The design heatmap: A simple visualization of $D$-optimality design problems

Tim Holland-Letz | Annette Kopp-Schneider

German Cancer Research Center, Division of Biostatistics, Heidelberg, Germany

Correspondence: Tim Holland-Letz, Division of Biostatistics, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Email: t.holland-letz@dkfz.de

This article has earned an open data badge "Reproducible Research" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract
Optimal experimental designs are often formal and specific, and not intuitively plausible to practical experimenters. However, even in theory, there often are many different possible design points providing identical or nearly identical information compared to the design points of a strictly optimal design. In practical applications, this can be used to find designs that are a compromise between mathematical optimality and practical requirements, including preferences of experimenters. For this purpose, we propose a derivative-based two-dimensional graphical representation of the design space that, given any optimal design is already known, will show which areas of the design space are relevant for good designs and how these areas relate to each other. While existing equivalence theorems already allow such an illustration in regard to the relevance of design points only, our approach also shows whether different design points contribute the same kind of information, and thus allows tweaking of designs for practical applications, especially in regard to the splitting and combining of design points. We demonstrate the approach on a toxicological trial where a $D$-optimal design for a dose–response experiment modeled by a four-parameter log-logistic function was requested. As these designs require a prior estimate of the relevant parameters, which is difficult to obtain in a practical situation, we also discuss an adaption of our representations to the criterion of Bayesian $D$-optimality. While we focus on $D$-optimality, the approach is in principle applicable to different optimality criteria as well. However, much of the computational and graphical simplicity will be lost.

KEYWORDS
design heatmaps, $D$-optimal design, dose–response studies, visualization, optimal design algorithms

1 INTRODUCTION

The practical motivation of our research is biomedical trials where dose–response functions are fitted to data sets obtained, for example, from cell culture experiments. Usually, nonlinear four parameter functions such as log-logistic or Weibull...
functions are used (see Ritz, 2010, and, e.g., Clothier, Gomez-Lechon, Honegger, Kinsner-Ovaskainen, & Kopp-Schneider, 2013), and the design problem is how to select the dose levels to be used for each of the experimental units. Here, $D$-optimal designs generally propose using only four different dose levels, while guidelines and practical rules of thumb more often recommend a larger number of equally spaced support points. In part, these less focused recommendations are justified by the dependence of optimal designs on a priori estimate of the underlying true parameter values, which can be quite imprecise in practice. However, it is possible to include this uncertainty in design theory as well by using an a priori distribution of parameter assumptions in a Bayes-like approach. Still, in practical situations, a compromise between the theoretical and the applied viewpoints must be found. Our work in this paper is thus motivated by the somewhat more general challenge of (i) communicating optimal design considerations to the applied scientists performing the actual experiments and (ii) adapting the designs to practical requirements.

While theoretical optimal designs for a fixed parameter assumption generally require a small number of support points, there are often additional or alternative design points that could be used instead with no or minimal loss in design efficiency. Using this knowledge, the theoretical optimal designs could be modified in order to obtain designs that form a compromise between statistical optimality and the preferences and requirements of the experimentalists. To support this, we propose two variations of a simple, two-dimensional visualization of the design space we call the design heatmap, which will illustrate the relative contributions of the available design points as well as the similarity of the information gained from different points. For our main proposal, the design heatmap is constructed by plotting the scalar products between the information matrices provided by pairs of design points, which, for $D$-optimality, we can show to be an approximation for the change in efficiency that will result from an exchange of these two design points in the optimal design.

Note that our approach is only indirectly related to Elfving-type representations (Dette, 1993; Elfving, 1952) and has a more limited interpretation. However, it does allow a low (i.e., two-) dimensional and computationally simple representation that enables fine tuning of designs.

The heatmap is not supposed to be a tool to actually determine optimal designs from scratch. Instead, it takes an existing optimal design obtained from any source (algorithmical or analytical) and illustrates the contributions and interrelations between all available design points compared to the optimal design. In practice, this usually means that, as the first step, a $D$-optimal design will have to be constructed, for example, by using an established algorithmic procedure (e.g., Yu, 2010a, or Fedorov, 1972). While the resulting design will not be unique, it will provide the $D$-optimal, unique information matrix. As the next step, this optimal information matrix allows computation of the design heatmap, which then in turn can be used to decide which design points can be joined, split, or moved in order to match practical preferences. Furthermore, the heatmap also serves as an accessible representation of the design problem in general, which can be shown to the experimentalists and incorporated into software applications such as R-shiny apps.

Note that while one of the immediate applications of the design heatmap is to extend the baseline versions of the common design algorithms to find optimal designs with smaller numbers of support points, more efficient solutions for this have already been proposed elsewhere (Yu, 2011, or Yang, Biedermann, & Tang, 2013, among others). We will therefore focus mainly on the illustrative aspects.

The paper is composed of three main sections. Section 2 introduces the optimal design background and the established algorithms, while Section 3 introduces the design heatmap, the underlying theory, and an application both to the example of a log-logistic dose–response model and to a standard linear model.

## 2 OPTIMAL DESIGN PROBLEMS AND DESIGN ALGORITHMS

### 2.1 Motivating example

Our main example is taken from biomedical dose–response trials, where the problem was parameter estimation for a 4-parameter log-logistic function (see Figure 5a) given by

$$\eta(x, \theta_1, \theta_2, \theta_3, \theta_4) = \theta_1 + \frac{(\theta_2 - \theta_1)}{1 + \exp(\theta_3(\log(x) - \log(\theta_4)))}.$$ 

Here, $\theta_1$ and $\theta_2$ represent the range of possible effects, $\theta_3$ defines the slope of the curve, and $\theta_4$ identifies the dose of 50% effect (the $ED_{50}$ dose). Assuming log-doses $\log(x)$ can be chosen between $-10$ and $10$ in log-dose steps of $0.1$ and the true parameter values are $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (0, 1, 1, 1)$, applying standard design algorithms (see Section 2.5) to find a
selection of log-dose levels \( \log(x) \) that is \( D \)-optimal for the estimation of \( \theta \) results in the following design:

\[
\begin{array}{ccccccccccc}
\text{Log Dose} & -10.0 & -9.9 & -9.8 & -9.7 & -1.1 & -1.0 & 1 & 1.1 & 9.7 & 9.8 & 9.9 & 10.0 \\
\text{Weights} & 0.13 & 0.07 & 0.03 & 0.01 & 0.10 & 0.15 & 0.15 & 0.10 & 0.01 & 0.03 & 0.07 & 0.13 \\
\end{array}
\] (1)

The design proposes using 12 different log dose levels between \(-10\) and 10 in the experiment and further proposes assigning between 1\% and 15\% of the available observations to these dose levels. Designs for different parameter assumptions will look very similar in structure (Holland-Letz & Kopp-Schneider, 2014).

We observe that the algorithm proposes using a moderately large number of different dose levels some of them at very small weights. This is generally not a very useful design for practical applications, but some slight adaptations to the algorithms (e.g., Yu, 2011, or Yang et al., 2013, for a slightly different approach) could be used to obtain a design with a smaller number of support points. From general design theory (Pukelsheim, 2006), it is known that, at least on a continuous design space, a four support point \( D \)-optimal design exists, and therefore a comparably efficient simplification with fewer support points should be possible even on a discrete design space. However, the design as it is shown above already provides the unique optimal information matrix and also serves well to illustrate that there might be many different possible designs of optimal or near optimal performance. Thus, it will be used as our main example. First, however, the necessary (well-known) theoretical background and notation must be introduced.

### 2.2 Model

We consider a nonlinear regression model, where \( n \) observations can be taken, and a vector of parameters \( \theta \) is to be estimated:

\[
Y_j = \eta(x_j, \theta) + \epsilon_j \quad j = 1, ..., n.
\] (2)

Observations can be taken at measurement conditions \( x_j, j = 1, ..., n \), selected from a finite set \( \mathcal{X} \) consisting of \( r \) different discrete measurement settings. While we primarily consider one-dimensional measurement conditions \( x_j \), these can also be multidimensional as long as \( \mathcal{X} \) remains a finite set. Any specific choice of measurement settings for all subjects is called the design \( \xi \) of the experiment. The parameter \( \theta = (\theta_1, ..., \theta_k) \in \Theta \), or a function of it, is the object of interest of the study. The function \( \eta : \mathcal{X} \times \Theta \rightarrow \mathbb{R} \) is known and twice continuously differentiable with respect to \( \theta \), while \( \epsilon_j \) represents the random error term for each individual subject, assumed to be independently normally distributed with expectation 0 and variance \( \sigma_j^2 \in \mathbb{R}, j = 1, ..., n \).

While the ideas presented here are generally applicable, we mainly focus on toxicological examples where the function \( \eta \) is usually the four parameter variant of either the log-logistic or the Weibull function.

Throughout this paper, we will use the concept of approximate designs and consider designs as probability measures on the (finite) design space \( \mathcal{X} \) (see Kiefer, 1974). As the design space is finite, a design \( \xi \) can then be specified just by a vector of weights \( w = (w_1, ..., w_r) \in \Xi \), where \( \Xi = \{(p_1, ..., p_r) \mid p_i \geq 0, i = 1, ..., r, \sum_{i=1}^{r} p_i = 1 \} \) and the \( i \)th component \( w_i \) represents the relative proportion of total observations to be taken at the point \( x_i \in \mathcal{X} \) \((i = 1, ..., r)\). From now on, we will use \( w \) to denote such a design. Furthermore, we will consider all vectors as column vectors unless otherwise specified.

It is well known that the variance–covariance matrix of a parameter estimate for \( \theta \) is asymptotically proportional to the inverse of the (Fisher) information matrix given by

\[
M(w, \theta) = \sum_{i=1}^{r} w_i A(x_i, \theta),
\] (3)

where \( A(x_i, \theta) = \frac{1}{\sigma_i^2} \frac{\partial \eta(x_i, \theta)}{\partial \theta} \frac{\partial \eta(x_i, \theta)}{\partial \theta}^T \) denotes the information gained from a single measurement at the design point \( i \).

Often, the matrix \( A(x_i, \theta) \) is split into the component functions \( f(x_i, \theta) := \frac{1}{\sigma_i} \frac{\partial \eta(x_i, \theta)}{\partial \theta} \), that is, \( A(x_i, \theta) = f(x_i, \theta)f(x_i, \theta)^T \), \( i = 1, ..., r \). We will denote by \( \mathcal{M} \) the space of all information matrices \( M(w, \theta) \) of the form (3).
Note that in a nonlinear situation, any optimal designs will usually depend on the unknown model parameters and require an initial guess for these parameters. The designs might be sensitive to misspecification of the parameters. However, they form the basis of many more sophisticated design strategies including variants of Bayesian (Chaloner & Larntz, 1989) or adaptive designs (Dragalin et al., 2010).

An optimal design for estimating the parameter $\theta$ maximizes an appropriate information function $\phi$ of the information matrix $M(\mathbf{w}, \theta)$. An information function is a scalar-valued matrix-function fulfilling certain regularity conditions, most importantly concavity and positive homogeneity (see Pukelsheim, 2006), and numerous criteria have been proposed for this purpose in the literature. We will focus on designs allowing a balanced estimation of parameters by maximizing the determinant of the information matrix, more precisely the information function

$$\phi_D(\mathbf{w}) = |M(\mathbf{w}, \theta)|^{1/k}. \quad (4)$$

These designs are called D-optimal designs.

Sometimes, previous information about $\theta$ is not very reliable, and a design with acceptable performance for several different values of $\theta$ is desired. To do so, an optimality criterion can be defined that measures the (weighted) average performance over different values of the parameters. Such a criterion is sometimes called an average-optimum or quasi-Bayes approach (; see also Dette, Pepelyshev, & Zhigljavsky, 2008, and Yu, 2010b, for the design algorithm), even though no true Bayesian data analysis is proposed. We will only consider this approach for D-optimality, where for parameter sets $\theta_1, \ldots, \theta_m$ with weights $\beta_1, \ldots, \beta_m$ and $\sum_{j=1}^{m} \beta_j = 1$, the information function is given by

$$\phi_{Bay_D}(\mathbf{w}) = \prod_{j=1}^{m} |M(\mathbf{w}, \theta_j)|^{\beta_j/k} \quad (5)$$

$$= \exp \left( \frac{1}{k} \sum_{j=1}^{m} \beta_j \log |M(\mathbf{w}, \theta_j)| \right).$$

In the literature, the exponential function is often dropped in this criterion, which does not change the optimal design. However, in this case $\phi_{Bay_D}$ would not be positively homogeneous, thus not an information function in the sense of Pukelsheim (2006), and efficiency would not be properly defined.

### 2.3 | Equivalence theory

In order to assess whether any given design is optimal for a given criterion $\phi$, we can use equivalence theory (see Pukelsheim, 2006), which provides precise conditions to check the optimality of a given design. The concrete conditions for D- and quasi-Bayes-D-optimality are the following:

**Theorem 2.1.** A design $\mathbf{w}$ is D-optimal in a regression model with information matrix of the form (3) if and only if the normality inequality

$$\frac{1}{k} f(x_i, \theta)^T M^{-1}(\mathbf{w}, \theta) f(x_i, \theta) \leq 1 \quad (6)$$

holds for all $x_i \in \mathcal{X}$. Moreover, there is equality in (6) at any support point of the design $\mathbf{w}$.

**Theorem 2.2.** A design $\mathbf{w}$ is quasi-Bayes-D-optimal in a regression model with different information matrices $M(\mathbf{w}, \theta_j)$, $j = 1, \ldots, m$, each of the form (3) if and only if the normality inequality

$$\frac{1}{k} \sum_{j=1}^{m} \beta_j f(x_i, \theta_j)^T M^{-1}(\mathbf{w}, \theta_j) f(x_i, \theta_j) \leq 1 \quad (7)$$

holds for all $x_i \in \mathcal{X}$. Moreover, there is equality in (7) at any support point of the design $\mathbf{w}$.
Thus, we can always confirm whether a specific given design is in fact optimal. Note that Theorem 2.1 implies that 
\[ \frac{1}{k} \text{tr}(\text{M}^{-1}(\mathbf{w}, \theta)A(x_i, \theta)) = 1 \] 
for any support point of a D-optimal design.

### 2.4 Algorithms

Finding the optimal designs is generally not possible analytically and requires an algorithmic approach. The two most common approaches are both based on changing a starting design in the direction of the design points of the largest improvement of the criterion function. The Fedorov–Wynn algorithm (Fedorov, 1972; Wynn, 1972) does so by iteratively replacing part of the existing design with the singular design point showing the largest derivative, while the class of multiplicative algorithms (with convergence proven by Yu, 2010a) multiplies every weight in the existing design simultaneously by a factor depending on the derivatives. Both algorithms in their basic form tend to propose a larger than necessary numbers of support points, but there are several adaptions to account for this problem (Gaffke & Schwabe, 2019; Martin & Camacho Gutierrez, 2015; Pronzato, 2013; Yang et al., 2013; Yu, 2011). In this paper, we will focus on the multiplicative algorithm, but our results do not depend on how an optimal design was obtained.

We follow Yu (2010a) and first define a subclass \( \mathcal{W} = \{ \mathbf{w} : w_i > 0, \sum_{i=1}^{r} w_i = 1 \} \) of designs assigning nonzero weight to every design point. Then, starting with the vector \( \mathbf{w}(0) \in \mathcal{W} \) (e.g., an equal weights design), we define for \( q = 0, 1, \ldots \) a vector of “new” weights \( \mathbf{w}^{(q+1)} = (w^{(q+1)}_1, \ldots, w^{(q+1)}_r) \in \mathcal{W} \) by

\[
    w^{(q+1)}_i = w^{(q)}_i \frac{d^\lambda_i}{\sum_{j=1}^{r} w^{(q)}_j d^\lambda_j}, \quad i = 1, \ldots, r, \tag{8}
\]

where \( \lambda \in (0, 1] \) denotes a calibration parameter depending on the information function, and

\[
    d_i = d_i(\mathbf{w}^{(q)}) = \text{tr}(\nabla \phi(\mathbf{w}^{(q)})) \quad i = 1, \ldots, r \tag{9}
\]

represents the differential of an information function \( \phi \) at the current design in direction of the individual information matrices. Here, \( \nabla \phi : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}^{k \times k} \) denotes the gradient matrix containing the elementwise derivatives of a matrix-function \( \phi \).

Once an optimal design is reached, the multiplier \( \frac{d^\lambda_i}{\sum_{j=1}^{r} w_j d^\lambda_j} \) will be \( \leq 1 \) for all available design points and equal to 1 for support points. Having a value of 1 is thus a necessary, but not sufficient criterion for a support point of an optimal design.

### 2.5 Application to the example

The end result of the multiplicative algorithm applied to the example from Section 2.1 with 5,000 iterations results in the design is shown in Figure 1 (left). The graph shows all possible dose levels (i.e., the design space) on the x-axis and the weights assigned to the dose levels by the algorithm on the y-axis. We observe that while weights very close to zero are assigned to most dose levels, relevant weights are assigned to at least 12 different dose levels distributed among four different areas in the design space. Including only those proposed dose levels with weights \( \geq 0.01 \) results in the rounded design already shown in Section 2.1.

The corresponding condition of the Equivalence Theorem 2.1 from Equation (6) for this design is shown in Figure 1 (right). Here, the term from the left part of Equation (6) is plotted on the y-axis for each element of the design space. As the values are \( \leq 1 \) for all elements of the design space, we have confirmed that the design is in fact optimal.

### 3 VISUALIZING DESIGN PROBLEMS

#### 3.1 Motivation

The main aim of this paper is to derive a graphical representation that shows both the relative importance and the relationship between the available design settings in the experiment. As an example, we will use the toxicological experiment from Section 2.1.
If an optimal design $\mathbf{w}$ is already available, equivalence theory will already allow a representation covering the first of these two aspects. For $D$-optimality, we can plot the main condition given in Equation (6), that is, the term
\[
\frac{1}{k} f(x, \Theta)^T M^{-1}(\mathbf{w}, \Theta) f(x, \Theta) = \frac{1}{k} \text{tr}(M^{-1}(\mathbf{w}, \Theta) A(x, \Theta))
\]
to obtain a graph similar to the one shown in Figure 1b for the example situation. As the gradient for the $D$ criterion is given by $\nabla \phi_D(\mathbf{w}) = \frac{1}{k} M^{-1}(\mathbf{w}, \Theta)$ (see Pázman, 1986), the condition is identical to the multiplier term $d_i$ used in the multiplicative algorithm (see Equation 9).

In addition to confirming the optimality of the underlying design, this plot also suggests that there are four different areas of interest in the design space. Furthermore, we might also suspect that two design points that both have criterion values close to one and are adjacent in the graph might be exchangeable with each other. Unfortunately, this cannot be concluded from this graph alone. Also, in some situations even distant points might provide similar information, such as in a linear model with $\eta(x_j, \Theta_1, \Theta_2) = \Theta_1 + \Theta_2 x_j^2$, where very large and very small values of $x_j$ will provide identical information about $\Theta$ (see Section 3.4). Thus, a more informative representation is required.

Observing the function $d_i$, we see that it is a scalar product between the individual information matrix of the point $x_i$ (i.e., the matrix $f(x_i, \Theta)^T f(x_i, \Theta)$) and the inverse information matrix of the existing design. If we want to generalize this to also supply information about the relationship between the information from different design points, we might look into the cross-scalar product between the information from different design points $x_i$ and $x_j$, specifically the term
\[
\frac{1}{k} f(x_i, \Theta)^T M^{-1}(\mathbf{w}, \Theta) f(x_j, \Theta).
\]
This term intuitively makes sense, as it is a measure of similarity between information matrices and, in fact, we can show that the term can be used to approximate the loss in efficiency experienced when the design point $x_i$ is replaced by $x_j$ in an optimal design. We will discuss the details of this approach in Section 3.2, but to do so a more general discussion of the replacement of design points is needed.

### 3.2 Replacement of design points

In this section, we will investigate how the performance of an optimal design $\mathbf{w}^*$ will be affected if a weight of $z$ for one of the support points (say, $x_t$) is replaced by a different potential support point (say, $x_u$). This can be either a full replacement ($z = w_t^*$) or only partial ($z < w_t^*$).

While a lot of theoretical work concerning this problem already exists in regard to algorithmic approaches (e.g., Mandal & Torsney, 2006; Yu, 2011), we will focus on ideas allowing a graphical representation.

For this, we will write $\text{supp}(\mathbf{w}^*)$ for the support points of the optimal design $\mathbf{w}^*$ and $\text{supp}(\mathbf{w}^*)$ for the extended support of the design $\mathbf{w}^*$, which also includes all additional design points achieving equality in the normality equation from the equivalence theorems, or, equivalently, achieving the same value of $d_i$ in Equation (9) as the actual support points. These points are design points that could possibly be support points of an optimal design, but are not included in the specific optimal design $\mathbf{w}^*$ (as optimal designs are not necessarily unique). If the optimal design $\mathbf{w}^*$ has been constructed from
an algorithmic procedure, in practical cases it often makes sense to apply a rounding procedure here and consider the support of $\mathbf{w}^*$ to be only those design points with a weight of at least, say, 0.01.

For notational convenience, from now on we will write $M = M(\mathbf{w}^*, \Theta)$ for the information matrix of the optimal design, $A_i = A(x_i, \Theta)$ for the elementary information matrices of the design points $x_i$ ($i = 1, \ldots, r$), and, specifically, $A_t = A(x_t, \Theta)$ for the information matrix of the removed design point $x_t$ and $A_e = A(x_e, \Theta)$ for the information matrix of the newly added design point $x_e$. Similarly, we write $f_i$ for the component function $f(x_i, \Theta)$, $f_e$ for $f(x_e, \Theta)$, and $f_t$ for $f(x_t, \Theta)$.

We will consider three different approaches how the effect of replacements can be expressed.

Of course, the first and immediate approach is to just directly calculate the efficiency of an alternate design $\tilde{\mathbf{w}}_z$ constructed by (partial) replacement of a weight $z$ from a single support point $x_t$ through the equivalent weight on point $x_e$. By definition, this efficiency is given by $\text{eff}(\tilde{\mathbf{w}}_z) = \frac{\phi(M(\tilde{\mathbf{w}}_z))}{\phi(M(\mathbf{w}^*))}$. (See also Fedorov, 1972, p. 161, for some alternative expressions for this quantity.)

In order to compare the effects of all possible replacements, we would have to consider three dimensions: the replaced point $x_t$, the replacement $x_e$, and the replaced weight $z$. In order to obtain a simpler, two-dimensional representation, we could alternatively consider the replacement of just a uniform weight $z_{\min}$, which is small enough to be taken from any of the support points $x_i$ and obtain the following lemma:

**Lemma 3.1.** Let $\mathbf{w}^* \in \Xi$ be a design which is optimal in regard to an information function $\phi$. Let $\mathbf{w}(t, e, z) = \tilde{\mathbf{w}}_z \in \Xi$ be an alternative design constructed from $\mathbf{w}^*$ as follows ($i, e = 1, \ldots, r$; $x_i \in \text{supp}(\mathbf{w}^*)$; $0 < z \leq w^*_t$):

$$
\tilde{w}_{z,i} := \tilde{w}_i(t, e, z) = \begin{cases} 
w^*_i - z & \text{if } i = t \\
w^*_i + z & \text{if } i = e \\
w^*_i & \text{if } i \not\in t, e
\end{cases} \quad (10)
$$

Furthermore, denote $z_{\min} = \{\min(w^*_i)|x_i \in \text{supp}(\mathbf{w}^*)\}$ the weight of the support point of $\mathbf{w}^*$ with the smallest weight. Then, the efficiency of $\tilde{\mathbf{w}}_z$ is

$$
\text{eff}(\tilde{\mathbf{w}}_z) = \begin{cases} 
\geq 1 + \frac{z}{z_{\min}} \left( \frac{\phi(M(\tilde{\mathbf{w}}_{z_{\min}}))}{\phi(M(\mathbf{w}^*))} - 1 \right) & \text{if } z < z_{\min} \\
\frac{\phi(M(\tilde{\mathbf{w}}_{z_{\min}}))}{\phi(M(\mathbf{w}^*))} & \text{if } z = z_{\min} \\
\leq 1 + \frac{z}{z_{\min}} \left( \frac{\phi(M(\tilde{\mathbf{w}}_{z_{\min}}))}{\phi(M(\mathbf{w}^*))} - 1 \right) & \text{if } z > z_{\min}
\end{cases} \quad (11)
$$

**Proof.** By definition $\text{eff}(\tilde{\mathbf{w}}_{z_{\min}}) = \frac{\phi(M(\tilde{\mathbf{w}}_{z_{\min}}))}{\phi(M(\mathbf{w}^*))}$. From the concavity of $\phi$, we conclude that, for an interpolation ($z \leq z_{\min}$), the equation

$$
\frac{z}{z_{\min}} \phi(M(\tilde{\mathbf{w}}_{z_{\min}})) + (1 - \frac{z}{z_{\min}}) \phi(M(\mathbf{w}^*))
= \phi(M(\mathbf{w}^*)) + \frac{z}{z_{\min}}(\phi(M(\tilde{\mathbf{w}}_{z_{\min}})) - \phi(M(\mathbf{w}^*))) \leq \phi(M(\tilde{\mathbf{w}}_z))
\Rightarrow 1 + \frac{z}{z_{\min}} \left( \frac{\phi(M(\tilde{\mathbf{w}}_{z_{\min}}))}{\phi(M(\mathbf{w}^*))} - 1 \right) \leq \frac{\phi(M(\tilde{\mathbf{w}}_z))}{\phi(M(\mathbf{w}^*))}
$$

supplies a lower limit for the efficiency. Similarly, extrapolation supplies an upper limit for values of $z \geq z_{\min}$.

Thus, we obtain the exact effect of a replacement of an identical weight $z_{\min}$ taken from any of the support points $x_i$. If a different weight $z$ is replaced instead, the lemma will supply either an upper or a lower limit for the resulting loss in efficiency. Of course, this approach still requires computation of the actual information functions for every possible combination of $t$ and $e$. Furthermore, if the weight of the replaced design point is larger than the smallest weight in the existing design, the lemma will only supply an upper limit for the resulting efficiency, which is of limited practical
usefulness. Still, it will be possible to construct an exact graphical representation using the constant replacement weight $z_{\text{min}}$.

The second possible approach again makes use of the concavity of information functions. That way, a direct approximation of the loss in efficiency caused by a replacement is possible, but unfortunately not very useful. For a concave information function $\phi$, we can write, for any new design point $x_e \in \mathcal{X}$ and for any $z \in [0, w_t]$:

$$
\phi(M + z A_e - z A_t) \leq \phi(M) + \text{tr}(\nabla \phi(M)(z A_e - z A_t)) = \phi(M) + z \text{tr}(\nabla \phi(M)A_e) - z \text{tr}(\nabla \phi(M)A_t) \leq \phi(M).
$$

(13)

If a support point of the optimal design is specifically replaced by a second point with an identical critical function $d_i$ (i.e., an element of $\text{supp}_x(w^*)$), then Equation (13) only supplies the trivial result that the resulting efficiency is 100% or less. This is especially not sufficient to conclude that the design is still optimal, as the new support point might contribute a different kind of information compared to the replaced one.

If, however, the new point is not part of the extended support, that is, has a lower value of the critical function $d_i$, Equation (13) does actually provide a useful upper limit for the efficiency of the resulting design. Thus, while this approach does not seem suitable for a stand-alone graphical representation, it does allow identifying replacements that will definitely reduce performance of a design.

As a third approach, we construct a derivative-based approximation for the loss of efficiency caused by a replacement of design points. It is based on a general lower limit for the efficiency of designs presented in Holland-Letz, Dette, and Renard (2012), Theorem 1. While this will only provide an approximate lower limit for the loss, the result is computationally efficient and allows a straightforward visualization of all possible replacements.

Theorem 3.2. Let $w^* \in \Xi$ be a design that is optimal in regard to an information function of either $\phi$ or $\log(\phi)$, with an extended support $\text{supp}_x(w^*)$ that includes $s$ design points. Let $\tilde{w}(t, e) = \tilde{w} \in \Xi$ be an alternative design constructed from $w^*$ as follows ($i = 1, \ldots, r; x_i \in \text{supp}_x(w^*); x_e \in \mathcal{X}; 0 < z \leq w^*_t$):

$$
\tilde{w}_i = \begin{cases} 
  w^*_i - z & \text{if } i = t \\
  w^*_i + z & \text{if } i = e \\
  w^*_i & \text{if } i \notin t, e
\end{cases}
$$

(14)

Then $\text{eff}(\tilde{w})$ is bounded by

$$
\text{eff}(\tilde{w}) \geq \left( \frac{\text{max}_{x_i \in \{x_e, \text{supp}_x(w^*)\}} \text{tr}(\nabla \phi(M)A_i) + \text{tr}(d \nabla \phi(M,z A_e - z A_t)A_i) + o(z)}{\text{tr}(\nabla \phi(M(\tilde{w}, \theta))M(\tilde{w}, \theta))} \right)^{-1},
$$

(15)

where the approximation is centered around $z = 0$, and the notation $d \nabla \phi(X, Y)$ represents the differential of the gradient matrix $\nabla \phi$ at $X$ in direction $Y$.

Proof. See the Appendix.

The quality of the approximation depends on the magnitude of $z A_e - z A_t$. Thus, it works very well for either small replacements (small $z$) or similar information matrices $A_e$ and $A_t$. As the latter is usually the case for most replacements, the approximation works well in practice.

While Theorem 3.2 is somewhat unwieldy, applying it to specific optimality criteria often results in much more straightforward relationships. For the most important practical case of $D$-optimality, we obtain the following corollary, dropping the $o(z)$ notation:

Corollary 3.3. Let $\phi = \phi_D = |M(w, \theta)|^{1/k}$, the information function corresponding to $D$-optimality. In this case, Equation (15) reduces to

$$
\text{eff}(\tilde{w}) \geq \left( 1 + zk \left( 1 - \frac{\text{tr}(M^{-1} A_e M^{-1} A_t)}{k^2} \right) \right)^{-1}.
$$

(16)
If \( \text{tr}(M^{-1}A_eM^{-1}A_t) = k^2 \), exact equality is achieved in Equation (16) and the resulting design is still optimal.

Proof. See the Appendix.

The approximate loss in \( D \)-efficiency only depends on the specific replacements through the term \( \text{tr}(M^{-1}A_eM^{-1}A_t) \) and no longer includes a maximum term. Primarily, we should consider replacements that are also an element of the extended support of the optimal design, that is, that have the same value of the critical function \( (d_t = d_e) \). As seen from Equation (13), replacing design points with points not fulfilling this criterion might at best provide a decent design, but never an optimal one. However, the corollary is valid in either case.

As an additional example, we also consider the case of quasi-Bayesian \( D \)-optimality:

**Corollary 3.4.** Let \( \phi = \phi_{BayD} = \exp\left( \frac{1}{k} \sum_{j=1}^{m} \beta_j \log |M(w, \theta_j)| \right) \), the information function corresponding to quasi-Bayes-\( D \)-optimality. In this case, Equation (15) reduces to

\[
\text{eff}(\tilde{w}) \geq \left( 1 + \frac{z}{k} \sum_{j=1}^{m} \beta_j \text{tr}(M(w, \theta_j)^{-1}(A_t(\theta_j) - A_e(\theta_j))M(w, \theta_j)^{-1}A_t(\theta_j)) \right)^{-1}. \tag{17}
\]

Proof. See the Appendix.

No exact equality result is available here.

### 3.3 | Design heatmaps

In the previous section, we have discussed three approaches towards handling replacements, two of which allow a graphical representation (Lemma 3.1 and Theorem 3.2).

#### 3.3.1 | Type A heatmaps

The first approach (called “Design Heatmap Type A”) involves plotting the direct efficiency change resulting from replacements, given by the term \( \frac{\phi(M(w, \theta_j))}{\phi(M(w^*))} \) as defined in Lemma 3.1, for all design points \( t \in \text{support supp}(w^*) \) of the optimal design on the \( x \)-axis versus all possible replacement points \( e \in \{1, ..., r\} \) on the \( y \)-axis. In the rounded example design, there were 12 design points, and the graph will thus show 12 design points on the \( x \)-axis, while the \( y \)-axis will show the full space of all \( r = 201 \) dose levels representing the possible replacements for each of these 12 points. As the smallest weight of the rounded initial design was 0.014, the replacement fraction is therefore set as \( z_{\text{min}} = 0.014 \) and every change in efficiency thus refers to the change caused by the replacement of a weight of 0.014. The resulting plot is shown in Figure 2. The actual \( D \)-efficiencies achieved by any replacement are shown as colors in the plot, and the coloring is scaled so that lossless replacements (efficiency of 1) are colored red, while increasing losses in efficiency change the coloring along the spectrum, with efficiencies of 0.995 or less uniformly shown as blue. Note that as we replace only a very small fraction of the observations (0.014) here, the change in efficiency is small.

We observe that the two central support point areas allow only limited flexibility in regard to their exact placement. The two adjacent design points proposed per area by the initial algorithmic solution (see Figure 1, left) seem to be basically interchangeable, but most other replacements will produce a relevant decrease in efficiency. Contrary to this, much more flexibility is available in regard to the boundary points, where, for example, the design point at a log dose of \(-10\) can be replaced by any other design point up to a log dose of at least \(-9\) or even up to \(-7\) without any relevant loss of efficiency, as the color remains nearly red up to this point. We also notice that moving a small weight from one original support point to another support point in a different region creates a limited but noticeable loss in efficiency.

As the Type A Heatmap plots all design points of an initial design against all possible replacements, it does depend on the number of support points of the original design. In the context of optimal design algorithms this means that, when the (rounded) support of an algorithmic design decreases with increasing iterations of the algorithm, the heatmap will change as well.
3.3.2 Type B heatmaps

As an alternative, we can also look at a derivative-based heatmap discussed as the third approach in Section 3.2. As shown in this section, in the case of $D$-optimality, the relative loss of efficiency caused by a replacement of design points is approximately determined by the term $\frac{1}{k^2} \text{tr}(M^{-1}A_i M^{-1}A_e)$, which can be rewritten as

$$\frac{1}{k^2} \text{tr}(M^{-1}A_i M^{-1}A_e) = \frac{1}{k^2} \text{tr}(M^{-1}f_i f_i^T M^{-1}f_e f_e^T) = \frac{1}{k^2} (f_e^T M^{-1} f_i)^2.$$  \hspace{1cm} (18)

Thus, it is indeed, as mentioned in Section 3.1, the squared scalar product between the vectors of derivatives of the original dose–response function $\eta$ in relation to the matrix $M^{-1}$, evaluated at the two design points marked for exchange. We will call Equation (18) the reduction factor. Introducing the matrix $F = F(\Theta) = (f_1, \ldots, f_r)$, the reduction factor for all possible replacements can be expressed as the second Hadamard product (i.e., the elementwise square, operator $\circ$) of the matrix

$$L = \frac{1}{k} F^T M^{-1} F,$$  \hspace{1cm} (19)
that is,

\[(\text{LoL})_{k,l} = \frac{1}{k^2} \left( f_k^T M^{-1} f_l \right)^2, \]  

which computationally is a very simple calculation. Of course, the results from Corollary 3.3 only apply to the entries of $L$ corresponding to design points actually part of the support of the optimal design.

Plotting LoL results in an alternative graphical representation we call the “Design Heatmap Type B.” Figure 3 (left) shows this heatmap for the log-logistic model from our example. Figure 3 (right) shows the same graphic, but with the color distribution magnified on values close to 1.

Here, both rows and columns represent the $r$ elements of the design space $\mathcal{X}$ and correspond to the design point to be replaced, and to the replacement, respectively. The color is determined by the exact value of the term $\text{LoL}$, with red again representing a value of 1, indicating a replacement without loss of efficiency. Lower values are represented by different colors, as shown in the legend, and indicate that a replacement will result in some loss of efficiency. Unfortunately, the relationship between the reduction factor and efficiency is not as direct as with the Type A heatmaps, but the ordering, at least for small replacements, is preserved.

The graph thus shows two different kinds of information. First, the diagonal helps to identify which design points are candidates for support points of an optimal design at all. In these positions, the plot shows the standardized scalar products of the information matrices of the design points with themselves. According to the equivalence theorem (see Section 2.3), a value of 1 in this criterion is a necessary condition for an optimal support point of a $D$-optimal design. As this corresponds to a red coloring, all potential support points will thus be marked red on the diagonal. Second, the nondiagonal elements represent the similarity of the information gained from the design point corresponding to the $x$-axis position to the information gained from the design point corresponding to the $y$-axis position. A red coloring there thus represents identical information gained from these two design points. Consequently, any larger group of design points colored red at every pairwise intersection indicates an area of nearly identical information, and observations can thus be moved freely within these areas without any relevant losses in efficiency.

In the example heatmap of Type B, we again observe that very high or low dose levels (above a log dose of ±9) form a solid red block and are thus basically interchangeable, while observing the two red areas in the center of Figure 1 we notice that the information provided by points in these areas is similar, but not identical. Furthermore, similarity decreases quickly with distance between two points. Accordingly, zooming into the right part of the figure shows us that the two pairs of central support points of our original design (Figure 1) can be combined into one, but a very slight loss in efficiency might result.

Note that this representation is symmetrical, and both the $x$- or the $y$-axis can be used to represent either the replaced point or the replacement. Also, Type B heatmaps only depend on the (unique) optimal information matrix $M$. Thus, they do not depend on the exact structure of the initial optimal design.
3.3.3 | Use of design heatmaps

Using the design heatmaps of either type, we can immediately determine which of the support points of a given design with many support points can be combined without any losses, which can be moved, and which can be joined with a small but acceptable loss of efficiency. If multiple replacements are planned, the heatmaps (as well as the information matrix of the resulting design) would remain constant after lossless replacements, but would have to be recalculated after any other replacement. However, as replacements are usually made with design points with similar information matrices, the original heatmaps still provide a good guideline even for multiple replacements. Note furthermore that the Type B Heatmap also includes entries for nonsupport points of the optimal design (i.e., those not marked red on the diagonal). These points are not candidates for replacement, but the depicted scalar products still provide some insight about the similarity of information matrices.

In general, there is no clear distinction which of the two types of heatmaps is superior. Both types have advantages and disadvantages:

- Type A heatmaps show the exact loss in efficiency for small replacements, while for Type B the relationship is only indirect and approximate.
- Type A heatmaps still accurately illustrate the effect of small replacements even with very different information matrices, where the approximation used by Type B is less effective.
- For larger replaced weights, the accuracy advantage of Type A is lost as only an upper limit for the replacement effect is provided.
- Type A heatmaps only show design points part of the initial optimal design, while Type B includes the whole design space.
- As they use scalar products, Type B heatmaps illustrate the similarity of all individual information matrices, even those not part of the optimal design.
- Type A heatmaps depend on the (nonunique) initial optimal design and are therefore not unique. Type B heatmaps are unique as they only depend on the unique optimal information matrix.
- Type B heatmaps are symmetric and show the condition of the equivalence theorem for D-optimality on the diagonal.

In our opinion, Type B heatmaps are also much accessible for practical users without in-depth design background, as they allow quick identification which of the different areas in the design space provide distinct relevant information and which do not. Also, they are extremely simple to compute, using only a single matrix multiplication of terms already available from the design algorithm.

3.4 | Alternative example: Linear model

In the toxicological example, most relevant design points of similar information are close to each other. However, this does not have to be the case. We illustrate this with another example that shows the additional information a design heatmap can provide in this situation. To do so, we plot the Type B heatmap for the previously mentioned linear model \( \eta(x_j, \theta_1, \theta_2) = \theta_1 + \theta_2 x^2_j \) (see Figure 4).

We observe that while a center design point is always required, the design points at the very edges of the design space in both directions can be replaced by each other without loss in efficiency.

3.5 | Design heatmaps for quasi-Bayes D-optimality

The previous examples for both types of heatmaps specifically refer to the most important practical case of non-Bayesian D-optimality. However, similar plots can be constructed for different criteria as well, as long as the relevant part of the corresponding corollary can be normalized to fall into the interval from 0 to 1.

As a third example, we thus consider the toxicological situation mentioned before, but instead of a fixed assumption for the parameter vector \( \theta \), we consider the quasi-Bayesian setup with three different possibilities for \( \theta \): \( \theta_1 = (0, 1, 1, 0.5) \), \( \theta_2 = (0, 1, 1, 1) \), and \( \theta_3 = (0, 1, 1, 2) \), each with a probability of 0.333. In the toxicological context, this represents three
**FIGURE 4**  
*D*-optimality Design Heatmap Type B for a linear model

**FIGURE 5**  
(Left): Log-logistic dose–response functions with $\theta_4 = 0.5$, $\theta_4 = 1$ and $\theta_4 = 2$. Parameters $(\theta_1, \theta_2, \theta_3)$ are given as (0,1,1).  
(Right): Initial optimal design with Bayes parameter assumptions

different assumptions in regard to the $ED_{50}$ levels of the substance. The corresponding true log-logistic dose–response curves are shown in Figure 5 (left). Furthermore, we only run the algorithm for 400 iterations.

The resulting design is shown in Figure 5 (right) and has a large number of support points, many of them spread around the central region with log-doses between $-1.5$ and $1.5$. Many of the design points have very low weights, and such a design would usually be impractical in real applications.

A Design Heatmap Type A based on these results and Lemma 3.1 is shown in Figure 6 (left). In this case, an alternative heatmap based on Corollary 3.4 would also be possible, for example, by plotting the terms $1 - \frac{1}{k^2} \sum_{j=1}^{m} \beta_j \text{tr}(M(w, \theta_j)^{-1}(A_j(\theta_j) - A_j(\theta_j))M(w, \theta_j)^{-1}A_j(\theta_j))$, which according to Corollary 3.4 should be 1 for lossless replacements of an existing support point and $<1$ for other replacements. However, this graph does not have
the useful properties of the previous Type B heatmaps and, specifically, is not symmetrical and also does not show the equivalence theorem conditions on the diagonal. Thus we propose to plot the scalar products instead, more precisely the terms \( \frac{1}{k} \sum_{j=1}^{m} \hat{\beta}_j f(x_t, \theta_j)' M^{-1} (\mathbf{w}, \theta_j) f(x_e, \theta_j) \). The resulting heatmap is shown in Figure 6 (right).

We observe that, compared to the \( D \)-optimal designs, a larger number of different support points will indeed be required in the central dose areas. However, the overly large number of different dose levels in this area proposed by the initial algorithm can be reduced to three different dose levels. In regard to the very large and small dose levels, the situation remains similar to the non-Bayesian case.

### 3.6 Algorithm to reduce the number of support points

Using either Lemma 3.1 or Corollary 3.3, we can formulate a simple algorithm to quickly and automatically reduce the number of number of support points of a given optimal design, while keeping the loss of efficiency to a minimum. Note that a more sophisticated adaption to the standard algorithms using similar principles has already been proposed by Yu (2011), and a much more efficient general algorithm to directly find the optimal design with the smallest possible support has been proposed by Yang et al. (2013). However, this algorithm is still a useful application of our results that can be applied to overly complex optimal designs obtained from any source and furthermore can also be used to find designs which are even simpler without being strictly optimal anymore.

Doing this in an automatic way requires setting a limit \( g \in (0, 1] \) for the relative approximate loss in efficiency where a replacement of design points is still acceptable. Generally, if the aim is just to reduce the number of support points without compromising efficiency, a value of \( g = 0.99 \) is suitable. If further simplifications are desired, even at the cost of efficiency, smaller values like \( g = 0.95 \) or even \( g = 0.80 \) might be chosen. While Corollary 3.3 only applies to \( D \)-optimality, a similar but computationally slightly more complex approach could be formulated for any optimality criterion using Lemma 3.1, Corollary 3.4, or Theorem 3.2. For \( D \)-optimality and Corollary 3.3., however, the procedure is the following:

1. Run any optimal design algorithm with a low number of iterations until a design \( \mathbf{w}^* \) with an efficiency close to 1 is obtained.
2. Calculate the matrix of scalar products \( L = \frac{1}{k} F^T M^{-1} F \).
3. Set a maximum value \( g \leq 1 \) of the scalar product determining which points should be combined (reduction factor). \( g = 1 \) indicates only lossless replacements.
4. Determine \( g_{\text{max}} = \max_{i,j \in \text{supp}(\mathbf{w}^*)} (L_{i} L_{j})_{i,j} \). Let \( L_{t,e} \) denote the position of the maximum.
5. If \( g_{\text{max}} < g \), end algorithm. Otherwise, determine whether \( \text{tr}(M^{-1} A_t) \geq \text{tr}(M^{-1} A_e) \). If yes, move all weight from point \( x_e \) to point \( x_t \); if no, move all weight from point \( x_t \) to \( x_e \).
6. Repeat steps 4 and 5 until \( g_{\text{max}} < g \).
Finally, rerun the original design algorithm using only the support points of the final design as new design space. That way, some fine tuning of weights can be achieved.

Step 7 is entirely optional, but helps in adjusting for the minor errors accumulated during repeated replacements. Of course, the final efficiency of the simplified design compared to the initial optimal design should be calculated afterwards in any case; also see Yang et al. (2013) for a more sophisticated approach towards the problem of adjusting weights on fixed design points.

As the algorithm will usually replace several support points in sequence, we might consider recalculating the matrix $L$ after nonlossless replacements (i.e., repeat steps 2, 4, and 5 instead of just 4 and 5). However, as usually support points are replaced by points of extremely similar information, the practical effects of this are negligible and reusing the original $L$ matrix seems to be sufficient.

Applying this algorithm to the example at the beginning (log-logistic function, $\theta = (0, 1, 1, 1)$), we first run the multiplicative algorithm for 5,000 iterations to obtain our original design shown in Figure 1, which has 12 support points with weights $> 0.01$:

$$\begin{align*}
\text{Log Dose} & \quad -10.0 & -9.9 & -9.8 & -9.7 & -1.1 & -1.0 & 1 & 1.1 & 9.7 & 9.8 & 9.9 & 10.0 \\
\text{Weights} & \quad 0.13 & 0.07 & 0.03 & 0.01 & 0.10 & 0.15 & 0.15 & 0.10 & 0.01 & 0.03 & 0.07 & 0.13
\end{align*}$$

The corresponding design heatmap has already been shown in Figure 3. We now run the reduction algorithm with a reduction parameter of $g = 0.99$ and obtain the following 4-point design, still with an efficiency of 0.999:

$$\begin{align*}
\text{Log Dose} & \quad -10.0 & -1.1 & 1.1 & 10 \\
\text{Weights} & \quad 0.25 & 0.25 & 0.25 & 0.25
\end{align*}$$

Thus, even combining only nearly equivalent support points does not necessarily cause a relevant loss in efficiency.

4 DISCUSSION AND SUMMARY

In this paper, we derived two different two-dimensional graphical representations for $D$-optimal designs, which illustrate the interrelations between the support points of an optimal design and all other potential design points. Our first approach plots the direct change of efficiency for any paired replacements of one design point by another, while the second approach shows the similarity of the information gained from design points by plotting the scalar products between the information matrices gained from any pair of points. As a nice side effect, the second approach will show the conditions for the well-known equivalence theorem for $D$-optimality on the diagonal. Both representations give an intuitive overview of the design problem and further show which alterations can be made to the design without major losses in efficiency. Both visualizations are easy to compute and remain two-dimensional even for models with larger numbers of parameters.

Design heatmaps for nonlinear models will depend on the unknown true parameters. However, at least for the log-logistic function the effect of different parameters will only affect the scale used for the design space and thus will not change the general structure of the heatmaps (see Holland-Letz and Kopp-Schneider, 2014). While our work was motivated by dose–response models, the results should be applicable for many other design questions involving nonlinear regression models.

In the future, we plan to include design heatmaps in R-shiny applications to support experimentalists in regard to experimental design.

ACKNOWLEDGMENTS

The authors would like to thank both the associate editor and two anonymous referees for very constructive comments on an earlier version of this paper.

Open access funding enabled and organized by Projekt DEAL.
CONFLICT OF INTEREST
The authors have declared no conflict of interest.

OPEN RESEARCH BADGES
This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge “Reproducible Research” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID
Tim Holland-Letz https://orcid.org/0000-0003-0348-8163
Annette Kopp-Schneider https://orcid.org/0000-0002-1810-0267

REFERENCES
Chaloner, K., & Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. Journal of Statistical Planning and Inference, 21, 191–208.

Clothier, R., Gomez-Lechon, M., Honegger, P., Kinsner-Ovaskainen, A., & Kopp-Schneider, A. (2013). Comparative analysis of eight cytotoxicity endpoints within the acutetox project. Toxicology in Vitro, 27, 1347–1356.

Dette, H. (1993). Elfving’s theorem for $D$-optimality. Annals of Statistics, 21, 753–766.

Dette, H., Pepelyshev, A., & Zhigljavsky, A. (2008). Improving updating rules in multiplicative algorithms for computing $D$-optimal designs. Computational Statistics and Data Analysis, 52(2), 312–320.

Elving, G. (1952). Optimal allocation in linear regression theory. Annals of Mathematical Statistics, 23, 255–262.

Fedorov, V. V. (1972). Theory of Optimal Experiments. New York, NY: Academic Press.

Gaffke, N., & Schwabe, R. (2019). Quasi-Newton algorithm for optimal approximate linear regression design: Optimization in matrix space. Journal of Statistical Planning and Inference, 198, 62–78.

Holland-Letz, T., Dette, H., & Renard, D. (2012). Efficient algorithms for optimal designs with correlated observations in pharmacokinetics and dose finding studies. Biometrics, 68, 138–145.

Holland-Letz Tim, Kopp-Schneider Annette (2015). Optimal experimental designs for dose–response studies with continuous endpoints. Archives of Toxicology, 89, (11), 2059–2068. https://doi.org/10.1007/s00204-014-1335-2.

Kiefer, J. (1974). General equivalence theory for optimum designs. Annals of Statistics, 2, 849–879.

Magnus, J. R., & Neudecker, H. (1988). Matrix differential calculus with applications in statistics and econometrics. New York, NY: John Wiley & Sons.

Mandal, S., & Torsney, B. (2006). Construction of optimal designs using a clustering approach. Journal of Statistical Planning and Inference, 136(3), 1120–1134.

Martin, R., & Camach Gutierrez, I. (2015). Combined algorithm to compute D-optimal designs. Journal of Computational and Applied Mathematics, 278, 248–257.

Pázmán, A. (1986). Foundations of optimum experimental design. Dordrecht, the Netherlands: D. Reidel.

Pronzato, L. (2013). A delimitation of the support of optimal designs for Kiefer’s phi-p-class of criteria. Statistics and Probability Letters, 83, 2721–2728.

Pronzato, L., & Pázman, A. (2013). Design of experiments in nonlinear models. New York, NY: Springer.

Puikisheim, F. (2006). Optimal design of experiments, classics in applied mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Ritz, C. (2010). Toward a unified approach to dose-response modeling in ecotoxicology. Environmental Toxicology and Chemistry, 29(1), 220–229.

Wynn, H. P. (1972). Results in the construction of $D$-optimum experimental designs. Journal of the Royal Statistical Society B, 34, 133–147.

Yang, M., Biedermann, S., & Tang, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. Journal of the American Statistical Association, 108(504), 1411–1420.

Yu, Y. (2010a). Monotonic convergence of a general algorithm for computing optimal designs. Annals of Statistics, 38, 1593–1606.

Yu, Y. (2010b). On a multiplicative algorithm for computing Bayesian D-optimal designs. arXiv, arXiv:1005.2355 [stat.CO].

Yu, Y. (2011). D-optimal designs via a cocktail algorithm. Statistics and Computing, 21, 475–481.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.
APPENDIX: Proofs

Proof of Theorem 3.2. From Holland-Letz et al. (2012), Theorem 1, we obtain that the efficiency of a design is bounded by

\[
\text{eff}(\mathbf{w}) \geq \left( \max_{i=1}^{r} \frac{\text{tr}(\nabla \phi(M(\mathbf{w}, \theta))A_i)}{\text{tr}(\nabla \phi(M(\mathbf{w}, \theta))M(\mathbf{w}, \theta))} \right)^{-1}.
\]  

(A1)

We immediately notice that multiplicative factors to \(\nabla \phi\) cancel out in this term. Thus, the efficiency bound will be the same for \(\phi\), \(\log(\phi)\) and \(e^{\phi}\).

Heuristically, this theorem means that the inefficiency of a given design can be assessed by the slope of the information function in direction of the information matrix of that individual design point that provides the largest improvement over the current design. In the case of an optimal design, no such improvement is possible, and the lower limit will be 1. We are now interested in the performance of an alternative design \(\tilde{\mathbf{w}}\) with information matrix \(M(\tilde{\mathbf{w}}) = M - zA_t + zA_e\).

Using a linear approximation, we obtain

\[
\text{tr}(\nabla \phi(M(\tilde{\mathbf{w}})))A_i) = \text{tr}(\nabla \phi(M))A_i) + \text{tr}(d\nabla \phi(M, zA_e - zA_t)A_i) + o(z),
\]  

(A2)

for \(z \to 0\), which can be inserted into Equation (A1).

However, obtaining the maximum in Equation (A1) requires checking and taking the maximum of the derivatives in the direction of every possible design point, that is, every element of \(\mathcal{X}\).

Fortunately, we can show that only the differentials in the direction of the information matrix of the extended support points of the optimal design as well as the replacement point are relevant. For this, consider an alternative reduced design space \(\tilde{\mathcal{X}}\) consisting of only the \(s\) design points of the extended support of the optimal design \(\mathbf{w}^*\) plus the replacement point and defined by \(\tilde{\mathcal{X}} = \{x_e\} \cap \{x_i | x_i \in \text{supp}(\mathbf{w}^*), i = 1, \ldots, s\}\). As \(\mathbf{w}^*\) corresponds to a valid design on \(\mathcal{X}\), and \(\tilde{\mathcal{X}}\) is a subset of \(\mathcal{X}\), the design \(\mathbf{w}^*\) will also be optimal on \(\tilde{\mathcal{X}}\). Furthermore, \(\tilde{\mathbf{w}}\) can also be rewritten as a valid design on \(\tilde{\mathcal{X}}\) with unchanged information matrix. Thus, Equation (A1) still holds, but the direction of the maximum derivative will be an element of \(\tilde{\mathcal{X}}\), that is, an element of the extended support of the optimal design.

Rewriting Equation (A1), we obtain

\[
\text{eff}(\tilde{\mathbf{w}}) \geq \left( \max_{x_i \in \{x_e, \text{supp}(\mathbf{w}^*)\}} \frac{\text{tr}(\nabla \phi(M))A_i) + \text{tr}(d\nabla \phi(M, zA_e - zA_t)A_i) + o(z)}{\text{tr}(\nabla \phi(M(\tilde{\mathbf{w}}, \theta))M(\tilde{\mathbf{w}}, \theta))} \right)^{-1},
\]  

(A3)

which proves the theorem.

Proof of Corollary 3.3. The gradient of the D-criterion is given by \(\nabla \phi_D(\mathbf{w}) = \frac{1}{k} |M(\mathbf{w}, \theta)|^{-1} M^{-1}(\mathbf{w}, \theta)\) (see Pázman, 1986). \(\frac{1}{k} |M(\mathbf{w}, \theta)|^{-1}\) is a multiplicative constant which cancels out in Equation (A1). Furthermore, the differential of \(M^{-1}\) is known to be \(dM^{-1} = -M^{-1} dMM^{-1}\) (see Magnus & Neudecker, 1988). Before entering these results into Theorem 3.2, however, we will show that in case of D-optimality it is possible to formulate a condition similar to Equation (15) using only the differential in the direction of the information matrix of the replaced design point \(x_t\).

To prove this, we make use of the observation by Pukelsheim (2006) that optimizing on a design space \(\mathcal{X}\) is equivalent to optimizing on the space of the induced information matrices of every design point. Thus, similar to the proof of Theorem 3.2, we define an alternative design space \(\tilde{\mathcal{X}} = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3\}\), which this time includes three available design points with corresponding information matrices \(A(\tilde{x}_1) = \frac{M - 1}{w_i} A_t\), \(A(\tilde{x}_2) = A_t\) and \(A(\tilde{x}_3) = A_e\), with \(A(\tilde{x}_1)\) representing the information of the optimal design without the support point \(x_t\) (scaled to a weight of 1), \(A(\tilde{x}_2)\) representing the information matrix of the support point \(x_t\) and \(A(\tilde{x}_3)\) representing the information matrix of the additional support point \(x_e\). Note that \(\tilde{x}_1\) represents observations taken at several different original design points. However, in the space of
information matrices the joint information from these observations can still be represented as a single design point, though with a rank of $\geq 1$.

Again, the space of information matrices induced by designs on $\tilde{\mathcal{X}}$ is a subset of the information matrix space induced by $\mathcal{X}$, and the information of both the original optimal design $\mathbf{w}^*$ and the replacement design $\tilde{\mathbf{w}}$ can be reproduced on this modified design space (using the weights of $(1 - \mathbf{w}^*_d, \mathbf{w}^*_d, 0)$ and $(1 - \mathbf{w}^*_d, \mathbf{w}^*_d - z, z)$, respectively). Thus, $\mathbf{w}^*$ is still optimal on $\tilde{\mathcal{X}}$, and Equation (15) can be applied to this design space.

As $\tilde{\mathcal{X}}$ has three elements, three differentials will be needed:

\[
1 + \frac{\text{tr}(d\nabla \phi(M, zA_e - zA_t)A(\tilde{x}_3))}{\text{tr}(\nabla \phi(M)M)} = 1 + \frac{-\text{tr}(M^{-1}(zA_e - zA_t)M^{-1}A(\tilde{x}_3))}{\text{tr}(M^{-1}M)}
= 1 + \frac{-\text{tr}(M^{-1}(zA_e - zA_t)M^{-1}A_e)}{k}
= 1 + \frac{z\text{tr}(M^{-1}A_e M^{-1}A_e) - z\text{tr}(M^{-1}A_e M^{-1}A_e)}{k}
= 1 + zk \left( \frac{\text{tr}(M^{-1}A_e M^{-1}A_e) - \text{tr}(M^{-1}A_e)^2}{k^2} \right)
\leq 1 + zk \left( \frac{\text{tr}(M^{-1}A_e)\text{tr}(M^{-1}A_e) - \text{tr}(M^{-1}A_e)^2}{k^2} \right)
= 1 + zk \frac{\text{tr}(M^{-1}A_e) k - \text{tr}(M^{-1}A_e)}{k},
\]

(A4)

\[
1 + \frac{\text{tr}(d\nabla \phi(M, zA_e - zA_t)A(\tilde{x}_2))}{\text{tr}(\nabla \phi(M)M)} = 1 + zk \left( 1 - \frac{\text{tr}(M^{-1}A_e M^{-1}A_e)}{k^2} \right)
\geq 1 + zk \left( 1 - \frac{\text{tr}(M^{-1}A_e)\text{tr}(M^{-1}A_e)}{k^2} \right)
= 1 + zk \frac{k - \text{tr}(M^{-1}A_e)}{k},
\]

(A5)

\[
1 + \frac{\text{tr}(d\nabla \phi(M, zA_e - zA_t)A(\tilde{x}_1))}{\text{tr}(\nabla \phi(M)M)} = 1 - \frac{\mathbf{w}^*_d}{1 - \mathbf{w}^*_d} zk \left( 1 - \frac{\text{tr}(M^{-1}A_e M^{-1}A_e)}{k^2} \right).
\]

(A6)

Here, we have used that due to the properties of scalar products, the scalar product $\text{tr}(M^{-1}A_e M^{-1}A_e)$ will always be smaller or equal to $\text{tr}(M^{-1}A_e)\text{tr}(M^{-1}A_e)$, which is equal to $k\text{tr}(M^{-1}A_e)$. As $\frac{\text{tr}(M^{-1}A_e)}{k} \leq 1$, the differential in the direction of $A(\tilde{x}_1) = A_I$ will be larger or equal to the differential in the direction of $A(\tilde{x}_3)$. As $\frac{\mathbf{w}^*_d}{1 - \mathbf{w}^*_d} > 0$, it will also be larger than the differential in the direction of $A(\tilde{x}_2)$. The differential in the direction of $A(\tilde{x}_2)$ is thus the largest of the three, proving Equation (16).

Finally, if $\text{tr}(M^{-1}A_e M^{-1}A_e) = \text{tr}(M^{-1}f_e f_t^T M^{-1} f_t f_t^T) = (f_t^T M^{-1} f_t)^2 = k^2$, this implies that $|f_t^T M^{-1} f_e| = f_t^T M^{-1} f_t = f_t^T M^{-1} f_e = k$, and the properties of scalar products allow us to conclude that $A_e = A_I$ and that the resulting design is still optimal.

\[
\text{Proof of Corollary 3.4. As } \phi_{\text{Bayes}} = \exp \left( \frac{1}{k} \sum_{j=1}^m \beta_j \log |M(\mathbf{w}, \theta_j)| \right) \text{ is still a concave function, every step of the proof of Theorem 1 from Holland-Letz et al. (2012) can be recreated analogously, resulting in a lower limit given by}
\]

\[
eff(\mathbf{w}) \geq \left( \max_{i=1}^r \frac{1}{\sum_{j=1}^m \beta_j (\nabla \phi(M(\mathbf{w}, \theta_j))A(\tilde{x}_i, \theta_j))} \right)^{-1}
\]

(A7)
for a generic quasi-Bayesian optimality criterion $\phi$ and

$$\text{eff}(\mathbf{w}) \geq \left( \frac{\max_{i=1}^{r} \text{tr} \sum_{j=1}^{m} \beta_j (M(\mathbf{w}, \theta_j)^{-1} A(\bar{x}_i, \theta_j))}{k} \right)^{-1}$$  \hspace{1cm} (A8)

for quasi-Bayesian $D$-optimality.

Now, applying a similar reasoning as in the proof of Corollary 3.3 to this situation results in an equivalent to Equation (A5) given by

$$1 + \frac{\sum_{j=1}^{m} \beta_j \text{tr}(d^\top \nabla \phi(M(\theta_j), zA_e(\theta_j) - zA_t(\theta_j))A(\bar{x}_2, \theta_j))}{k}$$

$$= 1 + \frac{z}{k} \sum_{j=1}^{m} \beta_j \text{tr}(M(\mathbf{w}, \theta_j)^{-1}(A_t(\theta_j) - A_e(\theta_j))M(\mathbf{w}, \theta_j)^{-1}A_t(\theta_j))$$

$$= 1 + \frac{z}{k} \sum_{j=1}^{m} \beta_j \text{tr}(M(\mathbf{w}, \theta_j)^{-1}A_t(\theta_j)M(\mathbf{w}, \theta_j)^{-1}A_t(\theta_j))$$

$$- \frac{z}{k} \sum_{j=1}^{m} \beta_j \text{tr}(M(\mathbf{w}, \theta_j)^{-1}A_e(\theta_j)M(\mathbf{w}, \theta_j)^{-1}A_t(\theta_j)),$$  \hspace{1cm} (A9)

with analogous equivalents to Equations (A6) and (A4) resulting in the same term but with multipliers $-1$ and $-\frac{\mathbf{w}^\top_j}{1-\mathbf{w}^\top_j}$ before the second term after the first equality. Unfortunately, it is not obvious in this case, which of the three terms will be the largest. However, the term

$$1 + \frac{z}{k} \left| \sum_{j=1}^{m} \beta_j \text{tr}(M(\mathbf{w}, \theta_j)^{-1}(A_t(\theta_j) - A_e(\theta_j))M(\mathbf{w}, \theta_j)^{-1}A_t(\theta_j)) \right|$$  \hspace{1cm} (A10)

will supply an upper limit for all of them, proving the corollary.  \hspace{1cm} $\square$