



HAL
open science

Design Optimization for dose-finding trials: A review

Jihane Aouni, Jean Noel Bacro, Gwladys Toulemonde, Pierre Colin, Loic Darchy, Bernard Sebastien

► **To cite this version:**

Jihane Aouni, Jean Noel Bacro, Gwladys Toulemonde, Pierre Colin, Loic Darchy, et al.. Design Optimization for dose-finding trials: A review. *Journal of Biopharmaceutical Statistics*, 2020, 30 (4), pp.662-673. 10.1080/10543406.2020.1730874 . hal-03082449

HAL Id: hal-03082449

<https://inria.hal.science/hal-03082449v1>

Submitted on 18 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Design Optimization for dose-finding trials: A review

Jihane AOUNI^{1,2}, jihane.aouni@sanofi.com

Jean Noel BACRO², jean-noel.bacro@umontpellier.fr

Gwladys TOULEMONDE^{2,3}, gwladys.toulemonde@umontpellier.fr

Pierre COLIN¹, pierre.colin@sanofi.com

Loic DARCHY¹, loic.darchy@sanofi.com

Bernard SEBASTIEN¹, bernard.sebastien@sanofi.com

¹Sanofi, Research and Development, 91385 Chilly-Mazarin, France

²IMAG, Univ Montpellier, CNRS, Montpellier, France

³Lemon, INRIA

December 27, 2019

Abstract

Dose selection is one of the most difficult and crucial decisions to make during drug development. As a consequence the dose-finding trial is a major milestone in the drug development plan and should be properly designed. This article will review the most recent methodologies for optimizing the design of dose-finding studies: all of them are based on the modeling of the dose-response curve, which is now the gold standard approach for analyzing dose-finding studies instead of the traditional ANOVA/multiple testing approach. We will address the optimization of both fixed and adaptive designs and briefly outline new methodologies currently under investigation, based on utility functions.

Keywords: Adaptive trials; Design optimization; Dose selection; Patient allocation; Utility functions

1 Introduction / Optimal design considerations

Selection of a dose (or doses) to carry into confirmatory phase III trials is one of the most difficult decisions that need to be made during drug development and it is generally agreed that the high attrition rate in phase III studies may, in part, be due to inadequate dose selection. Dose-finding studies therefore play a major role in drug development and should be carefully designed and conducted.

The scope of this review will be limited to the optimization of the design of parallel group phase IIB dose-finding studies, leaving dose escalation dose-finding studies out of the scope. We will also address the problem of the number of subjects per dose level (both in terms of total sample size and/or distribution of the subjects within treatment arms) in the dose-finding trial.

In addition, we will not consider the traditional ANOVA approach that uses contrast statistics to compare dose levels (e.g., differences between active doses and placebo); we will, instead, focus on the model-based approach, that is based on the estimation of the dose-response profile.

1.1 Dose-response modeling

The more recent approaches in dose selection have clearly moved the methodology from the multiple testing methods to dose-response "modeling" based methods such as the MCP-Mod method (Multiple Comparison Procedure and Modeling), see Bretz et al. (2005) and Aouni et al. (2019a). Modeling the dose-effect relationship then becomes a necessary, key task in drug development; for this purpose, various dose-response models, linear or nonlinear, have become standard and are well known: for instance, linear, logistic, E-max (simple, Sigmoid or standardized form), quadratic, exponential and linear in log-dose models, see Pinheiro et al. (2014), Pinheiro et al. (2006). These models are presented in Table 1. Throughout this publication, we will use d to refer to the dose under consideration and θ for the vector of parameters of the dose-response function $f(d, \theta)$ that represents the mean response for patients receiving dose d .

The linear model is the simplest one. Its simplicity is an advantage but it suffers from severe limitations (Vandenberg et al. (2012)). An example of a non-linear dose-response model is the E-max model: E_0 is the placebo effect, E_{max} is the asymptotic (associated to an "infinite dose") upper bound of effect, as compared to placebo, and ED_{50} is the dose giving half of E_{max} as change from placebo. A more complex form of this model is its Sigmoid version, where an additional parameter is included, ensuring greater shape flexibility, the "Hill" exponent (the slope), h , reflecting the shape of the dose-effect curve.

Other monotonic (as a function of dose) dose-response models can be considered: logistic, exponential and linear in log-dose models.

The quadratic model has the ability to describe a non-monotonic dose-response relationship, in a convex/U-shape (when $\beta_2 > 0$), or in a concave/umbrella-shape (when $\beta_2 < 0$).

Name	$f(d, \theta)$
Linear	$E_0 + \delta \times d$
E _{max}	$E_0 + E_{max} \times d / (ED_{50} + d)$
Sigmoid E _{max}	$E_0 + E_{max} \times d^h / (ED_{50}^h + d^h)$
Log linear	$E_0 + \delta \times \log(d + \text{offset})$
Exponential	$E_0 \times \exp(d/\delta)$
Quadratic	$E_0 + \beta_1 d + \beta_2 d^2$
Logistic	$E_0 + E_{max} / (1 + \exp[(ED_{50} - d)/\delta])$

Table 1: Examples of dose-response models.

1.2 Generalities on Optimal designs

In its widest sense, a design is defined by the doses considered in the trial, the total sample size and the relative allocation fraction to doses. Even if the number of doses and dose spacings are briefly discussed, we will essentially focus on optimization in the relative allocation fraction to doses so that, throughout this publication, design will be essentially defined by a vector $\xi = (w_1, \dots, w_K)^t$, representing the relative allocation fraction to doses.

Several optimality criteria can be defined leading to designs such as A, C, D, E, T and G-optimal designs studied by Rady et al. (2009), for instance. We will only briefly present D and C optimality criteria, which are, in practice, the most useful in the optimization of dose-finding study designs. It is important to note that the optimality criteria that we will discuss are 'local', in the sense that the designs are optimal for a given model, and for the assumed model parameters: some aspects related to model misspecification will be discussed in section 2.3.

D-optimal designs maximize the determinant of the Fisher information matrix related to the parameters defining the dose-response model. This corresponds to minimizing the volume of the confidence ellipsoid for the vector of parameters of the dose-response model, see Davison (2003) for instance. This criterion has the advantage of considering all model parameters, instead of focusing on one particular dose. This property can be described with the help of the equivalence theorem of Kiefer-Wolfowitz (Kiefer and Wolfowitz (1960)): an interpretation of this theorem states that, in the context of linear regression, a D-optimal design will minimize the maximum variability of the mean response estimation.

The objective of C-optimal design, in the context of dose-finding studies, is to minimize the variance of a specific function of the parameters, this function being most often defined by a target dose to estimate. As an example Bretz et al. (2010) define the minimum effective dose (MED), that is, the smallest dose producing a clinically relevant difference from placebo response, say Δ , and mention other doses of interest such those achieving 100p% of the maximum treatment effect in the observed dose range noted ED_p. In general those doses can easily be expressed as a function of the dose-response model parameters: for instance, in the context of an E_{max} model, one can easily see that

$$\text{MED}(\Delta) = \frac{\Delta \cdot ED_{50}}{E_{\max} - \Delta} \text{ (assuming that } \Delta < E_{\max} \text{) and } ED_p = ED_{50} \times \frac{p}{1-p}.$$

In practice, the optimization of the design is conducted using specific algorithms. We can mention the Fedorov-Wynn algorithm (Fedorov (1972), Wynn (1972)). Also the 'DoseFinding' R-package (Bornkamp et al. (2018)) provides a function to compute D-optimal designs.

1.3 Dose-finding based on a Pharmacometric approach

When it can be implemented, a pharmacometric approach based on exposure response modeling, instead of a dose-response modeling, can be very efficient. The exposure response modeling approach consists in relating, through a parametric model, the plasma concentration (or a summary of the kinetics of the plasma concentration such as the AUC or the Cmax) of the compound to a biomarker of interest or the clinical endpoint. This approach is described in detail by Holford (2006). Exposure response modeling can be more informative than dose-response modeling because individuals differ in their exposure even if they receive the same dose. Furthermore, when exposure response modeling is based on biomarker response, it can benefit from the mixed modeling framework for longitudinal analyses. Once estimated, exposure response models can be used to select the doses through simulations (see Holford (2006) for instance).

The exposure response approach is particularly useful when the recruitment of patients is challenging and the minimization of the study sample size is of major importance: this is the case for pediatric studies for instance.

In the context of exposure response analyses the same optimality criteria as those used for the dose-response models can be defined in the framework of mixed models. As an example, the search for D-optimal designs is implemented in PFIM software (Bazzoli et al. (2010)). Those optimal designs can be used to minimize the number of blood samples per patient while maintaining sufficient precision in the estimates of parameters of the pharmacokinetics models: this is particularly important for designing pediatric pharmacokinetics studies.

2 Design Optimization of fixed design

In this section we focus on randomized, double-blind, placebo-controlled, fixed dose, parallel group dose-finding trials. In particular we will not consider cross-over trials or dose titration designs as, despite their appealing statistical properties, they are more complicated to implement for phase II dose-finding trials with a long treatment duration (see Ting (2006), for a short discussion).

2.1 Choice of dose range

The first task is to identify the range of dose of interest, that we will note $[d_m, d^M]$: the dose will be selected within this dose range as well as the doses chosen for the dose-finding trial.

When $d_m = 0$, the dose-finding trial includes the placebo arm as control (it includes the case when all the doses, including the null dose, are administered on top of a background medication). When it is non ethical to include a placebo arm, then $d_m > 0$. In both situations there could be an active control arm in addition to the experimental drug dose arms.

The choice of d^M is a little bit less straightforward: the choice of the upper limit of the dose range is based on safety considerations and is related to the available information on the Maximal Tolerated Dose (MTD) as identified in phase I (Ting (2006)).

2.2 Number of doses, Dose spacings

The optimization of these two key components of trial designs are dependent on several aspects: the optimality criterion chosen, the primary objective of the trial (i.e. estimate the MED or EDp dose) and the true underlying dose-response.

Concerning the Number of doses and dose spacing, these two points are addressed by Bretz et al. (2010). Based on previous results obtained by Dette et al. (2008), where it is shown that the C-optimal (for MED and EDp definitions) design can be explicitly calculated for common dose-response models with two or three model parameters, the authors give explicit formula defining the optimal design (giving the required doses and the percentage of patients allocated to those doses), for the Emax model, considering the C-optimality criterion (both MED and EDp definitions) as well as the D-optimality criterion. Their main findings are as follows (we assume that $d_m = 0$): for the linear and log-linear models the optimal design requires two dose levels whereas it requires two or three dose levels for the Emax and exponential models.

In fact, the results provided by Bretz et al. (2010) or Dette et al. (2008) potentially identify the optimal design for the common dose-response models, but in practice, a direct application of these results for designing a dose-finding study can be difficult: a potential problem is that the optimal design only requires one intermediate dose between the placebo and the maximal dose (this can be problematic as it would then be impossible to explore alternative dose response relationships). To address this issue, one could envisage a more robust approach: optimize the design with respect to each model and keep the set of doses selected for the various models; again such an approach could not be really applicable as some of the intermediate doses identified can be 'odd' and not really feasible (this is also discussed in Bretz et al. (2010)); however we will address below, in more detail, robustness and model misspecification issues. As a consequence a more pragmatic and realistic approach consists in choosing a priori the doses to be tested, $d_0 (= 0)$, $d_1, d_2, \dots, d_K (= d^M)$, and optimizing the designs in patients allocation only, through weights $w_0, w_1, w_2, \dots, w_K$. Given the maximal dose, generally the MTD, the doses could be chosen based on pharmacokinetics in minimizing the overlap in exposure between consecutive doses. Another simple approach, although justified by none of the optimality

criteria, could be to choose equal spacing between consecutive doses. Other rules are proposed and discussed by Ting (2006).

2.3 Patient allocation

As mentioned above, we will now focus on the problem of optimizing patient allocation to the a priori chosen doses of the design: $0, d_1, \dots, d_K$. Mathematically, this optimization will consist in maximizing an objective function defined over a K -dimensional simplex defined as the set of vectors, $\xi = (w_k)_{k=0, \dots, K}$ with positive components such that $\sum_k w_k = 1$. The various methods that will be discussed here have potentially the ability to reduce the sample size of the dose-finding study by 40% or 45% (Bretz et al. (2010)) assuming that the true dose-response model is known and the true model parameter was used to perform the design optimization computation.

Optimality Criteria

For all optimality criteria that will be defined below, optimization of the design involves an objective function, $\Phi(\xi)$, representing the efficiency of the design, that depends on the total information matrix:

$$M_\theta(\xi) = \sum_{k=0}^K w_k \mathcal{I}(d_k; \theta)$$

where $\mathcal{I}(d_k, \theta)$ is the Fisher information matrix for one subject receiving dose d_k and with θ as the dose-response model parameter value.

For the D-optimality criterion the optimal design maximizes the design optimality criterion:

$$\Phi_D(\xi) = \log \det(M_\theta(\xi)).$$

As mentioned above, choosing the D-optimality criterion, will result in maximizing the overall precision of the parameter estimates, as assessed by the volume of the confidence ellipsoid for the model parameters.

If we are more interested in obtaining precise estimates of some target doses, such as the MED or EDp defined above, then the C-optimality criterion must be chosen. This can be formalized as follows: the target dose is a function, $\psi(\theta)$, of the model parameters (explicit definitions have been given above for the MED and EDp for the Emax model), therefore the optimal allocation is defined by the design vector, ξ , that minimizes the variance $var(\psi(\hat{\theta}))$. Based on the delta-method (Davison (2003)) the objective is to minimize

$$var(\psi(\hat{\theta})) \simeq \frac{\partial \psi^t}{\partial \theta} var(\hat{\theta}) \frac{\partial \psi}{\partial \theta}$$

which is approximately equivalent to minimizing

$$\frac{\partial \psi^t}{\partial \theta} (N \cdot \sum_k w_k \mathcal{I}(d_k; \theta))^{-1} \frac{\partial \psi}{\partial \theta} = \frac{\partial \psi^t}{\partial \theta} (N \cdot M_\theta(\xi))^{-1} \frac{\partial \psi}{\partial \theta}$$

where N is the total sample size.

Therefore the application of C-optimality criterion consists in maximizing, in ξ , the design optimality criterion (ignoring N that has no influence on the optimization):

$$\Phi_C(\xi) = \left(\frac{\partial \psi^t}{\partial \theta} M_\theta^{-1}(\xi) \frac{\partial \psi}{\partial \theta} \right)^{-1}.$$

Another example of C-optimality is used by Miller et al. (2007). Here the aim is to minimize the variance $var(\psi(\hat{\theta}, d))$ where ψ is a function of the parameter that also depends on the dose: $\psi(\theta, d) = f(d, \theta) - f(0, \theta)$, the mean difference versus placebo of the dose d . As defined above, the optimality criterion is then:

$$\Phi(\xi, d) = \left(\frac{\partial \psi(\theta, d)^t}{\partial \theta} M_\theta^{-1}(\xi) \frac{\partial \psi(\theta, d)}{\partial \theta} \right)^{-1}.$$

This optimality criterion can have a particular interest in the case when the dose-finding study is also the Proof of Concept (PoC) study. In this case, one of the primary objectives would be to characterize the effect of the maximal dose (d^M) and decide whether its effect is clinically relevant (subsequent analyses/studies would then be conducted) or whether the drug development must be stopped. In Miller et al. (2007) the authors propose to primarily consider this maximal dose and define

$$\Phi_1(\xi) = \Phi(\xi; d^M)$$

as optimality criterion of a design. This criterion has an appealing interpretation: if the criterion value of a particular design is increased by $x\%$ as compared to the balanced design, then the balanced design requires an increase in sample size of $x\%$ to provide the same precision in estimates of interest as the design under consideration.

In many situations, focusing on the highest dose only is not appropriate. In this situation, a criterion quantifying the quality of a design should consider a dose interval $[d_L, d_U]$: this interval could simply be the whole dose range $[0, d^M]$ or a sub-interval, depending on the context and the study specific objectives. In this case, the authors propose to define

$$\Phi_2(\xi) = \left(\int_{d_L}^{d_U} \Phi(\xi, x)^{-1} dx \right)^{-1}$$

and use it as optimality criterion of a design.

Robustness, Handling of Uncertainty

The various methods of design optimization described above can lead to substantial reduction in total sample size (as compared to the balanced design for instance) but the calculations depend on the knowledge of the true dose-response

function but also on the true values of the model parameters. Even though some knowledge is available at the beginning of phase II, the risk of misspecification in the dose-response models and/or the model parameter values cannot be neglected.

The problem of model misspecification is addressed by Bretz et al. (2010). This is of great importance as, according to Dette et al. (2008), optimal design specifications are extremely sensitive to model misspecification (whereas they show some robustness against parameter value misspecification). Indeed, authors report potentially severe loss of efficiency of the nominal optimal design in case of model misspecification. As an example, if the local MED-optimal design for the log-linear model is used for the logistic model, then its asymptotic variance is approximately 100 times larger than the value that would be obtained if the local MED-optimal design for logistic model was used, see Table 7 of Dette et al. (2008).

Bretz et al. (2010) propose several methods, applicable to any optimality criterion: D-optimality or C-optimality related to either MED or EDp. All of these methods require, as a first step, identifying or proposing m candidate models f_1, \dots, f_m , with their corresponding parameter values $\theta_1, \dots, \theta_m$, and then for each of them, identifying their optimal design ξ_j^* .

The first approach is a maximin approach: for each candidate design ξ its relative efficiency with respect to the optimal design of the models is computed:

$$eff_j(\xi) = \mathcal{C}_j(\xi) / \mathcal{C}_j(\xi_j^*)$$

where \mathcal{C}_j is the optimality criterion adapted to model j . Then the worst value,

$$\min\{eff_1(\xi), \dots, eff_m(\xi)\},$$

is computed. And the robust optimal design is the design

$$\xi^* = \arg \max_{\xi} \min\{eff_1(\xi), \dots, eff_m(\xi)\}$$

that maximizes this worst value. This optimization can be unconstrained but maybe a more pragmatic approach would be to conduct the optimization within a set of pre-defined easily feasible (from a logistic point of view) designs.

The second approach proposed by Bretz et al. (2010) has more a Bayesian or model averaging spirit. It consists in assigning model weights (α_j) to each model j , reflecting the belief in each of the models (but the natural choice $\alpha_j = 1/m$, $\forall j$, can be a pragmatic option). And then the robust optimal design is the one that maximizes the weighted

sum

$$\sum_{j=1}^m \alpha_j \log(\text{eff}_j(\xi)).$$

The problem of uncertainty in parameter values is addressed in Miller et al. (2007) using a Bayesian approach. The authors propose a priori typical dose-response profiles (called 'scenario') and assign parameter values representative of (or consistent with) these scenarios, and they also assign prior probabilities α_j to each of these scenarios, representing beliefs in these scenarios. And then using any of the design optimality criteria Φ , as defined above, the method consists in maximizing

$$\sum_{j=1}^m \alpha_j \Phi(\xi, \theta_j) / \Phi(\xi_B, \theta_j)$$

where ξ_B is the standard balanced design. This approach is more intuitive and less computationally demanding than specifying a full prior distribution π for the parameter values and computing the Bayesian average:

$$\int \Phi(\xi, \theta) / \Phi(\xi_B, \theta) d\pi(\theta).$$

As a conclusion, we think that in case of much uncertainty in the dose response relationship the primary goal is to consider designs robust against model misspecification. For this purpose both the maximin or weighted sum of the log of relative efficiencies of Bretz et al. (2010) could be used; the maximin approach is maybe more conservative (as it protects against a worst case scenario) and easier to explain to a clinical team whereas the weighted sum approach is more flexible as it allows, in case of an interim analysis, to update the weights (using a Bayesian approach for instance) and then optimize the design of the subsequent patient cohorts based on the interim analysis results.

2.4 Approach based on utility functions

In the methods described above, the objective of the dose-finding trials were not really to identify an "optimal" dose (in a sense that would need to be clarified) but to identify the lower dose that achieves some targeted level of efficacy: this is the case for the MED or EDp doses discussed in the preceding sections. In order to introduce notions of optimality for the doses, it is necessary to take into account safety considerations. A relevant way to do so is to consider utility functions and to assign utility values to the doses.

Utility functions are generally introduced and defined within the decision theory framework. Utility describes the preferences of the decision maker and classifies/orders decisions. A decision theory result suggests that, in an uncertain (random) environment, all decision rules can be compared and classified using the expectation of a certain function, called the utility function. Various levels of complexity can be introduced within the framework of utility functions: economical/financial aspects could be accounted for in addition to efficacy and safety considerations.

Utility functions of various kinds (including a safety / penalty term penalizing high doses, rewards, global costs

and/or costs per patient) have already been proposed and discussed in the context of drug development: we briefly summarize below the various approaches in the literature.

In Patel et al. (2013), utility functions are defined at the portfolio level: this work is beyond the scope of our review as the aim is to optimize the sample sizes allocated to the studies to the various drugs in a given portfolio in order to maximize the expected net present value. Other authors aim to optimize a given drug development plan using utility functions, sometimes defined as the expected net present value (as in Antonijevic et al. (2013), Kirchner et al. (2016), Patel et al. (2012)). In these works, the development plans are optimized with respect to phase II and phase III sample sizes, as well as Go/No-Go Decisions. In Foo and Duffull (2017), the aim is to optimize (in terms of number of patients, number of blood samples per patient and corresponding sampling schedule) an early phase pharmacokinetics study with the objective of minimizing study cost (study failure has a specific cost that is taken into account in the optimization). For early phase studies, we can mention Hee et al. (2016), where decisions theoretic based designs are reviewed.

The approach in Dragalin and Fedorov (2006) considers efficacy and safety simultaneously, described by two correlated binary variables. Defining a design as a probability measure, ξ , defined on a set of dose X , the authors define a penalty function

$$\varphi(x, \theta, C_E, C_T) = p_{10}(x, \theta)^{-C_E} (1 - p_{.1}(x, \theta))^{-C_T}$$

where $p_{10}(x, \theta)$ is the probability of being a responder for efficacy without adverse reactions at dose x (and for model parameter value θ), $p_{.1}(x, \theta)$ is the marginal probability of having an adverse reaction (whatever efficacy) at the dose x , and the C_E, C_T are positive control parameters to be chosen. This penalty function φ can be interpreted as the inverse of a utility function: it decreases as the efficacy increases and it increases as the safety deteriorates. The authors define an integrated penalty

$$\int \varphi(x, \theta, C_E, C_T) \xi(dx),$$

and then a penalized information criterion:

$$\frac{\int \mathcal{I}(x, \theta) \xi(dx)}{\int \varphi(x, \theta, C_E, C_T) \xi(dx)}$$

where $\mathcal{I}(x, \theta)$ is the information matrix associated to the two, efficacy and safety, binary variables. A D-optimal design is a design that would minimize

$$\det \left(\frac{\int \mathcal{I}(x, \theta) \xi(dx)}{\int \varphi(x, \theta, C_E, C_T) \xi(dx)} \right)^{-1} = \left(\int \varphi(x, \theta, C_E, C_T) \xi(dx) \right)^M \det \left(\int \mathcal{I}(x, \theta) \xi(dx)^{-1} \right)$$

where M is the number of parameters for the models. Because of the penalty term, the optimal design will put more weight on doses for which the penalty (combining both safety and efficacy) is not too large.

A related approach, but with a slightly different objective, can be described as follows. Starting from a utility function,

which can be similar to the inverse penalty in Dragalin and Fedorov (2006):

$$U(d; \theta) = U_E(d; \theta) \times U_T(d; \theta),$$

where the U_E , U_T are, respectively, efficacy and safety related utilities (U_E increases when the dose increases while U_T decreases when the dose increases). Then, instead of searching for an optimal design that would give more weight to doses with high utility values (like in Dragalin and Fedorov (2006)), one could search for designs that provide the most accurate estimates of the utility values for some selected set or range of doses. This would then lead to the performance of design optimization using a similar approach to the computation of C-optimal designs. Assuming the utility function is sufficiently smooth and concave, the optimal dose d^* is defined by

$$\frac{\partial U(d^*; \theta)}{\partial d} = 0.$$

Then with a maximum-likelihood estimate of θ , $\hat{\theta}$, the optimal dose is estimated by

$$\frac{\partial U(\hat{d}^*, \hat{\theta})}{\partial d} = 0.$$

By application of standard differential calculus results, we conclude that \hat{d} is a smooth function of θ , and that

$$\frac{\partial \hat{d}^*}{\partial \theta} = - \left(\frac{\partial^2 U}{\partial d^2} \right)^{-1} \frac{\partial^2 U}{\partial d \partial \theta}.$$

Using this relationship, we can then apply the methods for the determination of C-optimal designs with function $d^*(\theta)$ as a function determining the target dose (function $\psi(\cdot)$ of section 2.3). The application of this methodology is currently being assessed by the main author as part of her research work. But, to be applicable, such an approach relies on several assumptions: the parametric dose-response model should be perfectly specified and asymptotic normality should be achieved, whereas it is well documented that convergence issues might arise for the most complex models, like the logistic or sigmoidal Emax models, for instance.

3 Adaptive Design Optimization

In this section we will focus on the problem of conducting one or several interim analyses with the specific objective of improving the design of the dose-finding study and making better decisions at the end (choice of dose for phase III). In this framework, the interim analysis results in updating patient allocation to the active doses and placebo.

3.1 Two-stage adaptive designs

We first focus on the simplest case where there is only one interim analysis: the first cohort of patients is enrolled according to a standard design (a balanced design for example) and then, based on the interim results, the patient al-

location for the second part is optimized using one of the methodologies mentioned above. In practice, the following methods can be used:

- (i) consider only one single dose-response model (Emax for instance), use the interim data to estimate the parameters and optimize the second part of the trial accordingly.
- (ii) consider some uncertainty in model/parameter values: consider various dose-response models, estimate each model individually and then apply the maximin or the weighted efficiency approach of Bretz et al. (2010) or use the optimal Bayesian design approach of Miller et al. (2007) in updating the the prior beliefs for each scenario. This latter approach is detailed in Bornkamp et al. (2011).

When the data accumulated during the first part of the trial are included in the final analysis also, as in Miller et al. (2007), there are other constraints to consider when optimizing the weights in the first stage. Indeed, in this case for each dose arm j its weight for final analysis, w_j , is necessarily $\geq w_{j,0} = \frac{n_{j,interim}}{N_{total}}$, where $n_{j,interim}$ is the number of patients in arm j at interim analysis and N_{total} is the total sample size of the trial at the final analysis.

Therefore, when optimizing the efficiency of the design, the w_j weights are constrained to be $\geq w_{j,0}$ and not only to be ≥ 0 . Once the final weights are obtained, the number of patients to be enrolled for the second part of the trial is $n_{j,(2)} = N_{total} \times w_j - n_{j,interim}$.

Other operational constraints can be handled quite easily. In case, for instance, there is a need to have a pre-determined sample size or allocation rate in the placebo group, say w_0 , then the optimization would be conducted in the other allocation rates, with a new constraint on the sum of weights: $\sum_{k=1}^K w_k = 1 - w_0$.

According to Miller et al. (2007), the practical interest of performing an interim analysis with the only goal of optimizing the trial is questionable.

The first issue is related to the final statistical analysis: it is not fully clear to what extent statistical inference is disturbed by the adaptation following the interim analysis. This point is addressed using simulations by Miller et al. (2007): their results suggest a slight modification of the statistical properties of the estimates by the interim analysis.

In addition, the same authors conclude that there is no real evidence that design optimization following the interim analysis significantly improves the efficiency of the design. This lack of improvement brought by the interim design optimization is also noticed by MacCallum and Bornkamp (2015).

However, such a statement must be mitigated:

- (i) it is important to bear in mind that design optimization is not the sole objective of interim analyses. There are in fact many reasons that interim analyses are worthwhile: notably ethical reasons (ineffective or unsafe doses are dropped), or, for example, the need to bring information to the project team or management.
- (ii) there is evidence (see section 3.2) that multiple interim analyses improve the performance of the design.
- (iii) in Miller et al. (2007) the interim analysis is conducted when the third of the patients are enrolled: the results might have been different if the interim analysis had been conducted when half of the patients were enrolled for instance; in

fact, the timing of the interim analysis is a key parameter that drives the performance of the adaptive design, and it must be properly tuned and even optimized. Actually, the choice of the best timing should be made using simulations: the relative efficiency, with respect to a fixed design, of the adaptive design would be assessed for adaptation occurring at various timing such as, for instance, when 20%, 30%, 40%, 50%, 60%, 70% of the patients are enrolled, and the timing associated with the highest relative efficiency should be chosen.

In Antonijevic et al. (2010), using the utility function approach, the authors compare various scenarios defined by the number of doses in the phase II studies, the type of design (fixed or Bayesian adaptive dose allocation design) with the objective of optimizing the probability of success in phase III and the expected net value: the authors point out the importance of allocating a sufficient number of patients in phase II and conclude that the Bayesian adaptive designs generally outperform fixed dose allocation designs.

3.2 Designs with multiple interim analyses

The principle is similar to the case of a single adaptation: at each of the interim analyses, the vector of relative allocation fraction to doses is updated based on new parameter estimates or belief updates in the dose-response models for the Bayesian design case. Because of the operational complexities induced by the interim analyses, such adaptive designs are more appropriate when the interval between patient enrollment and assessment and analysis of the efficacy criterion is very short. These types of adaptive designs are assessed through simulations in Dragalin et al. (2010).

The authors consider various types of adaptive Bayesian or non-Bayesian designs, as well as various optimality criteria, and assess the efficiency of these designs with 0 (corresponding to a fixed design), 1, 2, 4 and 9 equally spaced interim analyses. Their findings confirm the improvement brought by design adaptation, especially when both the number of doses and the dose range are large. In particular, increasing the number of adaptations leads to a gain in precision when selecting target dose(s).

4 Conclusion

The randomized, double-blind, placebo controlled dose-finding study is a key milestone in drug development. Its aim is not only to provide strong evidence of drug effect, but also to give guidance for choosing the most relevant dose for the later phase of a drug development. Therefore it is of primary interest to optimize this trial and maximize the likelihood of success of the trial.

However, the optimal designs obtained by computation can be difficult to implement in practice. In addition, the dose-finding study does not only consist in the characterization of the primary endpoint dose-response relationship: another important objective is to provide some information on the safety profile for all doses used in the dose-finding

study. In this case, design optimization should take into account both efficacy and safety endpoints, like in Antonijevic et al. (2010). Current research conducted by the main author is focussed on this topic within the framework of utility functions (Aouni et al. (2019b)).

Acknowledgement *We would like to thank all the contributors and reviewers for their helpful comments and suggestions that led to an improved paper.*

Conflict of Interest

The authors have declared no conflict of interest.

References

- Antonijevic, Z., Manner, D., Pinheiro, J., Kimber, M., Burman, CF., and Bergenheim, K. (2013). Optimizing Drug Development Programs: Type 2 Diabetes Case Study. *Therapeutic Innovation Regulatory Science*, 47(3):363–374.
- Antonijevic, Z., Pinheiro, J., Fardipour, P., and Lewis, RJ. (2010). Impact of Dose Selection Strategies Used in Phase II on the Probability of Success in Phase III. *Statistics in Biopharmaceutical Research*, 2(4):469–486.
- Aouni, J., Bacro, JN., Toulemonde, G., Colin, P., Darchy, L., and Sebastien, B. (2019a). Assessing Dunnett and MCP-Mod based approaches in two-stage dose-finding trials. *Biostatistics and Health Sciences*, DOI: 10.21494/ISTE.OP.2019.0397, <https://www.openscience.fr/Evaluation-des-approches-basees-sur-le-test-de-Dunnett-et-MCP-Mod-dans-des>.
- Aouni, J., Bacro, JN., Toulemonde, G., and Sebastien, B. (2019b). Utility-Based Dose-Finding in Practice: Some Empirical Contributions and Recommendations. *Annals of Biostatistics & Biometric Applications*, DOI: 10.33552/ABBA.2019.03.000552, <https://irispublishers.com/abba/pdf/ABBA.MS.ID.000552.pdf>.
- Bazzoli, C., Retout, S., and Mentre, F. (2010). Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Computer Methods and Programs in Biomedicine*, 98:55–65.
- Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011). Response-adaptive dose-finding under model uncertainty. *The Annals of Applied Statistics*, 5(2B):1611–1631.
- Bornkamp, B., Pinheiro, J., and Bretz, F. (2018). *DoseFinding: Planning and Analyzing Dose Finding Experiments*. R package version 0.9-16.
- Bretz, F., Branson, M., and Pinheiro, J. (2005). Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies. *Biometrics*, 61(3):738–748.
- Bretz, F., Dette, H., and Pinheiro, J. (2010). Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine*, 29(7-8):731–742.
- Davison, AC. (2003). *Statistical Models*. Cambridge University Press.

- Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, 103(483):1225–1237.
- Dragalin, V., Bornkamp, B., Bretz, F., Miller, F., Padmanabhan, SK., Patel, N., Perevozskaya, I., Pinheiro, J., and JR, S. (2010). A Simulation Study to Compare New Adaptive Dose-Ranging Designs. *Statistics in Biopharmaceutical Research*, 2(4):487–512.
- Dragalin, V. and Fedorov, V. (2006). Adaptive designs for dose-finding based on efficacy–toxicity response. *Journal of Statistical Planning and Inference*, (136):1800–1823.
- Fedorov, VV. (1972). *Theory of optimal experiments*. Elsevier.
- Foo, LK. and Duffull, S. (2017). Designs to balance cost and success rate for an early phase clinical study. *Journal of Biopharmaceutical Statistics*, 27(1):148–158.
- Hee, SW., Hamborg, T., Day, S., Madan, J., Miller, F., Posch, M., Zohar, S., and Stallard, N. (2016). Decision-theoretic designs for small trials and pilot studies: a review. *Statistical methods in medical research*, 25(3):1022–1038.
- Holford, N. (2006). Dose Response: Pharmacokinetic–Pharmacodynamic Approach. In Ting, N., editor, *Dose Finding in Drug Development, Statistics for Biology and Health*, pages 73–88. Springer.
- Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 14:363–366.
- Kirchner, M., Kieser, M., Götte, H., and Schüler, A. (2016). Utility-based optimization of phase II/III programs. *Statistics in Medicine*, 35(2):305–316.
- MacCallum, E. and Bornkamp, B. (2015). Accounting for parameter uncertainty in two-stage designs for Phase II dose-response studies. *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, pages 427–450.
- Miller, F., Guilbaud, O., and Dette, H. (2007). Optimal Designs for Estimating the Interesting Part of a Dose-Effect Curve. *Journal of Biopharmaceutical Statistics*, 17(6):1097–1115.
- Patel, N., Bolognese, J., Chuang-Stein, C., Hewitt, D., Gammaitoni, A., and Pinheiro, J. (2012). Designing phase II trials based on program-level considerations: a case study for neuropathic pain. *Drug Information Journal*, 46(4):439–454.
- Patel, NR., Ankolekar, S., Antonijevic, Z., and Rajcic, N. (2013). A mathematical model for maximizing the value of phase III drug development portfolios incorporating budget constraints and risk. *Statistics in Medicine*, 32(10):1763–1777.
- Pinheiro, J., Bornkamp, B., Glimm, E., and Bretz, F. (2014). Model-based dose finding under model uncertainty using general parametric models. *Statistics in Medicine*, 33(10):1646–1661.
- Pinheiro, JC., Bretz, F., and Branson, M. (2006). Analysis of Dose-Response Studies—Modeling Approaches. In Ting, N., editor, *Dose Finding in Drug Development, Statistics for Biology and Health*, pages 146–171. Springer.
- Rady, EA., Abd El-Monsef, MME., and Seyam, MM. (2009). Relationships among several optimality criteria. *Interstat*,

15(6):1–11.

Ting, N. (2006). General Considerations in Dose–Response Study Designs. In Ting, N., editor, *Dose Finding in Drug Development, Statistics for Biology and Health*, pages 89–105. Springer.

Vandenberg, LN., Colborn, T., Hayes, TB., Heindel, JJ., Jacobs, DR., Lee, DH., Shioda, T., Soto, AM., Saal, FS., Welshons, WV., Zoeller, RT., and Myers, JP. (2012). Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. *Endocrine Reviews*, 33(3):378–455.

Wynn, HP. (1972). Results in the construction of D-optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, 34:133–147.