs, Biometrika 79 (1992) 763–770.
[29] K. Schorning, H. Dette, K. Kettelhake, W.K. Wong, F. Bretz, Optimal designs for active controlled dose-finding trials with efficacy-toxicity outcomes, Biometrika (2017) 1003–1010.
[30] R. Schwabe, Optimum Designs for Multi-Factor Models, Springer, New York, 1996.
[31] J. Stufken, M. Yang, On locally optimal designs for generalized linear models with group effects, Statist. Sinica 22 (2012) 1765–1786.
[32] A. Teixeira-Pinto, S.L.T. Normand, Correlated bivariate continuous and binary outcomes: Issues and applications, Stat. Med. 28 (2009) 1753–1773.
[33] H.-P. Wu, J. Stufken, Locally ϕ_p-optimal designs for generalized linear models with a single-variable quadratic polynomial predictor, Biometrika 101 (2014) 365–375.
[34] M. Yang, J. Stufken, Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences, Ann. Statist. 40 (2012) 1665–1681.
[35] R.-X. Yue, X. Liu, K. Chatterjee, D-optimal designs for multiresponse linear models with a qualitative factor, J. Multivariate Anal. 124 (2014) 57–69.
[36] Y. Zhai, Z. Fang, Locally optimal designs for binary dose–response models, Canad. J. Statist. 46 (2018) 336–354.