Noname manuscript No.
(will be inserted by the editor)

Optimal design to discriminate between rival copula models for a bivariate binary response

Laura Deldossi · Silvia Angela Osmetti · Chiara Tommasi

Received: date / Accepted: date

Abstract We consider a bivariate logistic model for a binary response and we assume that two rival dependence structures are possible. Copula functions are very useful tools to model different kinds of dependence with arbitrary marginal distributions. We consider Clayton and Gumbel copulae as competing association models. The focus is on applications in testing a new drug looking at both efficacy and toxicity outcomes. In this context, one of the main goals is to find the dose which maximizes the probability of efficacy without toxicity, herein called P-optimal dose. If the P-optimal dose changes under the two rival copulae, then it is relevant to identify the proper association model. To this aim, we propose a criterion (called PKL-) which enables us to find the optimal doses to discriminate between the rival copulae, subject to a constraint that protects patients against dangerous doses. Furthermore, by applying the likelihood ratio test for non-nested models, via a simulation study we confirm that the PKL-optimal design is really able to discriminate between the rival copulae.

Keywords Bivariate logistic model · Copula models · Cox's test · KL-optimality · Optimal experimental design · Efficacy-Toxicity response

Laura Deldossi · Silvia Angela Osmetti
Dipartimento di Scienze Statistiche
Università Cattolica del Sacro Cuore
Largo Gemelli, 1; 20123 Milan, Italy
Tel.: +39-02-72342764
Fax: +39-02-72343064
E-mail: laura.deldossi@unicatt.it; silvia.osmetti@unicatt.it

Chiara Tommasi
Dipartimento DEMM
Università degli Studi di Milano
Via Conservatorio, 7; 20122 Milan, Italy
Tel.: +39-02-50321537
Fax: +39-02-50321005
E-mail: chiara.tommasi@unimi.it

Mathematics Subject Classification (2000) 62K05 · 62F03

1 Introduction

In recent years, there has been an increasing interest in developing dose finding methods incorporating both efficacy and toxicity outcomes; see Dragalin et al (2008); Gao and Rosenberger (2013); Thall and Cook (2004); Thall (2012); Yuan and Guosheng (2009) among others. Up to our knowledge, in the literature, the association between efficacy and toxicity is always specified (except for some unknown parameters) through a bivariate model. For instance, among others, Dragalin et al (2008) propose a bivariate probit model for the selection of an efficacious and safe dose for a new anticoagulant compound to prevent thromboembolic disorders; Thall and Cook (2004) apply the bivariate binary Gumbel-Morgenstein model to identify rapid treatment of acute ischemic stroke; Yuan and Guosheng (2009) model toxicity and efficacy as time-to-event outcomes through the Clayton copula to investigate a novel mitotic inhibitor for treating prostate cancer; Tao et al (2013) propose a joint model for correlated efficacy-toxicity outcome constructed with Archimedean copula. However, as argued in Gao and Rosenberger (2013), assuming that the true efficacy-toxicity relation arises from a specific bivariate model might lead to unpleasant inferential consequences if the model is misspecified. Hence the motivation of this paper: to design an experiment with the aim of discriminating between rival bivariate models. More specifically, we consider a bivariate logistic model for a binary response and we use Clayton and Gumbel copulae (both allowing for positive association) as competing models for the dependence structure. We need to discriminate between the two rival models when the dose which maximizes the probability of efficacy without toxicity (called P-optimal dose) is different under the two models. The P-optimal dose is the safest and the most efficacious dose and it can be used as a benchmark for other doses; therefore, when this dose changes under the rival models it is necessary to clarify which is the true model and hence the true P-optimal dose.

In order to establish how the P-optimal dose depends on the assumed dependence structure (Clayton and Gumbel copulae) we have developed a robustness study. This study (which is reported in the Supplementary Material) shows that the P-optimal dose may change under different copula models; from here, the necessity to discriminate between competing copulae. To solve the discrimination problem, Perrone et al (2017) apply the D_s-criterion which can be used only for nested models; for this reason, they need to introduce the mixture copula model (which includes the rival copulae as special cases). In this paper, instead, we compare directly the competing models without using any other auxiliary reference model. More specifically, we modify the KL-optimality criterion proposed by López-Fidalgo et al (2007) in order to identify the correct dependence structure and at the same time protect patients against dangerous doses. In more detail, we propose a criterion (called

PKL-) which enables us to find doses which are "good" to discriminate between the rival copulae, subject to a constraint that protects patients against doses which are far away from the P-optimal dose. Finally, in order to assess the ability of the PKL-optimal design to select the right copula, we perform a simulation study where the likelihood ratio test for non-nested models is applied.

As previously recalled the D_s-criterion can be applied to discriminate between nested models. For separate models Atkinson and Fedorov (1975a,b) introduced the well known T-optimality, but it can be used only for regression Gaussian models. Some contributions to the theory of T-optimality, among others, are Ponce de Leon and Atkinson (1991), Uciński and Bogacka (2005), López-Fidalgo et al (2008) and Dette and Titoff (2009). Recently, Drovandi et al (2014) propose a sequential design based on the mutual information for model discrimination.

Let us note that the focus of this work is on applications in dose finding methods; the proposal, however, might be relevant for other application areas. For instance, in manufacturing industry to study the relationship between machine component failures under stress; see Kim and Flournoy (2015).

The paper is organized as follows. In Section 2 the bivariate copula model is introduced and the main definitions are given. Section 3 describes the binary model for efficacy-toxicity response through a copula function. Section 4 provides the definition of P-optimal dose and the motivation of the work. The PKL-optimality criterion is introduced in Section 5, where an equivalence theorem is also proved. Finally, in Section 6 we perform a simulation study to evaluate the performance of the PKL-optimum design to discriminate between the rival copulae. Concluding remarks follow in Section 7. Theoretical details are deferred to Appendices A and B.

2 Bivariate Copula-Based Model

Let (Y_1, Y_2) be a bivariate response variable with marginal distributions $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$, which depend on the unknown parameter vectors α and β, respectively. If Y_1 and Y_2 are not independent, then it is necessary to define a joint model for (Y_1,Y_2). Copula functions provide a rich and flexible class of models to obtain joint distributions for multivariate data.

A bivariate copula is a function $C : I^2 \to I$, with $I^2 = [0,1] \times [0,1]$ and $I = [0,1]$, that, with an appropriate extension of the domain in R^2, satisfies all the properties of a cumulative distribution function (cdf). In particular, it is the cdf of a bivariate random variable (U_1, U_2), with uniform marginal distributions in $[0,1]$:

$$C(u_1, u_2; \theta) = P(U_1 \leq u_1, U_2 \leq u_2; \theta), \quad 0 \leq u_1 \leq 1 \quad 0 \leq u_2 \leq 1,$$

where $\theta \in \Theta$ is a parameter measuring the dependence between U_1 and U_2. The importance of copulae in statistical modelling stems from Sklar's theorem (Nelsen, 2006), which states that a joint distribution can be expressed in terms

of marginal distributions and a function $C(\cdot,\cdot;\theta)$ that binds them together. In more detail, according to Sklar's theorem, if $F_{Y_1,Y_2}(y_1,y_2;\delta,\theta)$ is the joint cdf of (Y_1,Y_2), where $\delta=(\alpha,\beta)$, then there exists a copula function $C\colon I^2\to I$ such that

$$F_{Y_1,Y_2}(y_1,y_2;\delta,\theta)=C\big(F_{Y_1}(y_1;\alpha),F_{Y_2}(y_2;\beta);\theta\big),\quad y_1,y_2\in\mathbb{R}. \tag{1}$$

If $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$ are continuous functions then the copula $C(\cdot,\cdot;\theta)$ is unique. Conversely, if $C(\cdot,\cdot;\theta)$ is a copula function and $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$ are marginal cdfs, then $F_{Y_1,Y_2}(y_1,y_2;\delta,\theta)$ given in (1) is a joint cdf.

From (1) we have that a copula captures the dependence structure between the marginal probabilities. This idea allows researchers to consider marginal distributions and the dependence between them as two separate but related issues. Finally, let us recall that for each copula there exists a relationship between the parameter θ and Kendall's τ coefficient (see Nelsen (2006) pp. 158-170) and between θ and the lower and upper tail dependence parameters λ_l and λ_u (which measure the association in the tails of the joint distribution; see Nelsen (2006) pp. 214-216).

3 Binary Model for Efficacy and Toxicity

Let (Y_1,Y_2) be a binary efficacy-toxicity response variable; both Y_1 and Y_2 take values in $\{0,1\}$ (1 denotes occurrence and 0 denotes no occurrence). $\pi_1(x;\alpha)=P(Y_1=1|x;\alpha)$ and $\pi_2(x;\beta)=P(Y_2=1|x;\beta)$ are the marginal success probabilities of efficacy and toxicity, where $x\in\mathcal{X}$ denotes the dose of a drug. We consider a logistic model for both Y_1 and Y_2.

It is commonly accepted that efficacy and toxicity increase with dose. For efficacy, however, in order to allow a wide variety of possible dose-response relationships (including non-monotonic functions) a logistic model with a quadratic term is sometimes preferred; see Thall and Cook (2004). Then, we assume the following logistic models for efficacy and toxicity:

$$\pi_1(x;\alpha)=P(Y_1=1|x;\alpha)=\frac{e^{\alpha_0+\alpha_1x+\alpha_2x^2}}{1+e^{\alpha_0+\alpha_1x+\alpha_2x^2}},\quad \alpha=(\alpha_0,\alpha_1,\alpha_2),$$
$$\pi_2(x;\beta)=P(Y_2=1|x;\beta)=\frac{e^{\beta_0+\beta_1x}}{1+e^{\beta_0+\beta_1x}},\quad \beta=(\beta_0,\beta_1).$$

A copula approach is applied to define a bivariate binary logistic model for the efficacy-toxicity response. If $\delta=(\alpha,\beta)$ and $C(\cdot,\cdot;\theta)$ is a copula function which models the dependence between $\pi_1(x;\alpha)$ and $\pi_2(x;\beta)$, then the joint probability of (Y_1,Y_2) at an experimental condition x is

$$p^C_{y_1y_2}(x;\delta,\theta)=P(Y_1=y_1,Y_2=y_2|x;\delta,\theta),\qquad y_1,y_2=0,1. \tag{2}$$

From (2) and the copula representation (1), let

$$p^C_{11}(x;\delta,\theta)=P(Y_1=1,Y_2=1|x;\delta,\theta)=C\big(\pi_1(x;\alpha),\pi_2(x;\beta);\theta\big). \tag{3}$$

Equation (3) defines a class of models for the bivariate binary response: specifying $C(\cdot,\cdot;\theta)$ it provides a particular model.

Table 1: Joint probabilities for efficacy and toxicity.

	Toxicity		
Efficacy	1	0	
1	p^C_{11}	p^C_{10}	$\pi_1(x;\alpha)$
0	p^C_{01}	p^C_{00}	$1-\pi_1(x;\alpha)$
	$\pi_2(x;\beta)$	$1-\pi_2(x;\beta)$	1

From Table 1 we have that

$$p^C_{10}(x;\delta,\theta) = \pi_1(x;\alpha) - p^C_{11}(x;\delta,\theta), \tag{4}$$
$$p^C_{01}(x;\delta,\theta) = \pi_2(x;\beta) - p^C_{11}(x;\delta,\theta), \tag{5}$$
$$p^C_{00}(x;\delta,\theta) = 1 - \pi_1(x;\alpha) - \pi_2(x;\beta) + p^C_{11}(x;\delta,\theta). \tag{6}$$

Several bivariate copulae have been proposed in the literature (see for instance Nelsen (2006)). In this paper we consider only Clayton and Gumbel copulae which have been applied in the context of Optimal Design by Denman et al (2011) and Perrone and Müller (2016) and which have been used by Tao et al (2013) for modeling correlated efficacy-toxicity outcomes in a dose-finding clinical study.

Clayton and Gumbel copulae are recalled in Table 2.

Table 2: Copula functions

Copula	$C(u_1,u_2;\theta)$	$\theta\in\Theta$
Clayton	$(u_1^{-\theta}+u_2^{-\theta}-1)^{-1/\theta}$	$\theta\in(0,\infty)$
Gumbel	$\exp\left(-\left[\{-\ln(u_1)\}^\theta+\{-\ln(u_2)\}^\theta\right]^{1/\theta}\right)$	$\theta\in[1,\infty)$

Table 3 lists the herein considered values for θ, along with corresponding association measures: Kendall's τ and the lower and upper tail dependence coefficients, λ_l and λ_u.

Both these copulae allow only for positive association between variables ($\tau\geq 0$) but they exhibit strong *left* and strong *right* tail dependence, respectively. Their main characteristics are:

- As θ approaches zero the Clayton copula approaches the product copula $\Pi = u_1u_2$ (independence situation). For $\theta\to\infty$ the copula approaches the Frećhet-Hoeffding upper bound $M = min(u_1,u_2)$. In the bivariate case, the upper bound represents perfect positive dependence (i.e. comonotonicity or positive monotone functional dependence) between variables (see Nelsen

Table 3: Copula parameter values and the related dependence and tail dependence coefficients

Clayton	θ	$\tau = \theta/(\theta+2)$	$\lambda_l = 2^{(-1/\theta)}$	$\lambda_u = 0$
	2	0.500	0.707	0.000
	8	0.800	0.917	0.000
	18	0.900	0.962	0.000
Gumbel	θ	$\tau = 1 - 1/\theta$	$\lambda_l = 0$	$\lambda_u = 2 - 2^{(1/\theta)}$
	2	0.500	0.000	0.586
	5	0.800	0.000	0.851
	10	0.900	0.000	0.928

(2006) p.32 and p.187 for details). This copula exhibits strong left (lower) tail dependence, i.e. there exists a relationship between efficacy and toxicity when they assume their low values.

- For $\theta = 1$ the Gumbel copula corresponds to Π. For $\theta \to \infty$ the copula approaches the Frećhet-Hoeffding upper bound M. This copula exhibits strong right (upper) tail dependence, i.e. there exists a relationship between efficacy and toxicity when they assume their high values.

4 P-optimal dose and motivation of the paper

Researchers are usually interested in finding the P-optimal dose which maximizes the probability of efficacy without toxicity, i.e.

$$x_C^P = \arg\max_{x \in \mathcal{X}} p_{10}^C(x; \delta, \theta). \tag{7}$$

The computation of the P-optimal dose x_C^P is a deterministic problem that can be solved whenever the model for the data is known.

Equation (7) shows that x_C^P depends on the assumed dependence structure $C(\cdot, \cdot; \theta)$. To establish if the P-optimal dose changes considerably under different dependence structures, we have performed a robustness study exploring several scenarios δ (i.e. different values of the marginal parameters α and β). In order to obtain results which do not depend on the minimum (x_{min}) or maximum (x_{max}) doses, neither on the unit of measurement, we standardize the x according to this formula:

$$d = \frac{x - \frac{x_{min}+x_{max}}{2}}{\frac{x_{max}-x_{min}}{2}}, \tag{8}$$

hence the experimental domain $\mathcal{X}$ becomes the interval $\mathcal{D}= [-1, 1]$.

From the robustness study (see the Supplementary Material and Deldossi et al (2016)) we have that:

a) There are scenarios where the P-optimal dose does not change substantially under Clayton or Gumbel copulae and this common dose is obtained even assuming (incorrectly) independence (as in Scenario 1 of the Supplementary Material);
b) There are scenarios where the copula misspecification does not influence the P-optimal dose (as in the previous case), but we have a different P-optimal dose if we incorrectly assume independence (such as Scenario 2 of the Supplementary Material);
c) There are scenarios where the P-optimal dose changes considerably under different rival copulae (as in Scenario 3 of the Supplementary Material).

Therefore, it is necessary to discriminate between copulae in case c). This occurrence happens when the probability of toxicity overcomes the probability of efficacy at each dose (Fig. 5 in the Supplementary Material), which is quite common for instance in chemotherapeutic treatments.

In short, a pilot study, an expert opinion or past experiences suggest a value for δ and τ (as a consequence, from Table 3 the association parameters in the two rival copulae are also available); if the P-optimal doses obtained from (7) under distinct copula models are quite different, then it is necessary to select the most adequate dependence model. The identification of the true dependence structure, however, may be difficult because the competing models differ only for the tail dependence. In order to discriminate between rival copulae, we propose a constrained version of the KL-optimality criterion such that the corresponding optimum design is good to discriminate between Clayton and Gumbel copulae without exposing patients to unsafe doses.

5 Constrained KL-Optimality

An approximate design ξ with a finite number of support points is denoted as

$$\xi = \begin{Bmatrix} d_1 & \cdots & d_k \\ \omega_1 & \cdots & \omega_k \end{Bmatrix},$$

where $d_i \in \mathcal{D}$ is an experimental condition that the researcher can freely choose in the experimental domain $\mathcal{D}$ and $0 \leq \omega_i = \xi(d_i) \leq 1$, $i = 1, \ldots, k$, are weights summing up to 1 and representing the amount of experimental effort at each support point.

An experimental design is said "optimal" if it maximizes a concave optimality criterion function which reflects an inferential goal.

In what follows the indices Cl and G denote Clayton and Gumbel copulae, respectively. From now on we assume that nominal values for δ and τ are available (hence, θ_{Cl} and θ_G are known). In order to discriminate between the rival copulae, the following geometric mean of KL-efficiencies may be used as an optimality criterion:

$$\Phi_{KL}(\xi; \delta, \theta_{Cl}, \theta_G) = \{\mathrm{Eff}_{G,Cl}(\xi; \delta, \theta_{Cl})\}^{\gamma_1} \cdot \{\mathrm{Eff}_{Cl,G}(\xi; \delta, \theta_G)\}^{1-\gamma_1}, \quad 0 \leq \gamma_1 \leq 1, \tag{9}$$

where

$$\mathrm{Eff}_{C,J}(\xi;\delta,\theta_J)=\frac{I_{C,J}(\xi;\delta,\theta_J)}{I_{C,J}(\xi^*_{C,J};\delta,\theta_J)},\quad \xi^*_{C,J}=\arg\max_{\xi} I_{C,J}(\xi;\delta,\theta_J),\quad C,J=Cl,G.$$

The function

$$I_{C,J}(\xi;\delta,\theta_J)=\inf_{\theta_C}\int_{d\in\mathcal{D}}\mathcal{I}\left\{p^J_{y_1y_2}(d;\delta,\theta_J),p^C_{y_1y_2}(d;\delta,\theta_C)\right\}\,d\xi(d),$$

is the KL-criterion proposed by López-Fidalgo et al (2007), where

$$\mathcal{I}\left\{p^J_{y_1y_2}(d;\delta,\theta_J),p^C_{y_1y_2}(d;\delta,\theta_C)\right\}=\sum_{y_1,y_2\in\{0,1\}}p^J_{y_1y_2}(d;\delta,\theta_J)\log\frac{p^J_{y_1y_2}(d;\delta,\theta_J)}{p^C_{y_1y_2}(d;\delta,\theta_C)}$$

is the Kullback-Leibler divergence between the true model $p^J_{y_1y_2}(x;\delta,\theta_J)$ and $p^C_{y_1y_2}(x;\delta,\theta_C)$, defined in formulas (2)-(6), with $C,J=Cl,G$.

Unfortunately, maximizing (9) could provide optimal doses that are unsafe in the sense that they are very different from the P-optimal dose,

$$d^P_C=\arg\max_{d\in\mathcal{D}}p^C_{10}(d;\delta,\theta_C),\quad C=Cl,G. \tag{10}$$

To overcome this problem, we propose to maximize criterion (9) subject to a constraint on a function of the probability of efficacy without toxicity. In more detail, given a design ξ,

$$\Phi^P_C(\xi;\delta,\theta_C)=\int_{d\in\mathcal{D}}p^C_{10}(d;\delta,\theta_C)\,d\xi(d),\quad C=Cl,G \tag{11}$$

is the marginal probability of efficacy without toxicity (McGree and Eccleston, 2008), which is maximized by $\xi^P_C=\arg\max_\xi\Phi^P_C(\xi;\delta,\theta_C)$. It is easy to prove that ξ^P_C is the design which concentrates the whole mass at the optimal dose d^P_C given in (10). A measure of the "goodness" of a design ξ, in terms of safety and efficacy, is

$$0\le\mathrm{Eff}^P_C(\xi;\delta,\theta_C)=\frac{\Phi^P_C(\xi;\delta,\theta_C)}{\Phi^P_C(\xi^P_C;\delta,\theta_C)}\le 1,\quad C=Cl,G \tag{12}$$

which is herein called P-efficiency of ξ. Let us consider the following geometric mean of P-efficiencies

$$\Phi_P(\xi;\delta,\theta_{Cl},\theta_G)=\left\{\mathrm{Eff}^P_{Cl}(\xi;\delta,\theta_{Cl})\right\}^{\gamma_2}\cdot\left\{\mathrm{Eff}^P_G(\xi;\delta,\theta_G)\right\}^{1-\gamma_2},\quad 0\le\gamma_2\le 1; \tag{13}$$

we have that the larger $\Phi_P(\xi;\delta,\theta_{Cl},\theta_G)$, the safer and more efficacious ξ, under both the rival copulae.

Hence, in order to discriminate between the two competing models, we propose to maximize $\Phi_{KL}(\xi;\delta,\theta_{Cl},\theta_G)$ subject to the constraint

$$\Phi_P(\xi;\delta,\theta_{Cl},\theta_G)\ge c, \tag{14}$$

where c represents the value of the probability of efficacy without toxicity the researcher wants to exceed to protect patients. From Cook and Wong (1994), this constrained design problem is equivalent to the following compound criterion, which is called PKL-criterion:

$$\Phi_{PKL}(\xi;\delta,\theta_{Cl},\theta_G) = \{\Phi_{KL}(\xi;\delta,\theta_{Cl},\theta_G)\}^{\gamma_3}\cdot\{\Phi_P(\xi;\delta,\theta_{Cl},\theta_G)\}^{1-\gamma_3}, \quad 0 \le \gamma_3 \le 1. \tag{15}$$

For ease of notation, in what follows we omit δ from the argument of the functions, even if they depend on the model parameter δ.
Maximizing (15) is equivalent to maximize

$$\begin{aligned}\Psi_{PKL}(\xi;\theta_{Cl},\theta_G) &= \log\Phi_{PKL}(\xi;\theta_{Cl},\theta_G)\\ &= \gamma_3\log\Phi_{KL}(\xi;\theta_{Cl},\theta_G) + (1-\gamma_3)\log\Phi_P(\xi;\theta_{Cl},\theta_G).\end{aligned} \tag{16}$$

From Lemma 1 in Cook and Wong (1994) we may state the following theorem that relates the weight γ_3 in (15) with the constant c in (14):

Theorem 1 *Given $\gamma_3 \in (0,1)$, if $\xi_{PKL}^{\gamma_3} = \arg\max_\xi \Psi_{PKL}(\xi;\theta_{Cl},\theta_G)$ then*

$$\begin{aligned}\xi_{PKL}^{\gamma_3} &= \arg\max_{\xi}\Phi_{KL}(\xi;\theta_{Cl},\theta_G) \text{ subject to the constraint}\\ \Phi_P(\xi;\theta_{Cl},\theta_G) &\ge c_{\gamma_3}, \text{ where } c_{\gamma_3} = \Phi_P(\xi_{PKL}^{\gamma_3};\theta_{Cl},\theta_G).\end{aligned} \tag{17}$$

The PKL-optimum design $\xi_{PKL}^{\gamma_3} = \arg\max_\xi \Psi_{PKL}(\xi;\theta_{Cl},\theta_G)$ exists since criterion function (16) is concave, as it is a convex combination of concave optimality criteria (for a proof of the concavity of the KL-criterion see Tommasi (2007); it is also easy to prove that $\log\Phi_C^P(\xi;\delta,\theta_C)$ is concave as well). Furthemore, the following equivalence theorem may be stated:

Theorem 2 *A design $\xi_{PKL}^{\gamma_3}$ is PKL-optimum if and only if the following inequality is satisfied:*

$$\begin{aligned}&\gamma_3\left[\gamma_1\frac{\mathcal{I}\{p_{y_1y_2}^{Cl}(d;\theta_{Cl}),p_{y_1y_2}^{G}(d;\theta_G)\}}{I_{G,Cl}(\xi_{PKL}^{\gamma_3};\theta_{Cl})} + (1-\gamma_1)\frac{\mathcal{I}\{p_{y_1y_2}^{G}(d;\theta_G),p_{y_1y_2}^{Cl}(d;\theta_{Cl})\}}{I_{Cl,G}(\xi_{PKL}^{\gamma_3};\theta_G)}\right]\\ &+(1-\gamma_3)\left[\gamma_2\frac{p_{10}^{Cl}(d;\theta_{Cl})}{\Phi_{Cl}^P(\xi_{PKL}^{\gamma_3};\theta_{Cl})} + (1-\gamma_2)\frac{p_{10}^{G}(d;\theta_G)}{\Phi_G^P(\xi_{PKL}^{\gamma_3};\theta_G)}\right] - 1 \le 0, \quad d\in\mathcal{D}.\end{aligned} \tag{18}$$

The left-hand side of inequality (18) is the directional derivative of the PKL-criterion (16) evaluated at $\xi_{PKL}^{\gamma_3}$ in the direction of $\xi_d - \xi_{PKL}^{\gamma_3}$ (theoretical details are provided in Appendix A). The analytical expression of the directional derivative is useful to check the optimality of a design as well as to apply the first order algorithm in order to compute the PKL-optimum design numerically; see §3.2 in Fedorov and Hackl (1997) and Fedorov and Leonov (2014).

If a researcher aims at considering both the problems of model discrimination and parameter estimation at the same design stage, then a DKL-criterion could be used (Tommasi, 2009). Even in this case we suggest to penalize with respect to (13).

Remark 1. The optimality criterion (15) depends on the choice of γ_1, γ_2 and γ_3. Weight γ_1 reflects the relative importance of the two rival copula models. Let $\xi_{KL}^{\gamma_1} = \arg\max_\xi \Phi_{KL}(\xi;\theta_{Cl},\theta_G)$ be the best design to discriminate between the two copulae. Choosing γ_1 equal to 0.5 does not necessarily imply equal belief in the competing models, thus following Cook and Wong (1994) we suggest to choose the value γ_1^* such that $\mathrm{Eff}_{G,Cl}\big(\xi_{KL}^{\gamma_1^*};\theta_{Cl}\big) = \mathrm{Eff}_{Cl,G}\big(\xi_{KL}^{\gamma_1^*};\theta_G\big)$. In the same way, let $\xi_P^{\gamma_2} = \arg\max_\xi \Phi_P(\xi;\theta_{Cl},\theta_G)$ for a given γ_2. We suggest to use the value γ_2^* such that $\mathrm{Eff}_{Cl}^P(\xi_P^{\gamma_2^*};\theta_{Cl}) = \mathrm{Eff}_G^P(\xi_P^{\gamma_2^*};\theta_G)$.
Differently, $(1-\gamma_3)$ reflects the degree of protection from unsafe designs as expressed by the constraint (17): the smaller γ_3 the safer the optimal design. Therefore, the optimal design $\xi_{PKL}^{\gamma_3}$ and the threshold c_{γ_3} should be computed for several values of γ_3. Then, the researcher can choose the best PKL-optimum design depending on the degree of protection c_{γ_3} that he/she prefers.

For three values of $(\theta_{Cl};\theta_G)$ (corresponding to three different values of τ) and for several values of γ_3 ($\gamma_1 = \gamma_1^*$ and $\gamma_2 = \gamma_2^*$, as described in Remark 1), Table 4 reports: $\xi_{PKL}^{\gamma_3}$, the KL-efficiency, i.e.

$$\mathrm{Eff}_{KL}(\xi_{PKL}^{\gamma_3}) = \frac{\Phi_{KL}(\xi_{PKL}^{\gamma_3};\theta_{Cl},\theta_G)}{\Phi_{KL}\big(\xi_{KL}^{\gamma_1^*};\theta_{Cl},\theta_G\big)}, \tag{19}$$

and the threshold c_{γ_3} given in (17). For instance, if $\gamma_3 = 0.6$, $\xi_{PKL}^{\gamma_3}$ provides a good performance to discriminate between dependence structures, since the KL-efficiency of $\xi_{PKL}^{\gamma_3}$ is greater than 0.90 for all the values of $(\theta_{Cl};\theta_G)$. In addition, according to (17), $\xi_{PKL}^{\gamma_3}$ also guarantees a quite high probability (around 0.50) of efficacy without toxicity.

Results in Table 4 have been obtained by running a computer code written in Mathematica. The code is freely available upon request to the authors.

6 Simulation study

In order to assess the ability of the PKL-optimum design to discriminate between two competing copula models we employ the likelihood ratio test. In some sense, we apply Cox's test (see Cox (1961) and Cox (1962)) to compare non-nested[1] models, but instead of using the asymptotic distribution proposed by Cox, we consider the Monte Carlo distribution of the log-likelihood ratio.

For a specific Scenario δ and for a specific value of Kendall's τ coefficient, we generate M samples of size n, at the PKL-optimum design $\xi_{PKL}^{\gamma_3}$, from one of the two rival models. Then, we check how many times the likelihood ratio test provides an evidence in favour of each model.

[1] In non-nested hypotheses neither model can be obtained from the other by imposing a parametric restriction.

Table 4: PKL-optimal designs, their KL-efficiencies and thresholds c_{γ_3}

$(\theta_{Cl};\theta_G)$	γ_3	$\xi_{PKL}^{\gamma_3}$	$\text{Eff}_{KL}(\xi_{PKL}^{\gamma_3})$	c_{γ_3}
(2;2)	0.2	$\left\{\begin{matrix}-0.605 & -0.267\\ 0.530 & 0.470\end{matrix}\right\}$	0.446	0.767
	0.4	$\left\{\begin{matrix}-0.686 & -0.200\\ 0.478 & 0.522\end{matrix}\right\}$	0.747	0.620
	0.6	$\left\{\begin{matrix}-0.750 & -0.150\\ 0.465 & 0.535\end{matrix}\right\}$	0.911	0.504
	0.8	$\left\{\begin{matrix}-0.800 & -0.100\\ 0.460 & 0.540\end{matrix}\right\}$	0.986	0.410
	1	$\left\{\begin{matrix}-0.800 & -0.050\\ 0.440 & 0.560\end{matrix}\right\}$	1	0.368
(8;5)	0.2	$\left\{\begin{matrix}-0.600 & -0.157\\ 0.520 & 0.480\end{matrix}\right\}$	0.680	0.578
	0.4	$\left\{\begin{matrix}-0.663 & -0.132\\ 0.472 & 0.528\end{matrix}\right\}$	0.850	0.530
	0.6	$\left\{\begin{matrix}-0.700 & -0.100\\ 0.461 & 0.539\end{matrix}\right\}$	0.935	0.491
	0.8	$\left\{\begin{matrix}-0.750 & -0.063\\ 0.459 & 0.541\end{matrix}\right\}$	0.994	0.437
	1	$\left\{\begin{matrix}-0.793 & -0.050\\ 0.470 & 0.530\end{matrix}\right\}$	1	0.399
(18;10)	0.2	$\left\{\begin{matrix}-0.039 & 0.576\\ 0.583 & 0.417\end{matrix}\right\}$	0.590	0.795
	0.4	$\left\{\begin{matrix}-0.010 & 0.700\\ 0.517 & 0.483\end{matrix}\right\}$	0.829	0.689
	0.6	$\left\{\begin{matrix}-0.171 & 0.750\\ 0.472 & 0.528\end{matrix}\right\}$	0.910	0.647
	0.8	$\left\{\begin{matrix}-0.214 & 0.800\\ 0.441 & 0.559\end{matrix}\right\}$	0.983	0.558
	1	$\left\{\begin{matrix}-0.250 & 0.850\\ 0.413 & 0.587\end{matrix}\right\}$	1	0.504

6.1 Likelihood ratio test for rival copula-based models

Given δ, τ and a design ξ, let (y_{1i}, y_{2i}) for $i = 1, 2, ...n$ be a sample of efficacy and toxicity outcomes from one of the two rival models. Following Pesaran and Weeks (2001) the problem is to test both the following systems of hypotheses:

$$A)\ \begin{cases} H_{Cl} : \mathcal{F}_{Cl} = \{p^{Cl}_{y_1y_2}(d;\delta,\theta_{Cl}),\ \theta_{Cl} \in \Theta_{Cl}\} \\ H_G :\ \mathcal{F}_G = \{p^{G}_{y_1y_2}(d;\delta,\theta_G),\ \theta_G \in \Theta_G\} \end{cases}$$

$$B)\ \begin{cases} H_G :\ \mathcal{F}_G = \{p^{G}_{y_1y_2}(d;\delta,\theta_G),\ \theta_G \in \Theta_G\} \\ H_{Cl} : \mathcal{F}_{Cl} = \{p^{Cl}_{y_1y_2}(d;\delta,\theta_{Cl}),\ \theta_{Cl} \in \Theta_{Cl}\} \end{cases}$$

From now on, we omit the arguments d and δ for ease of notation. As test statistics, we consider the following log-likelihood ratios:

$$T_{ClG} = L_{Cl}(\widehat{\theta}_{Cl}) - L_G(\widehat{\theta}_G) \quad \text{and} \quad T_{GCl} = L_G(\widehat{\theta}_G) - L_{Cl}(\widehat{\theta}_{Cl}), \tag{20}$$

Table 5: PKL-optimal design, KL-efficiency and threshold c_{γ_3} for $(\theta_{Cl};\theta_G) = (8;5)$ and $\gamma_3 = 0.17$

$\xi_{PKL}^{\gamma_3}$	$\text{Eff}_{KL}(\xi_{PKL}^{\gamma_3})$	c_{γ_3}
$\left\{\begin{matrix} -0.587 & -0.187 \\ 0.531 & 0.469 \end{matrix}\right\}$	0.601	0.594

where $L_{Cl}(\theta_{Cl})$ and $L_G(\theta_G)$ are the log-likelihood functions under H_{Cl} and H_G, respectively, and $\widehat{\theta}_{Cl}$ and $\widehat{\theta}_G$ are the corresponding maximum likelihood estimators[2] of θ_{Cl} and θ_G.

Let p_{ClG} and p_{GCl} be the p-values of T_{ClG} and T_{GCl}, respectively. Given a significance level $\widetilde{\alpha}$, the test of hypothesis can lead to four different decisions:

a) If $p_{ClG} < \widetilde{\alpha}$ and $p_{GCl} \geq \widetilde{\alpha}$, we reject Clayton and accept Gumbel;
b) If $p_{ClG} \geq \widetilde{\alpha}$ and $p_{GCl} < \widetilde{\alpha}$, we accept Clayton and reject Gumbel;
c) If $p_{ClG} \geq \widetilde{\alpha}$ and $p_{GCl} \geq \widetilde{\alpha}$, we accept Clayton (or Gumbel) when $p_{ClG} > p_{GCl}$ (or $p_{GCl} > p_{ClG}$);
d) If $p_{ClG} < \widetilde{\alpha}$ and $p_{GCl} < \widetilde{\alpha}$, we reject Clayton (or Gumbel) when $p_{ClG} < p_{GCl}$ (or $p_{GCl} < p_{ClG}$).

In other words, we suggest to accept Clayton (or Gumbel) model whenever $p_{ClG} > p_{GCl}$ (or $p_{GCl} > p_{ClG}$).

In the case of non-nested models the log-likelihood ratio is not (asymptotically) distributed as a Chi-squared random variable (see for instance Cox (1962); Pesaran and Weeks (2001); Monfardini (2003)). Hence, we implement a Monte Carlo procedure to approximate the sample distribution of T_{ClG} and T_{GCl} and to compute the corresponding p-values, $\hat{p}_{ClG}$ and $\hat{p}_{GCl}$ under H_{Cl} and H_G, respectively. Differently, Cox (1961, 1962) proposed the asymptotic distribution of the log-likelihood ratio suitably standardized.

6.2 Simulation and results

For Scenario 3 of the Supplementary Material, $\delta = (1, 1.5, -3, 2.5, 5)$, and $\tau = 0.8$ we perform two Monte Carlo simulations, based on the generation of $M = 5000$ samples of size n from model (3) using a Clayton copula with $\theta_{Cl} = 8$ and a Gumbel copula with $\theta_G = 5$, respectively. The doses and the proportions of observations to be taken at each dose are given by the PKL-optimum design with $\gamma_3 = 0.17$, which is reported in Table 5. From the last two columns of Table 5 we can observe that this design is almost equally good for discriminating between rival copulae, according to the KL-efficiency (19), and for protecting patients against unsafe doses, according to the constraint (14).

[2] Observe that $\widehat{\theta}_{Cl}$ and $\widehat{\theta}_G$ are referred to as QML (quasi maximum likelihood) estimators when they are obtained under the not true hypotheses H_{CL} and H_G.

For the generation of the dichotomous response (Y_1, Y_2) in model (3) we consider the following latent response model with continuous dependent variable (Y_1^*, Y_2^*) (see Verbeek (2008), p.202). Let us assume that

$$Y_j = \begin{cases} 1 \; if \; Y_j^* > 0 \\ 0 \; if \; Y_j^* \leq 0 \end{cases} \quad j = 1, 2 \tag{21}$$

where, after the standardization (8),

$$Y_1^* = \alpha_0 + \alpha_1 d + \alpha_2 d^2 + \epsilon_1 = \eta_1(d; \delta) + \epsilon_1$$

$$Y_2^* = \beta_0 + \beta_1 d + \epsilon_2 = \eta_2(d; \delta) + \epsilon_2$$

and the random error (ϵ_1, ϵ_2) follows a bivariate standard logistic distribution with a dependence structure which fulfills Theorem 3 (see Appendix B).
In more detail:

We compute $\eta_1(d; \delta)$ and $\eta_2(d; \delta)$ at $d_1 = -0.587$ and $d_2 = -0.187$, which are the support points of the PKL-optimum design (see Table 5).

For M times we repeat the following steps:

1. We generate a random sample of n i.i.d. bivariate errors, $(\epsilon_{1i}, \epsilon_{2i})$, $i = 1, \ldots, n$, from the following cdf

   $$F_{\epsilon_1,\epsilon_2}(\tilde{\epsilon}_1, \tilde{\epsilon}_2; \theta_C) = F_{\epsilon_1}(\tilde{\epsilon}_1) + F_{\epsilon_2}(\tilde{\epsilon}_2) - 1 + C\big(1 - F_{\epsilon_1}(\tilde{\epsilon}_1), 1 - F_{\epsilon_2}(\tilde{\epsilon}_2); \theta_C\big),$$

   (see Equations (28) and (29)), where $F_{\epsilon_j}(\tilde{\epsilon}_j)$, $j = 1, 2$, denotes the marginal cdf of a standard logistic random variable and $C(\cdot, \cdot; \theta_C)$ is the Clayton copula with $\theta_{Cl} = 8$ (or the Gumbel copula with $\theta_G = 5$);
2. We compute

   $$\begin{cases} y_{1i}^* = \eta_1(d_1; \delta) + \epsilon_{1i} \\ y_{2i}^* = \eta_2(d_1; \delta) + \epsilon_{2i} \end{cases} \qquad i = 1, \cdots, n_1$$

   and

   $$\begin{cases} y_{1i}^* = \eta_1(d_2; \delta) + \epsilon_{1i} \\ y_{2i}^* = \eta_2(d_2; \delta) + \epsilon_{2i} \end{cases} \qquad i = 1, \cdots, n_2$$

   where n_1 and n_2 are obtained multipling $\xi(d_1) = 0.531$ and $\xi(d_2) = 0.469$ (given in Table 5) by n, and then using some rounding off rule (see for instance Chapter 12 in Pukelsheim (2006)).
3. For $i = 1, \cdots, n$ we obtain (y_{1i}, y_{2i}) by transforming (y_{1i}^*, y_{2i}^*) according to (21).
4. We compute the ML estimates of θ_{Cl} and θ_G to calculate the observed values of T_{ClG} and T_{GCl} given in (20).
5. We compute the Monte Carlo p-values (at the m-th step), $\hat{p}_{ClG}^m$ and $\hat{p}_{GCl}^m$ using the following subroutine:

   Subroutine (Monte Carlo p-value for T_{ClG})

   (a) Generate $R = 10000$ samples of size n from the model under H_{Cl} with $\theta_{Cl} = 8$;

   (b) For $r = 1, .., R$:

- Compute the estimates $(\widehat{\theta}^r_{Cl}, \widehat{\theta}^r_G)$ by maximizing the log-likelihood functions $L_{Cl}(\theta^r_{Cl})$ and $L_G(\theta^r_G)$ under H_{Cl} and H_G, respectively;
- Evaluate the log-likelihood ratio statistic

$$T^r_{ClG} = L_{Cl}(\widehat{\theta}^r_{Cl}) - L_G(\widehat{\theta}^r_G);$$

(c) Calculate the Monte Carlo p-value as

$$\hat{p}^m_{ClG} = \sum_{r=1}^{R} I(T^r_{ClG} \leq t^m_{ClG})/R$$

We can obtain the Monte Carlo p-value of T_{GCl} by reversing the role of Clayton and Gumbel models.

We calculate the percentages of correct selection of the true model, i.e. the percentage of times that $p^m_{ClG} > p^m_{GCl}$ for $m = 1, \ldots, M$, when the data are generated from the Clayton copula, and the percentage of times that $p^m_{GCl} > p^m_{ClG}$ for $m = 1, \ldots, M$, when the data are generated from the Gumbel copula.

The simulation results are reported in the third and the forth columns of Table 6.

Table 6: Monte Carlo simulation of the likelihood ratio test ($M = 5000$): data generated from Clayton and Gumbel copulae for $\tau = 0.8$, at the PKL-optimum design $\xi^{\gamma_3=0.17}_{PKL}$ (columns 3-4) and at the KL-optimum design $\xi_{KL} = \xi^{\gamma_3=1}_{PKL}$ (columns 5-6)

		$\xi^{\gamma_3=0.17}_{PKL}$		ξ_{KL}	
		True copula model (%)		**True copula model (%)**	
n	**Test decision**	Clayton	Gumbel	Clayton	Gumbel
100	Correct decision	70.58	71.86	72	71.88
	Wrong decision	29.42	28.14	28	28.12
200	Correct decision	81.46	81.4	82.28	84.04
	Wrong decision	18.54	18.6	17.72	15.96
500	Correct decision	91.34	91.06	95.56	95.4
	Wrong decision	8.66	8.94	4.44	4.6
1000	Correct decision	97.52	97.58	99.5	99.2
	Wrong decision	2.48	2.42	0.5	0.8

We can observe that the percentage of correct decision is always much greater than that of wrong decision. Its value is around 70% from $n = 100$ and it exceed 90% for $n = 500$. Furthermore, the percentage of wrong decision decreases substantially as n increases. Taking into account that the competing models differ only for the tail dependence, the obtained results are excellent. Perhaps continuous response variables guarantee better percentages even with a smaller sample size. This will be a matter of future research.

In order to compare the performance of the PKL-optimum design with the unconstrained KL-optimum one (reported in Table 4 for $(\theta_{Cl}; \theta_G) = (8; 5)$ and $\gamma_3 = 1$), we repeat the same simulation study generating data at the KL-optimum design. The corresponding percentages of correct decision and

wrong decision (listed in the last two columns of Table 6) show that the KL-optimum design is slightly better than its constrained PKL-version. Hence, we can conclude that the introduction of the penalization in the KL-criterion does not have a large negative effect on the discrimination ability.

7 Conclusion

In the last years, toxicity and efficacy are jointly studied in dose-finding methodologies. Many of these studies assume a specific dependence structure to model the relationship between the probabilities of efficacy and toxicity. Since the underlying dependence structure is sometimes unknown, our goal is to decide which specific copula is to be employed whenever two distinct copulae yield to a different P-optimal dose (the dose which maximizes the probability of efficacy without toxicity). More specifically, we consider as competing models the Clayton and Gumbel copulae, which both allow for positive association even if they differ for tail dependence. From a robustness study we observe that the P-optimal dose changes considerably (under the two copulae) when the probability of toxicity overcomes that of efficacy (at each dose). Hence, in this setting it is fundamental to determine the proper dependence structure in order to identify the P-optimal dose.

To this aim, we propose the PKL-criterion which is a constrained version of the KL-optimality and depends on the Scenario δ, the association parameter τ and the weights γ_1, γ_2, γ_3. To apply our method, we suggest the following scheme:

a) Guess a value for the Scenario δ and the association level τ from a pilot study, an expert opinion or past experiences.
b) Given δ and the copula parameters θ_{Cl} and θ_G corresponding to τ (see Table 3), compute the P-optimal doses under the two rival copulae applying (7). When the P-optimal doses are very different, then it is necessary to discriminate between the competing copulae.
c) Fix the weights γ_1, γ_2 and γ_3 as described in Remark 1.
d) Compute the PKL-optimum design applying (17).
e) Run the experiment in order to collect the data and finally apply the selection method (based on Cox's test) described in Section 6.

The PKL-optimum design is good to discriminate between the two rival copulae as well as to protect patients against unsafe doses. These two goals could be also achieved using the penalization approach described in Dragalin and Fedorov (2006) and Dragalin et al (2008) but, differently from their proposal, by choosing the value of γ_3 we can control the amount of protection against dangerous doses. A simulation study shows that the PKL-optimal design is really able to discriminate between the rival copulae despite the constraint introduced to avoid doses that are far away from the P-optimal one.

References

Atkinson AC, Fedorov VV (1975a) The design of experiments for discriminating between two rival models. Biometrika 62(1):57–70

Atkinson AC, Fedorov VV (1975b) Optimal design: Experiments for discriminating between several models. Biometrika 62(2):289–303

Cook R, Wong W (1994) On the equivalence of constrained and compound optimal designs. Journal of the American Statistical Association 89(426):687–692

Cox D (1961) TESTS OF SEPARATE FAMILIES OF HYPOTHESES, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistic and Probability. University of California Press: Berkeley

Cox D (1962) Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society B 24:406–424

Deldossi L, Osmetti SA, Tommasi C (2016) PKL-Optimality Criterion in Copula Models for efficacy-toxicity response. In mODa 11 - Advances in Model-Oriented Design and Analysis: Proceedings of the 11th International Workshop in Model-Oriented Design and Analysis, Kunert, J., Muller, C.H., Atkinson, A.C. (eds). Springer International Publishing: Heidelberg

Denman N, McGree J, Eccleston J, Duffull S (2011) Design of experiments for bivariate binary responses modelled by copula functions. Computational Statistics & Data Analysis 55(4):1509 – 1520

Dette H, Titoff S (2009) Optimal discrimination designs. Annals of Statistics 37(4):2056–2082

Dragalin V, Fedorov V (2006) Adaptive designs for dose-finding based on efficacy-toxicity response. Journal of Statistical Planning and Inference 136(6):1800 – 1823

Dragalin V, Fedorov V, Wu Y (2008) A two-stage design for dose-finding that accounts for both efficacy and safety. Statistics in Medicine 27(25):5156–5176

Drovandi CC, McGree JM, Pettitt AN (2014) A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design. Journal of Computational and Graphical Statistics 23(1):3–24

Fedorov V, Hackl P (1997) Model-Oriented Design of Experiments. Springer: New York

Fedorov V, Leonov S (2014) Optimal Design for Nonlinear Response Models. Chapman and Hall - CRC Press: Boca Raton

Gao L, Rosenberger W (2013) Adaptive Bayesian design with penalty based on toxicity-efficacy response. In mODa 10 - Advances in Model-Oriented Design and Analysis, Ucínski, D., Atkinson, A.C., Patan, M. (eds). Springer International Publishing: Heidelberg

Kim S, Flournoy N (2015) Optimal experimental design for systems with bivariate failures under a bivariate Weibull function. Journal of the Royal Statistical Society: Series C (Applied Statistics) 64(3):413–432

Ponce de Leon AC, Atkinson AC (1991) Optimum experimental design for discriminating between two rival models in the presence of prior information.

Biometrika 78(3):601–608

López-Fidalgo J, Tommasi C, Trandafir P (2007) An optimal experimental design criterion for discriminating between non-normal models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(2):231–242

López-Fidalgo J, Tommasi C, Trandafir PC (2008) Optimal designs for discriminating between some extensions of the Michaelis-Menten model. Journal of Statistical Planning and Inference 138(12):3797 – 3804

McGree J, Eccleston J (2008) Probability-based optimal design. Australian & New Zealand Journal of Statistics 50(1):13–28

Monfardini C (2003) An illustration of Cox's non-nested testing procedure for logit and probit models. Computational Statistics & Data Analysis 42(3):425 – 444

Nelsen R (2006) An Introduction to Copulas. Springer: New York

Perrone E, Müller W (2016) Optimal designs for copula models. Statistics 50(4):917–929

Perrone E, Rappold A, Müller WG (2017) Ds-optimality in copula models. Statistical Methods & Applications 26(3):403–418

Pesaran H, Weeks M (2001) Non-nested Hypothesis Testing: An Overview. In: A Companion to Theoretical Econometrics. BH Baltagi (eds). Blackwell Publishing Ltd: Malden, Ma, USA

Pukelsheim F (2006) Optimal design of experiments. SIAM: Philadelphia

Tao Y, Liu J, Li Z, Lin J, Lu T, Yan F (2013) Dose-finding based on bivariate efficacy-toxicity outcome using archimedean copula. PLoS ONE 8(11):1–6

Thall P (2012) Bayesian adaptive dose-finding based on efficacy and toxicity. Journal of Statistical Research 46:187–202

Thall P, Cook J (2004) Dose-finding based on efficacy-toxicity trade-offs. Biometrics 60(3):684–693

Tommasi C (2007) Optimal Designs for Discriminating among Several Non-Normal Models. In mODa 8 - Advances in Model-Oriented Design and Analysis: Proceedings of the 8th International Workshop in Model-Oriented Design and Analysis, López-Fidalgo, Jess; Rodrguez-Daz, Juan Manuel; Torsney, Ben (Eds.). Springer International Publishing: Heidelberg

Tommasi C (2009) Optimal designs for both model discrimination and parameter estimation. Journal of Statistical Planning and Inference 139(12):4123 – 4132

Uciński D, Bogacka B (2005) T-optimum designs for discrimination between two multiresponse dynamic models. Journal of the Royal Statistical Society Series B (Statistical Methodology) 67(1):3–18

Verbeek M (2008) A Guide to Modern Econometrics. John Wiley & Sons: Chichester

Yuan Y, Guosheng Y (2009) Bayesian dose finding by jointly modelling toxicity and efficacy as time-to-event outcomes. Journal of the Royal Statistical Society: Series C (Applied Statistics) 58(5):719–736

APPENDIX A: Theoretical details

Taking into account equations (9) and (13), criterion function (16) becomes

$$\Psi_{PKL}(\xi;\theta_{Cl},\theta_G) = \gamma_3 \left[\gamma_1 \log \frac{I_{G,Cl}(\xi;\theta_{Cl})}{I_{G,Cl}(\xi^*_{G,Cl};\theta_{Cl})} + (1-\gamma_1)\log \frac{I_{Cl,G}(\xi;\theta_G)}{I_{Cl,G}(\xi^*_{Cl,G};\theta_G)}\right] + (1-\gamma_3)\left[\gamma_2 \log \frac{\Phi^P_{Cl}(\xi;\theta_{Cl})}{\Phi^P_{Cl}(\xi^P_{Cl};\theta_{Cl})} + (1-\gamma_2)\log \frac{\Phi^P_G(\xi;\theta_G)}{\Phi^P_G(\xi^P_G;\theta_G)}\right]. \tag{22}$$

Except for a constant term, from (22) we have that

$$\Psi_{PKL}(\xi;\theta_{Cl},\theta_G) = \gamma_3\left[\gamma_1 \log I_{G,Cl}(\xi;\theta_{Cl}) + (1-\gamma_1)\log I_{Cl,G}(\xi;\theta_G)\right] + (1-\gamma_3)\left[\gamma_2 \log \Phi^P_{Cl}(\xi;\theta_{Cl}) + (1-\gamma_2)\log \Phi^P_G(\xi;\theta_G)\right] \tag{23}$$

The directional derivative of $\Psi_{PKL}(\xi;\theta_{Cl},\theta_G)$ at ξ in the direction of $\xi_d - \xi$ can be easily obtained from the expressions of the corresponding directional derivatives of $I_{i,j}(\xi;\theta_j)$, $i,j = G, Cl$ and $\Phi^P_C(\xi;\theta_C)$, $C = G, Cl$, respectively.

Assuming that the true model is $p^j_{y_1y_2}(x;\theta_j)$, we recall that

$$\partial I_{i,j}(\xi,\xi_d;\theta_j) = \mathcal{I}\{p^j_{y_1y_2}(d;\theta_j), p^i_{y_1y_2}(d;\theta_i)\} - I_{i,j}(\xi;\theta_j), \qquad i,j = G, Cl, \tag{24}$$

see López-Fidalgo et al (2007).

The directional derivative of $\Phi^P_C(\xi;\theta_C)$ at ξ in any direction $\bar{\xi} - \xi$ is:

$$\begin{aligned}\partial\Phi^P_C(\xi,\bar{\xi};\theta_C) &= \lim_{\alpha\to 0^+} \frac{\Phi^P_C\left[(1-\alpha)\xi + \alpha\bar{\xi};\theta_C\right] - \Phi^P_C(\xi;\theta_C)}{\alpha} \\ &= \lim_{\alpha\to 0^+} \frac{(1-\alpha)\Phi^P_C(\xi;\theta_C) + \alpha\Phi^P_C(\bar{\xi};\theta_C) - \Phi^P_C(\xi;\theta_C)}{\alpha} \\ &= \Phi^P_C(\bar{\xi};\theta_C) - \Phi^P_C(\xi;\theta_C), \qquad C = Cl, G,\end{aligned}$$

where the second equality is due to the linearity of the criterion $\Phi^P_C(\xi;\theta_C)$. Therefore, taking into account equation (11),

$$\partial\Phi^P_C(\xi,\bar{\xi};\theta_C) = \int_{d\in\mathcal{D}} \left[p^C_{10}(d;\theta_C) - \int_{d\in\mathcal{D}} p^C_{10}(d;\theta_C)\, d\,\xi(d)\right] d\bar{\xi}(d).$$

From this last expression, the directional derivative of $\Phi^P_C(\xi;\theta_C)$ at ξ in the direction of $\xi_d - \xi$ is

$$\partial\Phi^P_C(\xi,\xi_d;\theta_C) = p^C_{10}(d;\theta_C) - \int_{d\in\mathcal{D}} p^C_{10}(d;\theta_C)\, d\,\xi(d). \tag{25}$$

From (23), taking into account Equations (24) and (25), we have that

$$\begin{aligned}
\partial\Psi_{PKL}(\xi;\xi_d) &= \gamma_3\left[\gamma_1\frac{\mathcal{I}\{p^{Cl}_{y_1y_2}(d;\theta_{Cl}),p^{G}_{y_1y_2}(d;\theta_G)\}-I_{G,Cl}(\xi;\theta_{Cl})}{I_{G,Cl}(\xi;\theta_{Cl})}\right.\\
&\quad\left.+\ (1-\gamma_1)\frac{\mathcal{I}\{p^{G}_{y_1y_2}(d;\theta_G),p^{Cl}_{y_1y_2}(d;\theta_{Cl})\}-I_{Cl,G}(\xi;\theta_G)}{I_{Cl,G}(\xi;\theta_G)}\right]\\
&\quad+(1-\gamma_3)\left[\gamma_2\frac{p^{Cl}_{10}(d;\theta_{Cl})-\Phi^P_{Cl}(\xi;\theta_{Cl})}{\Phi^P_{Cl}(\xi;\theta_{Cl})}+(1-\gamma_2)\frac{p^{G}_{10}(d;\theta_G)-\Phi^P_{G}(\xi;\theta_G)}{\Phi^P_{G}(\xi;\theta_G)}\right].\\
&= \gamma_3\left[\gamma_1\frac{\mathcal{I}\{p^{Cl}_{y_1y_2}(d;\theta_{Cl}),p^{G}_{y_1y_2}(d;\theta_G)\}}{I_{G,Cl}(\xi;\theta_{Cl})}+(1-\gamma_1)\frac{\mathcal{I}\{p^{G}_{y_1y_2}(d;\theta_G),p^{Cl}_{y_1y_2}(d;\theta_{Cl})\}}{I_{Cl,G}(\xi;\theta_G)}\right]\\
&\quad+(1-\gamma_3)\left[\gamma_2\frac{p^{Cl}_{10}(d;\theta_{Cl})}{\Phi^P_{Cl}(\xi;\theta_{Cl})}+(1-\gamma_2)\frac{p^{G}_{10}(d;\theta_G)}{\Phi^P_{G}(\xi;\theta_G)}\right]-1.
\end{aligned}$$

APPENDIX B: Latent representation of the model

The random error (ϵ_1,ϵ_2) in the latent representation (21) is distributed as a bivariate standard logistic distribution which fulfills the following theorem (for ease of notation, in what follows we omit d and δ).

Theorem 3 *If $\hat{C}(\cdot,\cdot;\theta_C)$ is the copula that defines the cdf of the bivariate error (ϵ_1,ϵ_2), according to the Sklar's theorem, then $P(Y_1=1,Y_2=1;\theta_C)=C(\pi_1,\pi_2;\theta_C)$ where $C(\cdot,\cdot;\theta_C)$ is the survival copula of $\hat{C}(\cdot,\cdot;\theta_C)$. Vice versa, if $P(Y_1=1,Y_2=1;\theta_C)=C(\pi_1,\pi_2;\theta_C)$ where $C(\cdot,\cdot;\theta_C)$ is a copula function, then the cdf of (ϵ_1,ϵ_2) is defined by the survival copula of $C(\cdot,\cdot;\theta_C)$.*

Proof Let us recall that given a copula $\hat{G}(\cdot,\cdot;\theta)$, the corresponding survival copula is

$$G(u,v;\theta)=u+v-1+\hat{G}(1-u,1-v;\theta). \tag{26}$$

(see Nelsen (2006)). In addition, let $F_{\epsilon_1,\epsilon_2}(\cdot,\cdot;\theta_C)$ and $F_{\epsilon_i}(\cdot)$ be the joint cdf and the marginal cdf of the errors ϵ_j with $j=1,2$.

If we assume that $F_{\epsilon_1,\epsilon_2}(\tilde{\epsilon}_1,\tilde{\epsilon}_2;\theta_C)=\hat{C}(F_{\epsilon_1}(\tilde{\epsilon}_1),F_{\epsilon_2}(\tilde{\epsilon}_2);\theta_C)$, where $\hat{C}(\cdot,\cdot;\theta_C)$ is a copula function, then

$$\begin{aligned}
P(Y_1=1,Y_2=1;\theta_C) &= P(Y_1^*>0,Y_2^*>0;\theta_C)=P(\epsilon_1>-\eta_1,\epsilon_2>-\eta_2;\theta_C)\\
&= 1-P(\epsilon_1<-\eta_1)-P(\epsilon_2<-\eta_2)+P(\epsilon_1<-\eta_1,\epsilon_2<-\eta_2;\ \theta_C)\\
&= \overline{F}_{\epsilon_1}(-\eta_1)+\overline{F}_{\epsilon_2}(-\eta_2)-1\\
&\quad+\hat{C}\big(1-\bar{F}_{\epsilon_1}(-\eta_1),1-\bar{F}_{\epsilon_2}(-\eta_2)\,;\,\theta_C\big)
\end{aligned} \tag{27}$$

where $\overline{F}_{\epsilon_j}(\cdot)=1-F_{\epsilon_j}(\cdot)$ is the survival function of ϵ_j, $j=1,2$.
The first statement of the theorem is proved by comparing Equations (26) and (27).

On the other side, if $P(Y_1 = 1, Y_2 = 1; \theta_C) = C(\pi_1, \pi_2; \theta_C)$, where $C(\cdot, \cdot; \theta_C)$ is a copula function, then

$$\begin{aligned} P(Y_1 = 0, Y_2 = 0; \theta_C) &= 1 - \pi_1 - \pi_2 + C(\pi_1, \pi_2; \theta_C) \\ &= F_{\epsilon_1}(-\eta_1) + F_{\epsilon_2}(-\eta_2) - 1 \\ &+ C\big(1 - F_{\epsilon_1}(-\eta_1), 1 - F_{\epsilon_2}(-\eta_2)\,;\, \theta_C\big). \end{aligned} \tag{28}$$

By comparing (28) with (26) it is easy to show that $P(Y_1 = 0, Y_2 = 0; \theta_C)$ is defined by the survival copula of $C(\cdot, \cdot; \theta_C)$. However it should be noted also that

$$\begin{aligned} P(Y_1 = 0, Y_2 = 0; \theta_C) &= P(Y_1^* < 0, Y_2^* < 0; \theta_C) = P(\epsilon_1 < -\eta_1, \epsilon_2 < -\eta_2; \theta_C) \\ &= F_{\epsilon_1, \epsilon_2}(-\eta_1, -\eta_2; \theta_C). \end{aligned} \tag{29}$$

Thus, the second statement of the theorem follows from Equations (28) and (29).

SUPPLEMENTARY MATERIAL: P-optimal dose under different scenarios

Let $d \in [-1, 1]$ denote a dose as defined in (8). To understand how the P-optimal dose given in (10) changes accordingly to the assumed dependence structure $C(\cdot, \cdot; \theta_C)$, we have considered several different settings for δ. Herein, we describe just three scenarios as representatives of three different cases:

a) It is not necessary to take into consideration the dependence structure: $p_{10}^{Cl}(d; \delta, \theta_{Cl})$ and $p_{10}^{G}(d; \delta, \theta_G)$ give P-optimal doses close to that obtained in the independence case;
b) It is necessary to model the dependence but $p_{10}^{Cl}(d; \delta, \theta_{Cl})$ and $p_{10}^{G}(d; \delta, \theta_G)$ give almost the same P-optimal dose, hence discrimination is unnecessary;
c) It is relevant to discriminate between Clayton and Gumbel copulae as $p_{10}^{C}(d; \delta, \theta_C)$ leads to different P-optimal doses for $C = Cl, G$.

Let Scenario 1 be $\delta = (1, 1.5, -0.5, -2, 1.5)$. As shown in Fig. 1 under this scenario the marginal probability of efficacy is greater than 0.5 and that of toxicity is less than 0.4, at each dose. It follows that the whole design region

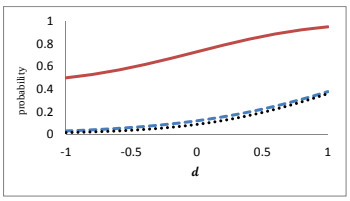

Fig. 1: Marginal probabilities of efficacy (red solid line) and toxicity (blue dashed line) and their joint probability in the independence case (black dotted line), for $\delta = (1, 1.5, -0.5, -2, 1.5)$ related to Scenario 1.

$\mathcal{D} = [-1, 1]$ may represent the so-called *therapeutic region* defined by the the minimum effective dose (MED) and the maximum tolerated dose (MTD) (Dragalin et al, 2008).

Let us recall that to measure the goodness of a dose d with respect to the P-optimal dose d_C^P we use the P-efficiency defined in (12):

$$\text{Eff}_C^P(d; \delta, \theta_C) = \frac{p_{10}^C(d; \delta, \theta_C)}{p_{10}^C(d_C^P; \delta, \theta_C)}.$$

Table 7 reports the P-optimal dose d_C^P, the P-efficiency of d_Π^P (optimal dose in the independence case) and the P-efficiency of $d_{C_F}^P$ (optimal dose under a misspecified copula) for Scenario 1. We can observe that the P-optimal dose under the independence assumption is quite similar to those obtained assuming different copula functions (and/or different values of the dependence parameter θ_C). Hence, the P-efficiencies of d_Π^P are all close to 1.

Table 7: P-optimal dose d_C^P, P-efficiency of d_Π^P (optimal dose in the independence case) and P-efficiency of $d_{C_F}^P$ (optimal dose under a misspecified copula) under Scenario 1

Scenario 1						
True Copula (C_T)	θ_{C_T}	d_C^P	$\text{Eff}_{C_T}^P(d_\Pi^P)$	**False Copula** (C_F)	θ_{C_F}	$\text{Eff}_{C_T}^P(d_{C_F}^P)$
	2	0.2538	0.9999		2	0.9999
Clayton	8	0.2479	0.9998	**Gumbel**	5	1
	18	0.2479	0.9998		10	0.9999
	2	0.2467	0.9998		2	0.9999
Gumbel	5	0.2479	0.9998	**Clayton**	8	1
	10	0.2479	0.9998		18	1
Independence copula Π		0.2654	1			

Actually, the region where the joint probability (3) takes its values is

$$\pi_1(x;\alpha)\cdot\pi_2(x;\beta) \le p_{11}^C(x;\delta,\theta) \le \min\{\pi_1(x;\alpha);\pi_2(x;\beta)\}, \tag{30}$$

(see Nelsen (2006) p.30). From (30) we have that the farther p_{11}^C is from the lower bound which corresponds to independence between efficacy and toxicity, the larger should be the effect of the dependence structure. According to (30), we can observe from Fig. 1 that p_{11}^C may assume values only in the area included between the blue dashed line and the black dotted one. As a consequence, the dependence structure (i.e. the copula function) cannot separate p_{11}^C too much from $\pi_1\cdot\pi_2$ and thus the probabilities of efficacy without toxicity p_{10}^C for the Clayton and the Gumbel copulae with the same τ are overlapping, as shown in Fig. 2.

Hence, for Scenario 1, clinicians can avoid to model toxicity and efficacy jointly by using Clayton or Gumbel copulae: the P-optimal dose can be obtained under the independence assumption. From our simulations, it seems that this kind of results holds when the marginal probability of efficacy is uniformly greater than the marginal probability of toxicity.

Consider now Scenario 2 defined by $\delta = (-1, 3, 0, -1, 4)$ where the marginal probability of efficacy and toxicity are quite similar for $d \le 0$, while for $d > 0$ the probability of toxicity is greater than that of efficacy (see Fig. 3 where the area included between the red solid line and the black dotted one defines the region where p_{11}^C may assume values). From the results reported in Table 8 we can observe that the losses in P-efficiency of d_Π^P increase with the association between efficacy and toxicity.

Differently from the previous scenario, in this case clinicians should model efficacy and toxicity jointly, since the P-efficiency of d_Π^P is quite low under both the rival copulae, except for θ_{Cl}=2 and θ_G=2 (compare the shapes of p_{10}^C in Fig. 4 under the different dependence structures). From the right-hand side of Table 8, however, we can observe that the P-efficiency of the P-optimal dose under a misspecified copula, $d_{C_F}^P$, is large and increases with θ_C. Hence, even if it is relevant to take into consideration the dependence structure (because

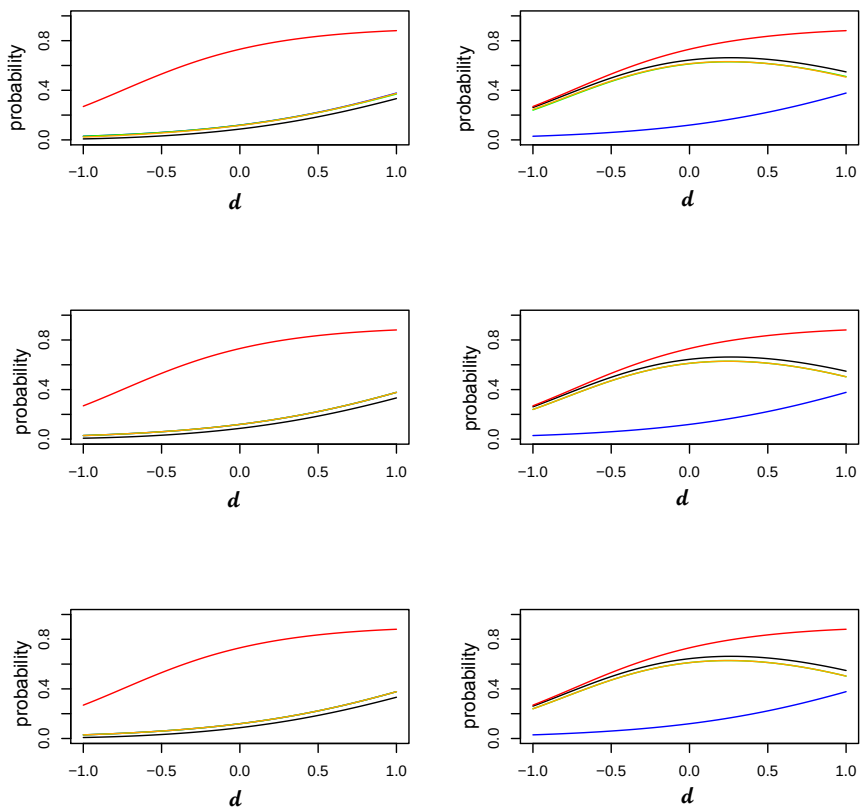

Fig. 2: Marginal probabilities of efficacy (red line) and toxicity (blue line); $p_{11}^{C}(d;\delta,\theta_C)$ (left-side) and $p_{10}^{C}(d;\delta,\theta_C)$ (right-side) for: the independence situation (black line), $C = Cl$ (green line) and $C = G$ (orange line), with three different values of τ (see Table 3): $\tau = 0.5$ (first row), $\tau = 0.8$ (second row) and $\tau = 0.9$ (last row), at $\delta = (1, 1.5, -0.5, -2, 1.5)$ (Scenario 1).

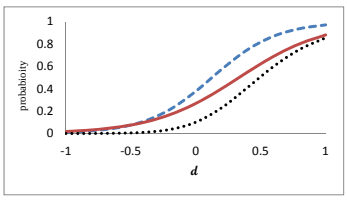

Fig. 3: Marginal probabilities of efficacy (red solid line) and toxicity (blue dashed line) and their joint probability in the independence case (black dotted line), for $\delta = (-1, 3, 0, -1, 4)$ related to Scenario 2.

Table 8: P-optimal dose d_C^P, P-efficiency of d_Π^P (optimal dose in the independence case) and P-efficiency of $d_{C_F}^P$ (optimal dose under a misspecified copula) under Scenario 2

Scenario 2						
True Copula (C_T)	θ_{C_T}	d_C^P	$\text{Eff}_{C_T}^P(d_\Pi^P)$	**False Copula** (C_F)	θ_{C_F}	$\text{Eff}_{C_T}^P(d_{C_F}^P)$
	2	0.3249	0.9708		2	0.8666
Clayton	8	-0.3180	0.5184	**Gumbel**	5	0.9049
	18	-0.3551	0.0859		10	0.9656
	2	0.0366	0.9179		2	0.7603
Gumbel	5	-0.1562	0.5796	**Clayton**	8	0.9225
	10	-0.2757	0.2355		18	0.9771
Independence copula Π		0.1993	1			

to ignore it leads to a wrong optimal dose d_π^P), the choice of the copula seems to be indifferent (orange and green lines in Fig. 4 are almost overlapping).

Finally consider Scenario 3 defined by $\delta = (1, 1.5, -3, 2.5, 5)$, where the marginal probability of toxicity is greater than that of efficacy, as shown in Fig. 5 where the area included between the red solid line and the black dotted one defines the region where p_{11}^C may assume values. Actually, we have losses in the P-efficiency of both d_Π^P and $d_{C_F}^P$, as shown in Table 9.

Table 9: P-optimal dose d_C^P, P-efficiency of d_Π^P (optimal dose in the independence case) and P-efficiency of $d_{C_F}^P$ (optimal dose under a misspecified copula) under Scenario 3

Scenario 3						
True Copula (C_T)	θ_{C_T}	d_C^P	$\text{Eff}_{C_T}^P(d_\Pi^P)$	**False Copula** (C_F)	θ_{C_F}	$\text{Eff}_{C_T}^P(d_{C_F}^P)$
	2	-0.3760	0.9215		2	0.7747
Clayton	8	-0.2234	0.5982	**Gumbel**	5	0.2308
	18	-0.0825	0.3282		10	0.0333
	2	-0.5551	0.9516		2	0.7601
Gumbel	5	-0.6229	0.7567	**Clayton**	8	0.0957
	10	-0.6606	0.4914		18	0.0002
Independence copula Π		-0.479	1			

From Fig. 6 we have that the probabilities of efficacy without toxicity, p_{10}^C under Clayton and Gumbel copulae, reach their maximum value at different doses (even if both of them are flat).

Therefore, for this kind of scenarios, it is necessary to correctly identify the true dependence copula model in order to assess the P-optimal dose.

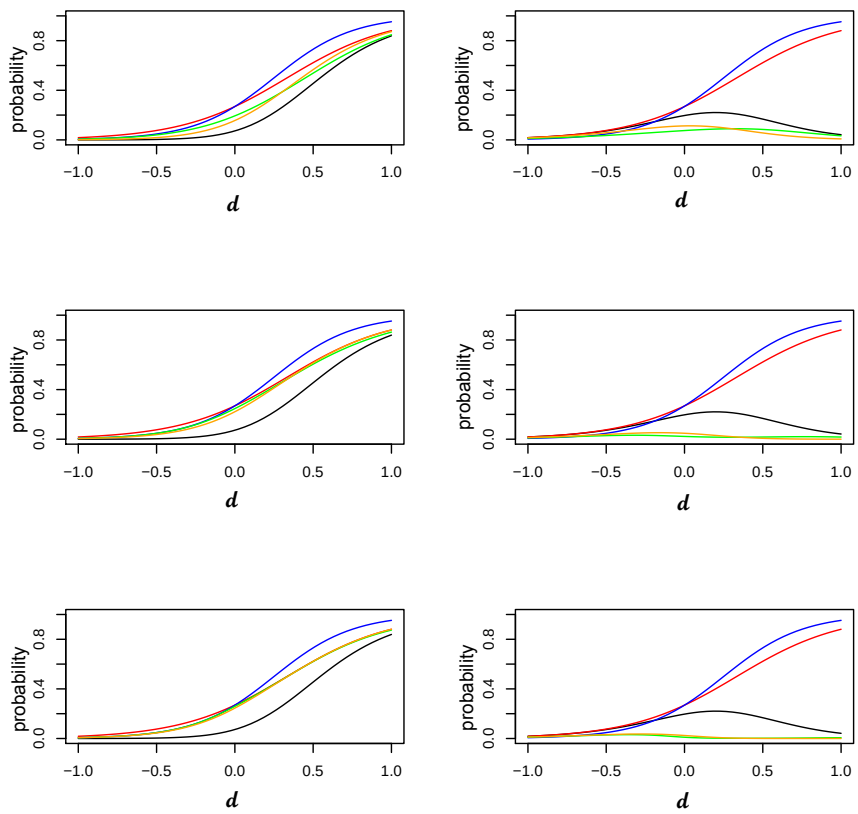

Fig. 4: Marginal probabilities of efficacy (red line) and toxicity (blue line); $p_{11}^{C}(d;\delta,\theta_C)$ (left-side) and $p_{10}^{C}(d;\delta,\theta_C)$ (right-side) for: the independence situation (black line), $C = Cl$ (green line) and $C = G$ (orange line), with three different values of τ (see Table 3): $\tau = 0.5$ (first row), $\tau = 0.8$ (second row) and $\tau = 0.9$ (last row), at $\delta = (-1, 3, 0, -1, 4)$ (Scenario 2).

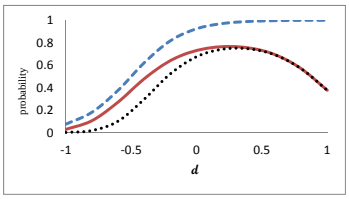

Fig. 5: Marginal probabilities of efficacy (red solid line) and toxicity (blue dashed line) and their joint probability in the independence case (black dotted line), for $\delta = (1, 1.5, -3, 2.5, 5)$ related to Scenario 3.

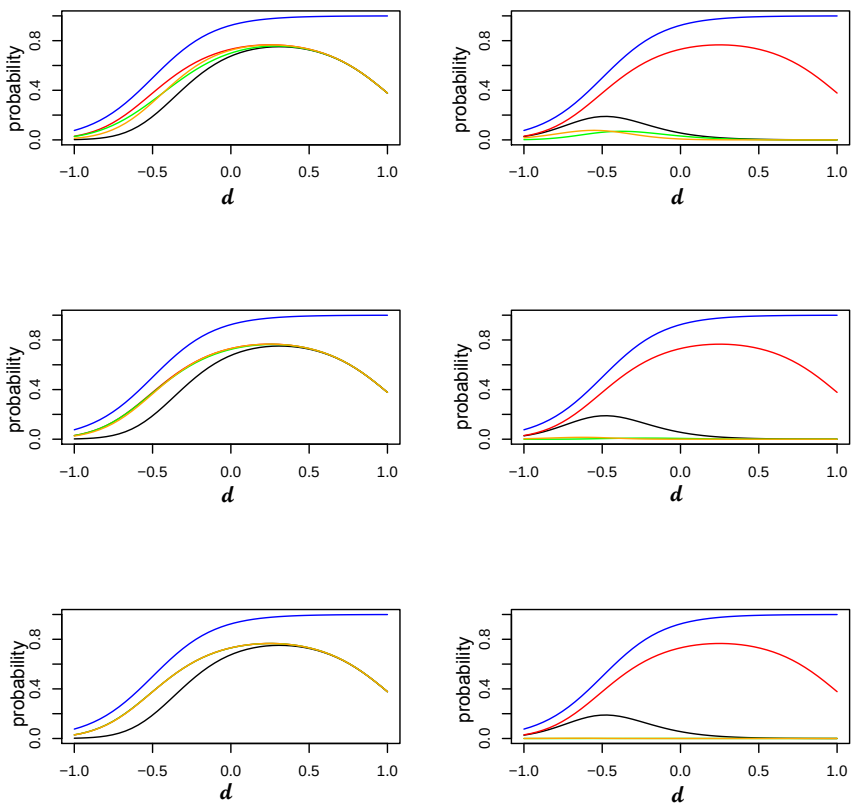

Fig. 6: Marginal probabilities of efficacy (red line) and toxicity (blue line); $p_{11}^{C}(d;\delta,\theta_C)$ (left-side) and $p_{10}^{C}(d;\delta,\theta_C)$ (right-side) for: the independence situation (black line), $C = Cl$ (green line) and $C = G$ (orange line), with three different values of τ (see Table 3): $\tau = 0.5$ (first row), $\tau = 0.8$ (second row) and $\tau = 0.9$ (last row), at $\delta = (1, 1.5, -3, 2.5, 5)$ (Scenario 3).