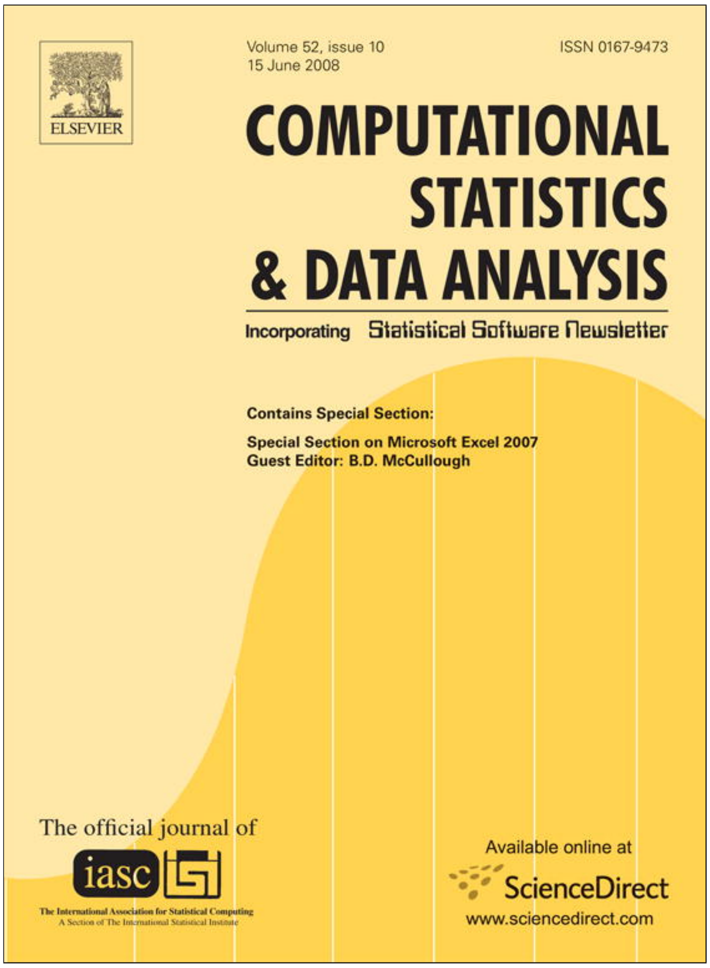
Volume 52, issue 10
15 June 2008
ISSN 0167-9473
ELSEVIER
COMPUTATIONAL STATISTICS & DATA ANALYSIS
Incorporating Statistical Software Newsletter
Contains Special Section:
Special Section on Microsoft Excel 2007
Guest Editor: B.D. McCullough
The official journal of
iasc
The International Association for Statistical Computing
A Section of The International Statistical Institute
Available online at
ScienceDirect
www.sciencedirect.com

Computational Statistics and Data Analysis 52 (2008) 4625–4634

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Optimization in a multivariate generalized linear model situation

S. Mukhopadhyay [a,*], A.I. Khuri [b]

[a] Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

[b] Department of Statistics, University of Florida, P.O. Box 118545, Gainesville, FL 32611-8545, USA

ARTICLE INFO

Article history:
Received 3 September 2007
Received in revised form 1 April 2008
Accepted 1 April 2008
Available online 6 April 2008

ABSTRACT

The purpose of this article is to find the settings of the factors which simultaneously optimize several mean responses in a multivariate generalized linear model (GLM) environment. The generalized distance approach, initially developed for the simultaneous optimization of several linear response surface models, is adapted to this multivariate GLM situation. An application of the proposed methodology is presented in the special case of a bivariate binary distribution resulting from a drug testing experiment concerning two responses, namely, the efficacy and toxicity of a particular drug combination. One of the objectives of this application is to find the dose levels of two drugs that simultaneously maximize their therapeutic effect and minimize any possible toxic effects. A second application is presented in the case of a multivariate gamma distribution.

1. Introduction

The problem of simultaneous optimization of several responses using linear multiresponse models was addressed by Harrington (1965), Derringer and Suich (1980), and Khuri and Conlon (1981) under the usual assumptions of normality and homogeneity of error variances. However, situations may arise, especially in clinical and epidemiological studies, where such assumptions cannot be made on the responses under consideration. For example, the response data may be discrete, possibly correlated, and/or exhibit heterogeneous variances. In such situations, optimization using generalized linear models (GLMs) would be more appropriate. These models have proved to be very useful in several areas of application, for example, in biomedical fields, entomology and climatology. The use of GLMs has also become very important in industrial experiments. For example, in the semiconductor industry, the number of defects on wafers in a manufacturing process, and the resistivity of the test wafer are all data on response variables that do not follow the normal distribution. Resistivity is well known to have a heavy right-tail distribution that is more appropriately modeled as a gamma random variable [see Myers et al. (2002, Section 5.8)]. Lewis et al. (2001) presented several examples of industrial experiments with non-normal responses. Hamada and Nelder (1997) provided convincing arguments for the use of GLMs in the analysis of quality-improvement experiments. They listed several potential applications of GLMs for industry, including their use in reliability-improvement experiments, and in the analysis of data from variation-reduction experiments in which modeling both the mean and dispersion of a process is important.

Unfortunately, the problem of optimization in a generalized linear model (GLM) situation is not as well developed as in the case of linear models. Very limited work has been done in a GLM setup in the case of a single response, and no work has been undertaken in the multivariate case. In single-response GLMs, Paul and Khuri (2000) used *modified ridge analysis* to carry out optimization of a single response. Ridge analysis is a well-known procedure for the optimization of a second-degree response surface model over a spherical region of interest. Khuri and Myers (1979) introduced a modification of this procedure by incorporating a certain constraint on the prediction variance and they called it modified ridge analysis.

* Corresponding author. Tel.: +91 22 2576 7495; fax: +91 22 2572 3480.
E-mail addresses: siulimukherjee@gmail.com (S. Mukhopadhyay), ufakhuri@stat.ufl.edu (A.I. Khuri).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.04.001

In this article, we discuss the determination of the settings of the factors which simultaneously optimize several responses represented by a multivariate GLM. The proposed optimization algorithm is based on the generalized distance approach of Khuri and Conlon (1981). In this algorithm, we first compute the estimated mean responses using the corresponding multiresponse data. The next step is to obtain optima for the individual estimated mean responses. If all the individual optima are attained at the same set of conditions on the control variables or factors, then an "ideal optimum" is said to have been reached. Unfortunately, such an optimum rarely occurs in practice. Optimal conditions for one response may be far from optimal or even physically impractical for the other responses. We therefore resort to finding compromise conditions on the control variables that are "favorable" to all the responses. The deviation from the ideal optimum is measured by a distance function expressed in terms of the estimated mean responses along with their variance–covariance matrix. By minimizing such a distance function we arrive at a set of conditions for a "compromise simultaneous optimum".

One area of application of the generalized distance approach is in Phase I/II dose-finding clinical trials (Thall and Russell, 1998; Bekele and Shen, 2005; Zohar and O'Quigley, 2006) where the primary objective is to find the ideal dose of a drug or a combination of drugs with maximum therapeutic effects and lowest toxic effects. Combined drug therapy is nowadays more frequently used in the treatment of various diseases. It is advantageous to use combined therapy since many therapeutic drugs are usually toxic at high doses. However, when given at low doses in combination with another drug, they can have the same therapeutic effect, but with fewer side effects. Cannabinoids and opioids both produce analgesia and are of particular interest in the context of clinical pain management (Gennings et al., 1994). An example of an opioid is morphine and an example of cannabinoids is $\triangle^9$-tetrahydro-cannabinol ($\triangle^9$-THC), which is the active ingredient in marijuana. THC is used as an appetite stimulant in AIDS-wasting patients and as an anti-emetic for cancer chemotherapy (Nelson et al., 1994; Timpone et al., 1997; Beal et al., 1997). Recent clinical reports support the use of cannabinoids and opioids for peripheral inflammatory pain (Sawynok, 2003). Though both morphine and THC are analgesic in nature, high doses of these drugs, which may be required to treat chronic or severe pain, are accompanied by undesirable side effects. Thus, administration of low doses of THC in combination with morphine seems to be an alternative regimen that enhances the analgesic potency of morphine while reducing undesirable side effects (Cichewicz, 2004). Evidence in recent years studying the synergy between opioids and cannabinoids shows promise for combination pain therapy as well as novel treatments for opioid addiction and abuse (Cichewicz, 2004).

In this article, we search for the so-called "ideal dose" in a drug combination study of the effects of $\triangle^9$-THC and morphine sulfate on male mice. Both drugs are analgesic in nature, but are also associated with some adverse side effects, like hypothermia. We study the effect of the combined doses of the two drugs on the standard binary response of success or failure of the drugs and some measure of their side effects. This results in two responses, namely, efficacy and toxicity of the drugs. The efficacy response is 1 if the drug used has the desired therapeutic effect; the toxicity response is 1 if the drug causes some undesirable side effects. Due to the nature of the data as described, doing statistical analysis using standard linear models will give inadequate results. For such data, GLMs would be more appropriate. Since several responses (for example, efficacy and toxicity) are measured from each subject, multivariate GLMs should be considered. Thus in this article, we propose to find the settings of the factors (dose levels of morphine sulfate and $\triangle^9$-THC) that simultaneously yield optimal, or near optimal, values of the responses (efficacy and toxicity) under consideration in a multivariate GLM setup.

In many industrial situations, several characteristics may be measured on a single product and all such characteristics jointly determine the usefulness and marketability of the product. For example, an industrial engineer may want to study the influence of cutting speed and depth of cut on the life of a tool and the rate at which it loses metal. Often, these quality characteristics are measured by non-normally distributed variables. In such situations, one may be interested in finding the settings of the factors which simultaneously optimize all the measured characteristics of the product. For example, in a particular chemical experiment, a resin is required to have a certain minimum viscosity, high softpoint temperature, and high percentage yield [see Chitra (1990, p. 107)]. The actual realization of the optimum depends on the nature and distribution of the responses, and the form of the fitted models. Hence, the use of GLMs in the simultaneous optimization of several responses is quite imperative. Another potential application of such an endeavor is in the joint modeling of the mean and dispersion of a process, as was mentioned earlier, where it may be of interest to maximize the process mean while minimizing the process variance. Obviously, this can be very beneficial in improving the quality and marketability of a product [see Vining et al. (2000) and Lee and Nelder (2000)].

The remainder of this article is organized as follows. We define the multivariate generalized linear model setup in Section 2. Section 3 is concerned with the simultaneous optimization of a multiresponse function. Section 4 outlines the steps leading to optimization of multiple responses in a multivariate GLM setup. Two special cases, the bivariate binary distribution and the multivariate gamma distribution, are discussed in Section 5. Section 6 presents two numerical examples based on data from these distributions to illustrate the application of the proposed methodology. Concluding remarks are given in Section 7.

2. Multivariate generalized linear models

Consider a multiresponse situation involving q responses. The data set under consideration consists of n independent q-dimensional random variables $\mathbf{y}_1, \ldots, \mathbf{y}_n$. The distribution of $\mathbf{y}_j$ $(j = 1, \ldots, n)$ belongs to the exponential family with the density function

$$\delta(\mathbf{y}_j \mid \boldsymbol{\theta}_j, \phi) = \exp[\phi\{\mathbf{y}_j'\boldsymbol{\theta}_j - b(\boldsymbol{\theta}_j)\} + c(\mathbf{y}_j, \phi)], \quad j = 1, \ldots, n, \tag{1}$$

where $b(\cdot)$, $c(\cdot)$ are known scalar functions and ϕ is a dispersion parameter, possibly unknown. The vector $\boldsymbol{\theta}_j$, $j = 1, \ldots, n$, consists of q elements. The mean vector, $\boldsymbol{\mu}_j$, and variance–covariance matrix, $\boldsymbol{\Sigma}_j$, of $\mathbf{y}_j$ are [see Fahrmeir and Tutz (2001, p. 433)]

$$\boldsymbol{\mu}_j = \frac{\partial b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}, \qquad \boldsymbol{\Sigma}_j = \frac{1}{\phi}\frac{\partial^2 b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'}, \quad \text{respectively.} \tag{2}$$

The linear predictor, $\boldsymbol{\eta}(\mathbf{x}) = \mathbf{Z}'(\mathbf{x})\boldsymbol{\beta}$, is related to the mean response $\boldsymbol{\mu}(\mathbf{x})$ by a link function $\boldsymbol{\eta}(\mathbf{x}) = \mathbf{g}[\boldsymbol{\mu}(\mathbf{x})]$, where $\mathbf{x} = (x_1, \ldots, x_k)'$, $\mathbf{Z}(\mathbf{x}) = \oplus_{i=1}^{q} \mathbf{f}_i(\mathbf{x})$, $\mathbf{f}_i(\mathbf{x})$ is a known vector function of $\mathbf{x}$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_q')'$ is a $\tilde{p}$-dimensional vector of unknown parameters, and $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ip_i})'$ is a vector of unknown parameters for the ith response of order $p_i \times 1$ $(i = 1, \ldots, q)$, where $\sum_{i=1}^{q} p_i = \tilde{p}$, and $\mathbf{g} : R^q \to R^q$. The inverse of $\mathbf{g}$ is assumed to exist and is denoted by $\mathbf{h}$.

Using the maximum likelihood estimate (MLE), $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$, we can obtain an estimate of the mean response $\boldsymbol{\mu}(\mathbf{x})$ given by

$$\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}[\mathbf{Z}'(\mathbf{x})\hat{\boldsymbol{\beta}}]. \tag{3}$$

The variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ is approximately given by [see Fahrmeir and Tutz (2001, p. 106)]

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) \doteq (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}, \tag{4}$$

where $\mathbf{X} = [\mathbf{Z}(\mathbf{x}_1) : \ldots : \mathbf{Z}(\mathbf{x}_n)]'$ and $\mathbf{W}$ is a block-diagonal matrix of the form, $\mathbf{W} = \mathrm{diag}[\mathbf{W}_1, \ldots, \mathbf{W}_n]$, $\mathbf{W}_j = (\frac{\partial \eta_j}{\partial \mu_j'}\boldsymbol{\Sigma}_j\frac{\partial \eta_j'}{\partial \mu_j})^{-1}$, $j = 1, \ldots, n$. Here, $\boldsymbol{\eta}_j = \boldsymbol{\eta}(\mathbf{x}_j)$ $(j = 1, \ldots, n)$ is the linear predictor evaluated at $\mathbf{x}_j = (x_{j1}, \ldots x_{jk})'$, $\boldsymbol{\mu}_j = \boldsymbol{\mu}(\mathbf{x}_j)$, is the mean response at $\mathbf{x}_j$, and $\frac{\partial \eta_j}{\partial \mu_j'}$ is the first-order partial derivative matrix of $\boldsymbol{\eta}(\mathbf{x})$ with respect to $\boldsymbol{\mu}'(\mathbf{x})$ evaluated at $\mathbf{x}_j$ $(j = 1, \ldots, n)$.

The distribution of the MLE, $\hat{\boldsymbol{\beta}}$, is asymptotically normal with mean $\boldsymbol{\beta}$ and a variance–covariance matrix $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$ [see Fahrmeir and Tutz (2001, p. 106)] for large n. Based on Wald's (1943) results, an approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$ is given by

$$C = \{\boldsymbol{\gamma} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})'[(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})](\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}) \le \chi^2_{\alpha,\tilde{p}}\}, \tag{5}$$

where $\tilde{p}$ is the total number of parameters and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) \doteq (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}$; $\hat{\mathbf{W}}$ is an estimate of $\mathbf{W}$ obtained by estimating each element of $\mathbf{W}$ using maximum likelihood estimation.

Using $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ and a first-order approximation of the relation $\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}[\hat{\boldsymbol{\eta}}(\mathbf{x})]$, we get an approximate expression for $\mathrm{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})]$ given by

$$\mathrm{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})] \doteq \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})}[\mathbf{Z}'(\mathbf{x})\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{Z}(\mathbf{x})]\frac{\partial \boldsymbol{\mu}'(\mathbf{x})}{\partial \boldsymbol{\eta}(\mathbf{x})}. \tag{6}$$

An estimator of $\mathrm{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})]$ is therefore of the form

$$\widehat{\mathrm{Var}}[\hat{\boldsymbol{\mu}}(\mathbf{x})] \doteq \widehat{\frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})}}[\mathbf{Z}'(\mathbf{x})\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})\mathbf{Z}(\mathbf{x})]\widehat{\frac{\partial \boldsymbol{\mu}'(\mathbf{x})}{\partial \boldsymbol{\eta}(\mathbf{x})}}, \tag{7}$$

where $\widehat{\frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})}}$ is obtained by estimating each element of $\frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})}$ using maximum likelihood estimation.

3. Simultaneous optimization of a multiresponse function

In this section, we discuss a procedure for the simultaneous optimization of several estimated mean responses in a multivariate GLM setup. For this purpose we use the generalized distance approach of Khuri and Conlon (1981).

3.1. *Generalized distance approach in a multivariate GLM*

Let $\hat{\kappa}_i$ $(i = 1, \ldots, q)$ be the optimum value of $\hat{\mu}_i(\mathbf{x})$ optimized individually over an experimental region, R, where $\hat{\mu}_i(\mathbf{x})$, $i = 1, \ldots, q$, is the i-th element of $\hat{\boldsymbol{\mu}}(\mathbf{x})$. If all the estimated mean responses attain their individual optima at the same setting of $\mathbf{x}$, then the problem of simultaneous optimization is considered solved and no further work is needed. If not, then we resort to finding compromise conditions on the factors that are favorable to all the responses. To find these compromise conditions, we use the generalized distance approach. This approach amounts to finding the conditions on $\mathbf{x}$ that minimize the distance between the vector of the estimated mean responses and the vector of their corresponding individual optimum values. We denote such a distance function by $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \hat{\boldsymbol{\kappa}}]$, where $\hat{\boldsymbol{\kappa}} = (\hat{\kappa}_1, \ldots, \hat{\kappa}_q)'$ is the vector of all the q individual optima. The function $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \hat{\boldsymbol{\kappa}}]$ measures the distance of $\hat{\boldsymbol{\mu}}(\mathbf{x})$ from $\hat{\boldsymbol{\kappa}}$. Several choices of ρ are possible. We choose here a distance function given by

$$\rho_1[\hat{\boldsymbol{\mu}}(\mathbf{x}), \hat{\boldsymbol{\kappa}}] = [(\hat{\boldsymbol{\mu}}(\mathbf{x}) - \hat{\boldsymbol{\kappa}})'\{\widehat{\text{Var}}[\hat{\boldsymbol{\mu}}(\mathbf{x})]\}^{-1}(\hat{\boldsymbol{\mu}}(\mathbf{x}) - \hat{\boldsymbol{\kappa}})]^{1/2}, \tag{8}$$

where $\widehat{\text{Var}}[\hat{\boldsymbol{\mu}}(\mathbf{x})]$ is the estimate of $\text{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})]$ as in formula (7). Another distance function that has some intuitive appeal, particularly to those who like to consider relative changes from the individual optima, is

$$\rho_2[\hat{\boldsymbol{\mu}}(\mathbf{x}), \hat{\boldsymbol{\kappa}}] = \left[\sum_{i=1}^{q} \frac{(\hat{\mu}_i(\mathbf{x}) - \hat{\kappa}_i)^2}{\hat{\kappa}_i^2}\right]^{1/2}. \tag{9}$$

It should be noted that in formulas (8) and (9), $\hat{\boldsymbol{\kappa}}$ has been treated as a fixed non-stochastic point in a q-dimensional Euclidean space. If we recall, $\hat{\boldsymbol{\kappa}}$ is the vector of all the q individual optima of the random variables $\hat{\mu}_1(\mathbf{x}), \ldots, \hat{\mu}_q(\mathbf{x})$. Therefore, $\hat{\kappa}_1, \ldots, \hat{\kappa}_q$ are random variables themselves. Thus, before minimizing ρ, we have to take into account the variability associated with $\hat{\boldsymbol{\kappa}}$. To do so, we apply the following procedure given by Khuri and Conlon (1981): Let $\zeta_i(\boldsymbol{\beta})$ be the true optimum value of the i-th ($i = 1, \ldots, q$) mean response optimized individually over R, and let $\boldsymbol{\zeta}(\boldsymbol{\beta}) = [\zeta_1(\boldsymbol{\beta}), \ldots, \zeta_q(\boldsymbol{\beta})]'$. Our objective is to find an $\mathbf{x} \in R$ such that $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\zeta}(\boldsymbol{\beta})]$ is minimized over R. Since $\boldsymbol{\zeta}(\boldsymbol{\beta})$ is unknown, minimizing $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\zeta}(\boldsymbol{\beta})]$, which is a function of $\boldsymbol{\zeta}(\boldsymbol{\beta})$, is impossible. Instead, we minimize an upper bound on $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\zeta}(\boldsymbol{\beta})]$ as follows: Let D_ζ be a confidence region for $\boldsymbol{\zeta}(\boldsymbol{\beta})$. Then, whenever $\boldsymbol{\zeta}(\boldsymbol{\beta}) \in D_\zeta$,

$$\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\zeta}(\boldsymbol{\beta})] \le \max_{\boldsymbol{\xi} \in D_\zeta} \rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\xi}]. \tag{10}$$

The right-hand side of (10) serves as an overestimate of $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\zeta}(\boldsymbol{\beta})]$. From (10) it follows that

$$\min_{\mathbf{x} \in R} \rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\zeta}(\boldsymbol{\beta})] \le \min_{\mathbf{x} \in R} \max_{\boldsymbol{\xi} \in D_\zeta} \rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\xi}].$$

We therefore minimize the right-hand side of (10) over the region R, thus adopting a conservative distance approach to our minimization problem.

3.2. *Construction of a confidence region on $\boldsymbol{\zeta}(\boldsymbol{\beta})$*

Rao (1973, p. 240), shows that for any continuous function $l(\boldsymbol{\beta})$ defined on a subset of the $\tilde{p}$-dimensional Euclidean space that contains C,

$$\text{P}\{\min_{\boldsymbol{\gamma} \in C} l(\boldsymbol{\gamma}) \le l(\boldsymbol{\beta}) \le \max_{\boldsymbol{\gamma} \in C} l(\boldsymbol{\gamma})\} \ge 1 - \alpha, \tag{11}$$

where, if we recall, C is the $100(1-\alpha)\%$ confidence region on $\boldsymbol{\beta}$ given in (5).

The interval $[\min_{\boldsymbol{\gamma} \in C} l(\boldsymbol{\gamma}), \max_{\boldsymbol{\gamma} \in C} l(\boldsymbol{\gamma})]$ defines a conservative confidence interval on $l(\boldsymbol{\beta})$. Since the individual optimum, $\zeta_i(\boldsymbol{\beta})$, $(i = 1, \ldots, q)$ of the ith true mean response, $\mu_i(\mathbf{x})$, can be expressed as a continuous function of the parameter vector $\boldsymbol{\beta}$, the above result in (11) can be used to obtain a conservative confidence interval on $\zeta_i(\boldsymbol{\beta})$ using the confidence region in (5). Moreover, since whenever $\boldsymbol{\beta} \in C$,

$$\zeta_i(\boldsymbol{\beta}) \in [\min_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma}), \max_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma})], \quad \text{for all } i = 1, \ldots, q,$$

then

$$\text{P}[\min_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma}) \le \zeta_i(\boldsymbol{\beta}) \le \max_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma}) \mid i = 1, \ldots, q] \ge \text{P}[\boldsymbol{\beta} \in C] \approx 1 - \alpha.$$

It follows that $[\min_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma}), \max_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma})]$, $i = 1, \ldots, q$, form conservative simultaneous confidence intervals on the $\zeta_i(\boldsymbol{\beta})$ with a joint coverage probability approximately greater than or equal to $1 - \alpha$. Now, let us denote the interval $[\min_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma}), \max_{\boldsymbol{\gamma} \in C} \zeta_i(\boldsymbol{\gamma})]$ by $D_i(C)$, $i = 1, \ldots, q$. Since $\zeta_i(\boldsymbol{\beta}) \in D_i(C)$ for all $i = 1, \ldots, q$ if and only if $\boldsymbol{\zeta}(\boldsymbol{\beta}) \in \times_{i=1}^{q} D_i(C)$, where $\times_{i=1}^{q} D_i(C)$ denotes the Cartesian product of the $D_i(C)$, then

$$\text{P}[\boldsymbol{\zeta}(\boldsymbol{\beta}) \in \times_{i=1}^{q} D_i(C)] = \text{P}[\zeta_i(\boldsymbol{\beta}) \in D_i(C) \mid i = 1, \ldots, q] \ge \text{P}[\boldsymbol{\beta} \in C] \approx 1 - \alpha. \tag{12}$$

Consequently, $\times_{i=1}^{q} D_i(C)$ forms a rectangular conservative confidence region on $\boldsymbol{\zeta}(\boldsymbol{\beta})$. We therefore choose this region to be the one described in Section 3.1, that is, D_ζ.

4. Optimization algorithm

In this section, we outline the steps leading to the optimization of a multiresponse function in a multivariate GLM situation.

1. The MLE of $\boldsymbol{\beta}$ and the approximate $100(1-\alpha)\%$ confidence region of $\boldsymbol{\beta}$ given by C (formula (5)) are obtained.
2. Using the MLE of $\boldsymbol{\beta}$ from step 1, the estimated mean responses $\hat{\mu}_i(\mathbf{x}),\ i = 1, \ldots, q$ (formula (3)) are obtained.
3. The estimated mean responses are individually optimized over the experimental region R to obtain the vector $\hat{\boldsymbol{\kappa}}$ of estimated individual optima.
4. If all the estimated mean responses attain their individual optima at the same setting of $\mathbf{x}$ in R, then the optimization problem is considered to be solved, otherwise, proceed to step 5.
5. A distance measure $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \hat{\boldsymbol{\kappa}}]$ is chosen according to formulas (8) or (9).
6. A rectangular confidence region $D_\zeta = \times_{i=1}^{q} D_i(C)$ (formula (12)) with confidence coefficient of at least $(1-\alpha)$ is constructed. To obtain the individual confidence intervals, $D_i(C)$, on $\zeta_i(\boldsymbol{\beta})$, which is the true optimum value of $\mu_i(\mathbf{x})$ over R, the following steps are taken:
   (a) For each of N_1 values of $\boldsymbol{\gamma}$ randomly selected from C, we compute the optimum of $\mu_i(\mathbf{x})$ over R (an actual numerical value for N_1 is given in Example 1 in Section 6.1). We denote this optimum by $\zeta_i(\boldsymbol{\gamma})$.
   (b) Next, we find the minimum and maximum of the N_1 values obtained for $\zeta_i(\boldsymbol{\gamma})$ in (a) to get the approximate lower and upper bounds of $D_i(C)$, respectively.
7. For each of N_2 values of $\mathbf{x}$ selected from R, we compute the maximum of the distance function, $\rho[\hat{\boldsymbol{\mu}}(\mathbf{x}), \boldsymbol{\xi}]$, with respect to $\boldsymbol{\xi} \in D_\zeta$. An actual numerical value for N_2 is given in Example 1 in Section 6.1.
8. The minimum of the N_2 values obtained for the maximum distance in step 7 is computed.

Note that increasing the number of selected values of $\boldsymbol{\gamma}$ and $\mathbf{x}$ beyond the chosen values for N_1 and N_2 should not cause any appreciable change in the optimization results.

5. Special cases

5.1. Special case 1: The bivariate binary distribution

In many experimental situations, several responses may be observed for the same subject. A typical example is from a drug testing experiment, where in addition to the standard binary response of success or failure of the drug, some measure of the side effects of the drug may be of interest. This results in two responses, efficacy and toxicity of the drug. The efficacy response is 1 if the drug used has the desired therapeutic effect; the toxicity response is 1 if the drug causes some undesirable side effects. The two responses are assumed to be correlated as they come from the same subject. In this example, the determination of the dose level that simultaneously maximizes the efficacy and minimizes the toxicity effects of the drug is of particular interest.

Consider a bivariate binary response situation in which m_j subjects, or experimental units, are tested at the jth run (jth level of $\mathbf{x}$), $j = 1, \ldots, n$. The measurement taken from the wth subject at the jth run for the efficacy response is termed y_{jw1} ($j = 1, \ldots, n$; $w = 1, \ldots, m_j$) and the corresponding toxicity response is termed as y_{jw2} ($j = 1, \ldots, n$; $w = 1, \ldots, m_j$). The response y_{jw1} is 1 or 0 depending on whether the drug used has the desired therapeutic effect or not; the toxicity response y_{jw2} is 1 or 0 depending on whether the drug causes some undesirable side effects or not. Thus we define $\mathbf{y}_{jw} = (y_{jw1}, y_{jw2})'$ as the vector of bivariate responses from the wth subject at the jth experimental run. If $\mathbf{y}_{jw}$, for $w = 1, \ldots, m_j$, are independent then the p.m.f. (probability mass function) of $\mathbf{y}_j = (\mathbf{y}'_{j1}, \ldots, \mathbf{y}'_{jm_j})'$ is given by

$$\delta_j(\mathbf{y}_j) \propto \pi_{j1}^{\sum_{w=1}^{m_j} y_{jw1}y_{jw2}} \pi_{j2}^{\sum_{w=1}^{m_j} y_{jw1}(1-y_{jw2})} \pi_{j3}^{\sum_{w=1}^{m_j} (1-y_{jw1})y_{jw2}} \pi_{j4}^{\sum_{w=1}^{m_j} (1-y_{jw1})(1-y_{jw2})}, \tag{13}$$

where the four cell probabilities $\boldsymbol{\pi}_j = (\pi_{j1}, \pi_{j2}, \pi_{j3}, \pi_{j4})'$ for the j-th level of $\mathbf{x}$ can be expressed as $\pi_{j1} = P(y_{jw1} = 1, y_{jw2} = 1)$, $\pi_{j2} = P(y_{jw1} = 1, y_{jw2} = 0)$, $\pi_{j3} = P(y_{jw1} = 0, y_{jw2} = 1)$, $\pi_{j4} = P(y_{jw1} = 0, y_{jw2} = 0)$ (these probabilities are the same for all $w = 1, \ldots, m_j$). Note that $\pi_{j1} + \pi_{j2} + \pi_{j3} + \pi_{j4} = 1, j = 1, \ldots, n$. We can then write the above probability mass function as

$$\delta_j(\mathbf{z}_j) \propto \pi_{j1}^{z_{j1}} \pi_{j2}^{z_{j2}} \pi_{j3}^{z_{j3}} \left[1 - \sum_{i=1}^{3} \pi_{ji}\right]^{z_{j4}}, \tag{14}$$

where $z_{j1} = \sum_{w=1}^{m_j} y_{jw1}y_{jw2}, z_{j2} = \sum_{w=1}^{m_j} y_{jw1}(1-y_{jw2}), z_{j3} = \sum_{w=1}^{m_j} (1-y_{jw1})y_{jw2}$ and $z_{j4} = m_j - \sum_{i=1}^{3} z_{ji} = \sum_{w=1}^{m_j} (1-y_{jw1})(1-y_{jw2})$. Thus,

z_{j1} = number of subjects experiencing both efficacious and toxic effects of the drug,
z_{j2} = number of subjects experiencing only efficacious effect of the drug, but no harmful side-effects,
z_{j3} = number of subjects experiencing only toxic effect of the drug, but no efficacious effect of the drug,
z_{j4} = number of subjects experiencing neither efficacious nor toxic effects of the drug.

We note from the form of the p.m.f in (14) [see Casella and Berger (2002, p. 180) and Fahrmeir and Tutz (2001, p. 70, formula 3.12)] that the responses $\mathbf{z}_j = (z_{j1}, z_{j2}, z_{j3})'$ follow a *multinomial distribution* with parameters m_j and $\boldsymbol{\pi}_j = (\pi_{j1}, \pi_{j2}, \pi_{j3})'$, $j = 1, \ldots, n$, that is, $\mathbf{z}_j \sim \text{multinomial}(m_j, \boldsymbol{\pi}_j)$. Thus each z_{ji} $(i = 1, 2, 3)$ follows a binomial(m_j, π_{ji}). Therefore, $z_{j1}+z_{j3}$ follows a binomial$(m_j, \pi_{j1}+\pi_{j3})$, where $z_{j1}+z_{j3}$ = number of subjects experiencing toxic effect of the drug (irrespective of the efficacious effect). Let us denote z_{j2} by E_j (E for efficacy), and let $T_j = z_{j1} + z_{j3}$ (T for toxicity), $N_j = z_{j4}$ and $\pi_{jE} = \pi_{j2}$, $\pi_{jT} = \pi_{j1} + \pi_{j3}$, $\pi_{jN} = \pi_{j4}$. So, $(E_j, T_j)' \sim \text{multinomial}(m_j, \pi_{jE}, \pi_{jT})$.

The mean of $(E_j, T_j)'$ is $\boldsymbol{\mu}_j = (\mu_{jE}, \mu_{jT})'$, where $\mu_{jE} = m_j\pi_{jE}$, $\mu_{jT} = m_j\pi_{jT}$ and the variance–covariance matrix of $(E_j, T_j)'$ is $\boldsymbol{\Sigma}_j$, where

$$\boldsymbol{\Sigma}_j = m_j \begin{pmatrix} \pi_{jE}(1-\pi_{jE}) & -\pi_{jE}\pi_{jT} \\ -\pi_{jT}\pi_{jE} & \pi_{jT}(1-\pi_{jT}) \end{pmatrix}, \quad j = 1, \ldots, n. \tag{15}$$

The corresponding link function used here is [see Agresti (2002, pp. 267–274) and Fahrmeir and Tutz (2001, p. 73)]

$$\eta_i(\mathbf{x}) = \log\left(\frac{\pi_i(\mathbf{x})}{1-\pi_E(\mathbf{x})-\pi_T(\mathbf{x})}\right) = \mathbf{f}_i'(\mathbf{x})\boldsymbol{\beta}_i, \quad i = E, T, \tag{16}$$

where $\mathbf{f}_i(\mathbf{x})$ is a known vector function of $\mathbf{x}$ defined in Section 2, and $\pi_E(\mathbf{x})$ corresponds to the first element of $\boldsymbol{\pi}_{jE,T} = (\pi_{jE}, \pi_{jT})'$, but is evaluated at $\mathbf{x}$ instead of the j-th run, and similarly $\pi_T(\mathbf{x})$ corresponds to the second element of $\boldsymbol{\pi}_{jE,T}$.

Hence, we can write,

$$\hat{\pi}_i(\mathbf{x}) = \frac{\exp[\mathbf{f}_i'(\mathbf{x})\hat{\boldsymbol{\beta}}_i]}{1+\exp[\mathbf{f}_E'(\mathbf{x})\hat{\boldsymbol{\beta}}_E]+\exp[\mathbf{f}_T'(\mathbf{x})\hat{\boldsymbol{\beta}}_T]}, \quad i = E, T,$$

where $\hat{\boldsymbol{\beta}}_i$ is the MLE of $\boldsymbol{\beta}_i$ for $i = E, T$. The first-order partial derivative of $\boldsymbol{\pi}$ with respect to $\boldsymbol{\eta}'$ evaluated at $\mathbf{x}$, as required in formulas (6) and (7), is

$$\frac{\partial\boldsymbol{\pi}(\mathbf{x})}{\partial\boldsymbol{\eta}'(\mathbf{x})} = \begin{pmatrix} \pi_E(\mathbf{x})[1-\pi_E(\mathbf{x})] & -\pi_E(\mathbf{x})\pi_T(\mathbf{x}) \\ -\pi_T(\mathbf{x})\pi_E(\mathbf{x}) & \pi_T(\mathbf{x})[1-\pi_T(\mathbf{x})] \end{pmatrix},$$

where $\boldsymbol{\pi}(\mathbf{x}) = [\pi_E(\mathbf{x}), \pi_T(\mathbf{x})]'$, and $\boldsymbol{\eta}(\mathbf{x}) = [\eta_E(\mathbf{x}), \eta_T(\mathbf{x})]'$.

5.2. *Special case 2: The multivariate gamma distribution*

In this section, we consider a situation involving q mutually independent responses that have a gamma distribution with the same shape parameter and different scale parameters. Hence, the probability density function of $\mathbf{y}_j = (y_{j1}, \ldots, y_{jq})'$, $(j = 1, \ldots, n)$ is given by [see Jearkpaporn et al. (2005, pp. 479–482)]

$$\delta(\mathbf{y}_j) = \prod_{i=1}^{q} \frac{1}{\lambda_{ji}^{\nu}\Gamma(\nu)} y_{ji}^{\nu-1} \exp^{-y_{ji}/\lambda_{ji}}, \quad j = 1\ldots, n. \tag{17}$$

The mean of $\mathbf{y}_j$ is $\boldsymbol{\mu}_j = (\mu_{j1}, \ldots, \mu_{jq})'$, where $\mu_{ji} = \lambda_{ji}\nu$, and its variance–covariance matrix is $\boldsymbol{\Sigma}_j$ [see Jearkpaporn et al. (2005, p. 479)], where

$$\boldsymbol{\Sigma}_j = \begin{pmatrix} \dfrac{{\mu_{j1}}^2}{\nu} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \dfrac{{\mu_{jq}}^2}{\nu} \end{pmatrix}, \quad j = 1, \ldots, n. \tag{18}$$

We assume that the dispersion parameter ϕ is equal to one.

The link function used here is the log link (Jearkpaporn et al., 2005, p. 479;Fahrmeir and Tutz, 2001, p. 23)

$$\eta_i(\mathbf{x}) = \log[\mu_i(\mathbf{x})] = \mathbf{f}_i'(\mathbf{x})\boldsymbol{\beta}_i, \quad i = 1, \ldots, q, \tag{19}$$

where $\mathbf{f}_i'(\mathbf{x})$ is a known vector function of $\mathbf{x}$ defined in Section 2, and $\mu_i(\mathbf{x})$ corresponds to the i-th element of $\boldsymbol{\mu}_j = (\mu_{j1}, \ldots, \mu_{jq})'$, but is evaluated at $\mathbf{x}$ instead of the j-th run. Hence, we can write,

$$\hat{\mu}_i(\mathbf{x}) = \exp[\mathbf{f}_i'(\mathbf{x})\hat{\boldsymbol{\beta}}_i], \quad i = 1, \ldots, q,$$

where $\hat{\boldsymbol{\beta}}_i$ is the MLE of $\boldsymbol{\beta}_i$ for $i = 1, \ldots, q$. The gamma distribution is known to be a good model for many nonnegative continuous random variables with a long right tail (Jearkpaporn et al., 2005, p. 479).

Table 1
Experimental design (3 × 6 factorial) and response values (Example 1)

X_1	X_2	x_1	x_2	E_j	T_j	N_j	m_j
2	0.5	−1	−1	5	0	1	6
2	1	−1	−0.933	2	0	4	6
2	2.5	−1	−0.72	2	0	4	6
2	5	−1	−0.38	4	1	1	6
2	10	−1	0.31	5	1	0	6
2	15	−1	1	3	3	0	6
4	0.5	0	−1	5	0	1	6
4	1	0	−0.93	6	0	0	6
4	2.5	0	−0.72	5	1	0	6
4	5	0	−0.38	3	3	0	6
4	10	0	0.31	3	3	0	6
4	15	0	1	3	3	0	6
6	0.5	1	−1	6	0	0	6
6	1	1	−0.93	6	0	0	6
6	2.5	1	−0.72	6	0	0	6
6	5	1	−0.38	6	0	0	6
6	10	1	0.31	1	5	0	6
6	15	1	1	0	6	0	6

x_1 and x_2 are coded levels of $X_1 =$ morphine sulfate and $X_2 = \triangle^9$-THC, respectively.

6. Numerical examples

In this section, we consider two examples to illustrate the use of the generalized distance to finding a simultaneous optimum. The first example is based on a combination drug therapy study. We have two responses, efficacy and toxicity, that follow a bivariate binary distribution. In the second example, we have used two simulated responses that follow a multivariate gamma distribution. In the latter example, we consider the simultaneous minimization of the two responses.

6.1. Example 1: A drug combination study

The data set considered here is taken from Gennings et al. (1994, pp. 429–451). In a combination drug therapy study on male mice, the pain relieving (analgesic) ability of two drugs, namely, $\triangle^9$-tetrahydro-cannabinol ($\triangle^9$-THC) and morphine sulfate, are studied. Though both drugs are analgesic (i.e., provide pain relief), they are also associated with adverse side effects. The two responses of interest are therefore $y_1 =$ pain relief and $y_2 =$ side effect. The response y_1 takes the value 1 if a mouse takes more than 8 s to flick its tail when placed under a heat lamp, and the value 0, otherwise. The response value $y_1 = 1$ is considered good because the mouse does not feel pain when placed under a heat lamp for at least 8 s due to the pain relieving (analgesic) ability of the drugs. The side-effect response, y_2, was determined by recording the rectal temperature of the mouse after 60 min following drug administration. This response is equal to 1 when the rectal temperature of the mouse drops below 35 °C (resulting in hypothermia) after administration of the drugs, and is equal to zero, otherwise. Hypothermia ($y_2 = 1$) is a harmful side effect. Thus each of the response variables has two levels and all four level combinations are possible. The purpose of the investigation is to study how the associated probabilities concerning pain relief and hypothermia are related to dose levels of the two drugs.

For the pain relief and side-effect responses, 18 groups of mice (six animals per group) were randomly assigned to receive the treatment combinations from a 3 × 6 factorial design, where a treatment combination consists of a single injection using one of three levels of morphine sulfate (2, 4, 6 mg/kg) in addition to one of 6 levels of $\triangle^9$-THC (0.5, 1.0, 2.5, 5.0, 10.0, 15.0 mg/kg). Thus we have 18 runs with 6 experimental units (mice) in each run. The design and resulting data are presented in Table 1. Note that in Table 1, x_1 and x_2 are coded levels of morphine sulfate and $\triangle^9$-THC, respectively.

Recall that at the jth run ($j = 1, \ldots, 18$), E_j is the number of mice experiencing pain relief, but no undesirable side effects, T_j is the number of mice experiencing undesirable side effects. We fit the following additive model for the linear predictors corresponding to the data in Table 1,

$$\eta_E(\mathbf{x}) = \beta_1 + \beta_2 x_1 + \beta_3 x_2$$
$$\eta_T(\mathbf{x}) = \beta_4 + \beta_5 x_1 + \beta_6 x_2. \quad (20)$$

The likelihood ratio (LR) test was used to test for statistical significance of all effects. Both backward elimination and forward selection were used to determine whether the main effects (x_1 and x_2) and their interaction should be retained in the model. The LR test for retaining the interaction effect was not significant. Next, both x_1 and x_2 were tested for elimination. The LR tests for retaining both main effects were significant. The link functions are the same as those described in formula (16). The parameter estimates and their standard errors for the above model are shown in Table 2. These estimates and the estimated variance–covariance matrix of $\hat{\boldsymbol{\beta}}$, given in Table 3, are used jointly to find C, the approximate 95% confidence region on $\boldsymbol{\beta}$ [see formula (5)].

Table 2
Maximum likelihood estimates and standard errors (Example 1)

Parameter	Estimate	Std. error	*P*-value
β_1	5.2598	1.4791	0.0004
β_2	2.9670	1.0568	0.0050
β_3	2.5255	1.3044	0.0528
β_4	4.1942	1.4967	0.0051
β_5	3.6733	1.1234	0.0011
β_6	4.7620	1.3814	0.0006

Scaled deviance $= 68.3053$, DF $= 30$.

Table 3
Estimated variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ (Example 1)

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
$\hat{\beta}_1$	2.1877	1.0781	1.4488	2.1667	1.1003	1.4590
$\hat{\beta}_2$	1.0781	1.1169	0.1649	1.0755	1.1139	0.1676
$\hat{\beta}_3$	1.4488	0.1649	1.7014	1.4320	0.1872	1.7020
$\hat{\beta}_4$	2.1667	1.0755	1.4320	2.2402	1.0686	1.4062
$\hat{\beta}_5$	1.1003	1.1139	0.1872	1.0686	1.2619	0.2502
$\hat{\beta}_6$	1.4590	0.1676	1.7020	1.4062	0.2502	1.9084

Table 4
The individual optima and the confidence region D_ζ (Example 1)

		Location of individual optima		$D_i(C)$[a]	
Estimated mean response	Individual estimated optima ($\hat{\kappa}$)	x_1	x_2	Lower bound	Upper bound
$\hat{\pi}_E$	0.9372 (max)	0.5454	−1.0000	0.6454	0.9886
$\hat{\pi}_T$	0.0080 (min)	−1.0000	−1.0000	0.0005	0.1150

[a] Coverage probability is approximately greater than or equal to 0.95.

Table 5
Simultaneous optima using the distance function in formula (8) (Example 1)

Minimax ρ_1		2.7100
Simultaneous optima	$\hat{\pi}_E$	0.7712
	$\hat{\pi}_T$	0.0696
Location of simultaneous optima	x_1	−1.0000 ($X_1 = 2.0000$)
	x_2	−0.2833 ($X_2 = 5.6962$)

Let $\hat{\kappa}_i$ $(i = E, T)$ be the optimum value of $\hat{\pi}_i(\mathbf{x})$ optimized individually. Thus, $\hat{\kappa}_E$ is the maximum value of $\hat{\pi}_E(\mathbf{x})$ over the experimental region $R = \{x_1, x_2 \mid -1 \le x_1 \le 1, -1 \le x_2 \le 1\}$, and $\hat{\kappa}_T$ is the minimum value of $\hat{\pi}_T(\mathbf{x})$ over R. Table 4 reports these individual optima.

If we recall from Section 3, the components of $\hat{\boldsymbol{\kappa}} = (\hat{\kappa}_E, \hat{\kappa}_T)'$ are random variables, thus we have to take into account the variability in $\hat{\boldsymbol{\kappa}}$ before using the distance metric. Using formula (12), the confidence region D_ζ is the Cartesian product of the individual confidence intervals, $D_i(C)$, on $\zeta_i(\boldsymbol{\beta})$, the true optimum value of $\pi_i(\mathbf{x})$ over R, $i = E, T$. To find $D_i(C)$, the following steps are taken: For each of 1000 values of $\boldsymbol{\gamma}$ randomly selected from C, we compute the optimum of $\pi_i(\mathbf{x})$, $i = E, T$, over the experimental region R. We denote this optimum by $\zeta_i(\boldsymbol{\gamma})$. Next, we find the minimum and maximum of the 1000 values of ζ_i, $i = E, T$, to get the approximate lower and upper bounds of $D_i(C)$, respectively. The intervals, $D_i(C)$, for $i = E, T$ are given in Table 4 which also gives the individual optima $\hat{\kappa}_i$ $(i = E, T)$. The next step is to use the confidence intervals $D_i(C)$ to get the confidence region D_ζ.

For each of 1000 values of $\mathbf{x}$ selected from the region R by a grid search, we compute the maximum of the distance function, $\rho_1[\hat{\boldsymbol{\pi}}(\mathbf{x}), \boldsymbol{\xi}]$, in formula (8) with respect to $\boldsymbol{\xi} \in D_\zeta$. Let us denote this maximum by $\rho_{1\max}(\mathbf{x})$. We then find the minimum of the 1000 values of this maximum distance to approximately arrive at the minimax distance. This distance, corresponding simultaneous optima of $\hat{\pi}_E(\mathbf{x})$, $\hat{\pi}_T(\mathbf{x})$, and their locations are given in Table 5. Note that increasing the number of selected values of $\boldsymbol{\gamma}$ and $\mathbf{x}$ above 1000 does not cause any appreciable change in the results.

From Table 5 we see that the location of $\mathbf{x}$ corresponding to the minimax distance is $X_1 = 2.0000$, $X_2 = 5.6962$ (X_1 and X_2 are the uncoded, or original, levels of morphine sulfate and $\triangle^9$-THC, respectively). This result indicates that pain relief is maximized and the side effect experienced by the mice is minimized when the dose levels of the drugs are 2 mg/kg of morphine sulfate and 5.6962 mg/kg of $\triangle^9$-THC. The probability of efficacy at the simultaneous optimum is above 77% and the probability of toxicity is below 7%.

Table 6
Experimental design and response values (Example 2)

x_1	x_2	Y_1	Y_2
−1	−1	2.0684	0.0425
−1	0	8.1015	0.0143
−1	1	0.3016	0.3348
0	−1	3.8800	1.6180
0	0	5.6472	0.5675
0	1	0.2195	0.3657
1	−1	0.0776	1.7917
1	0	0.1145	1.6073
1	1	0.0114	23.2277

Table 7
Maximum likelihood estimates and standard errors (Example 2)

Parameter	Estimate	Std. error	*P*-value
β_1	0.1196	0.3825	0.7545
β_2	−2.0039	0.7508	0.0076
β_3	−1.3102	0.7627	0.0858
β_4	−0.1069	0.3102	0.7304
β_5	2.0862	0.4107	<.0001
β_6	0.5011	0.3291	0.1279

Scaled deviance $= 20.9807$; DF $= 12$.

Table 8
The individual optima and the confidence region D_ζ (Example 2)

		Location of individual minima		$D_i(C)$[a]	
Estimated mean response	Individual estimated minimum ($\hat{\kappa}$)	x_1	x_2	Lower bound	Upper bound
$\hat{\mu}_1$	0.0410	1	1	0.0013	1.1729
$\hat{\mu}_2$	0.0676	−1	−1	0.0104	0.4145

[a] Coverage probability is approximately greater than or equal to 0.95.

Table 9
Simultaneous optima using the distance function in formula (9) (Example 2)

Minimax ρ_2		265.1834
Simultaneous optima	$\hat{\mu}_1$	0.2396
	$\hat{\mu}_2$	1.9492
Location of simultaneous optima	x_1	0.1379
	x_2	0.9710

6.2. *Example 2: Multivariate gamma*

The implementation of the generalized distance function approach to find the simultaneous optimum in a multivariate gamma setup is illustrated through a simulated example. Two responses following a multivariate gamma distribution, as described in Section 5.2, are generated using model (21) (see below). The link functions used are the same as those described in formula (19). We shall consider the simultaneous minimization of these two responses. The design settings and the simulated data are given in Table 6.

The following models for the linear predictors were fitted using the data in Table 6,

$$\eta_1(\mathbf{x}) = \beta_1 + \beta_2 x_1 + \beta_3 x_2,$$
$$\eta_2(\mathbf{x}) = \beta_4 + \beta_5 x_1 + \beta_6 x_2. \tag{21}$$

The parameter estimates and their standard errors for the above models are shown in Table 7. These estimates and the estimated variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ are used jointly to find C, the approximate 95% confidence region on $\boldsymbol{\beta}$.

The experimental region R is, $R = \{\mathbf{x} \mid -1 \le x_1 \le 1, -1 \le x_2 \le 1\}$. The confidence region D_ζ described in (12) with a 95% coverage probability is given in Table 8, which also includes the individual minima $\hat{\kappa}_1$ and $\hat{\kappa}_2$ for the two responses along with their locations in R.

Using the distance measure ρ_2 given in formula (9), we obtain the location of the simultaneous minima of the two responses to be 0.1379 for x_1 and 0.9710 for x_2, as can be seen from Table 9.

6.3. Computer programs used for the selection process and the optimization

RANDOM_DATA is a set of MATLAB programs used to generate data points randomly from a given region. We used RANDOM_DATA to generate $\boldsymbol{\gamma}$'s randomly from the ellipsoidal region, C. The required MATLAB files can be downloaded from http://people.scs.fsu.edu/~burkardt/m_src/random_data/random_data.html.

A set of MATLAB routines called HEX_GRID was used to select $\mathbf{x}$'s from the experimental region R. HEX_GRID computes points on a hexagonal grid defined on a rectangular region. The required MATLAB files can be downloaded from http://people.scs.fsu.edu/~burkardt/m_src/hex_grid/hex_grid.html.

A computer program called MCS (Huyer and Neumaier, 1999), was used in all the above optimizations. MCS is a MATLAB program for constrained global optimization using function values only. It is based on a multilevel coordinate search that balances global and local searches. The local search for the optimum is done via a sequential quadratic programming procedure. The required MATLAB files can be downloaded from http://www.mat.univie.ac.at/~neum/software/mcs/. To run the programs, the two MATLAB 5 programs MINQ (bound constrained quadratic program solver) and GLS (global line search) are also required. Both MINQ and GLS can be found in the same web site.

SAS 9.1.3 (SAS Institute Inc., 2000–2004) was used to calculate the parameter estimates and the variance–covariance matrix of the estimates.

7. Summary

In many environmental and toxicological studies, it is of interest to investigate the effect of several drugs on two or more responses that can be obtained from the same subject [see, for example, Solana et al. (1990), Gennings et al. (1994), and Moser et al. (2005)]. In such and similar situations, multivariate GLM techniques can be used to model the multiresponse data. This is particularly true given the discrete nature of the data and the fact that the responses are quite often correlated and should therefore not be modelled independently. The fitted multivariate GLM can then be used to determine optimum conditions on the model's control variables that simultaneously optimize the responses. The numerical examples presented in Section 6 demonstrate that the proposed methodology of the generalized distance approach is not limited by the number of responses or the control variables in the multiresponse system. In addition, heterogeneity in the variances, the correlations that may exist among the responses, and the random nature of the estimated individual optima are all taken into consideration in the search for the compromise simultaneous optimum.

References

Agresti, A., 2002. Categorical Data Analysis, 2nd edition. Wiley, New York.
Beal, J., Olson, R., Lefkowitz, L., Laubenstein, L., Bellman, P., Yangco, B., Morales, J., Murphy, R., Powderly, W., Plasse, T., Mosdell, K., Shepard, K., 1997. Long-term efficacy and safety of dronabinol for acquired immunodeficiency syndrome-associated anorexia. J. Pain Symptom Manag. 14, 7–14.
Bekele, N., Shen, Y., 2005. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. Biometrics 61, 344–354.
Casella, G., Berger, R.L., 2002. Statistical Inference, 2nd edition. Duxbury Press, Pacific Grove, CA.
Chitra, S.P., 1990. Multi-response optimization for designed experiments. Amer. Stat. Assoc. Proc. Stat. Comput. Sect. 107–112.
Cichewicz, D., 2004. Synergistic interactions between cannabinoid and opioid analgesics. Life Sci. 74, 1317–1324.
Derringer, G., Suich, R., 1980. Simultaneous optimization of several response variables. J. Qual. Technol. 12, 214–219.
Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd edition. Springer, New York.
Gennings, C., Carter Jr., W.H., Martin, B.R., 1994. Drug interactions between morphine and marijuana. In: Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L., Greenhouse, J. (Eds.), Case Studies in Biometry. Wiley, New York, pp. 429–451.
Hamada, M., Nelder, J.A., 1997. Generalized linear models for quality-improvement experiments. J. Qual. Technol. 29, 292–304.
Harrington, E.C., 1965. The desirability function. Ind. Quality Control 21, 494–498.
Huyer, W., Neumaier, A., 1999. Global optimization by multilevel coordinate search. J. Global Optim. 14, 331–355.
Jearkpaporn, D., Montgomery, D., Runger, G., Borror, C., 2005. Model-based process monitoring using robust generalized linear models. Int. J. Prod. Res. 43, 1337–1354.
Khuri, A.I., Conlon, M., 1981. Simultaneous optimization of multiple responses represented by polynomial regression functions. Technometrics 23, 363–375.
Khuri, A.I., Myers, R.H., 1979. Modified ridge analysis. Technometrics 21, 467–473.
Lee, Y., Nelder, J.A., 2000. Joint modeling of the mean and dispersion for the analysis of quality improvement experiments. In: Park, S.H., Vining, G.G. (Eds.), Statistical Process Monitoring and Optimization. Dekker, New York, pp. 387–394.
Lewis, S.L., Montgomery, D.C., Myers, R.H., 2001. Examples of designed experiments with nonnormal responses. J. Qual. Technol. 33, 265–278.
Moser, V.C., Casey, M., Hamm, A., Carter Jr., W.H., Simmons, J.E., Gennings, C., 2005. Neurotoxicological and statistical analyses of a mixture of five organophosphorus pesticides using a ray design. Toxicol Sci. 86, 101–115.
Myers, R.H., Montgomery, D.C., Vining, G.G., 2002. Generalized Linear Models. Wiley, New York.
Nelson, K., Walsh, T., Deeter, P., 1994. A phase II study of delta-9 tetrahydrocannabinol for appetite stimulation in cancer-associated anorexia. J. Palliative Care 10, 14–19.
Paul, S., Khuri, A.I., 2000. Modified ridge analysis under nonstandard conditions. Commun. Statist.-Theory Meth. 29, 2181–2200.
Rao, C.R., 1973. Linear Statistical Inference and its Application. Wiley, New York.
SAS Institute Inc., 2000–2004. SAS 9.1.3 Help and Documentation, Author, Cary, NC.
Sawynok, J., 2003. Topical and peripherally acting analgesics. Pharmacol. Rev. 55, 1–20.
Solana, R.P., Harris, L.W., Carter, W.H.J., Talbot, B.G., Carchman, R.A., Gennings, C., 1990. Evaluation of a two-drug combination pretreatment against organophosphorus exposure. Toxicol. Appl. Pharmacol. 102, 421–429.
Thall, P., Russell, K., 1998. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. Biometrics 54, 251–264.
Timpone, J., Wright, D., Li, N., Egorin, M., Enama, M., Mayers, J., Galetto, G., 1997. The safety and pharmacokinetics of single-agent and combination therapy with megestrol acetate and dronabinol for the treatment of HIV wasting syndrome. Aids Res. Hum. Retrov. 13, 305–315.
Vining, G.G., Schaub, D.A., Modigh, 2000. Experimental strategies for estimating mean and variance function. In: Park, S.H., Vining, G.G. (Eds.), Statistical Process Monitoring and Optimization. Dekker, New York, pp. 291–304.
Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54, 426–482.
Zohar, S., O'Quigley, J., 2006. Optimal designs for estimating the most successful dose. Statist. Med. 25, 4311–4320.