Key properties of D-optimal designs for event-related functional MRI experiments with application to nonlinear models

Darcie A. P. Delzell, Richard F. Gunst, William R. Schucany, Patrick S. Carmack, Qihua Lin, Jeffrey S. Spence and Robert W. Haley

To properly formulate functional magnetic resonance imaging (fMRI) experiments with complex mental activity, it is advantageous to permit great flexibility in the statistical components of the design of these studies. The length of an experiment, the placement of various stimuli and the modeling approach used all affect the ability to detect mental activity. Major advances in understanding the implications of various designs of fMRI experiments have taken place over the last decade. Nevertheless, new and increasingly difficult issues relating to the modeling of hemodynamic responses and the detection of activated brain regions continue to arise because of the increasing complexity of the experiments. In this article, the D-optimality criterion is used in conjunction with a genetic algorithm to create probability-based design generators for the selection of designs in event-related fMRI experiments where the hemodynamic response function is modeled with a function that is nonlinear in the parameters. The designs produced by these generators are shown to perform well compared with locally D-optimal designs and provide insight into optimal design characteristics that investigators can utilize in the selection of interstimulus intervals. Designs with these characteristics are shown to be applicable to fMRI studies involving one or two stimulus types. The designs are also shown to be robust with respect to misspecification of an AR(1) error autocorrelation and compare favorably with a maximin procedure. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: hemodynamic response function (HRF); genetic algorithm; probabilistic experimental design; ISI selection; stimulus timing; brain imaging; maximin

1. Introduction

Functional magnetic resonance imaging (fMRI) [1–3] is a noninvasive brain imaging technique that produces a series of three-dimensional images acquired over time by analyzing the two-dimensional time series of brain activity recorded in individual locations in the brain. This allows for the observation of brain activation over a continuous time span and it is primarily utilized to identify areas of the brain associated with specific experimental tasks. Subjects are placed into a magnet and images are acquired by the MRI scanner while the subject is exposed to stimuli, such as flashing checkerboards, music, or speech, usually at predetermined time intervals.

[a]Department of Mathematics and Computer Science, Wheaton College, 501 College Ave., Wheaton, IL 60187, USA
[b]Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275-0332, USA
[c]Department of Mathematics, University of Central Arkansas, 201 Donaghey Avenue, Conway, AR 72035, USA
[d]Department of Clinical Science, Biostatistics Division, University of Texas, Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX, 75390-8830, USA
[e]Department of Internal Medicine, Epidemiology Division, University of Texas, Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX, 75390-8874, USA
[a*Correspondence to: Darcie A. P. Delzell, Department of Mathematics and Computer Science, Wheaton College, 501 College Ave., Wheaton, IL 60187, USA.
[e-mail: darcie.delzell@wheaton.edu]
An fMRI signal does not record neuronal activation directly, but instead reflects biological processes known to accompany neuronal activity. Blood oxygenation level dependent (BOLD) fMRI is an effective imaging modality for identifying the strength and extent of brain activation following a stimulus [1, Ch. 1]. Neural activity is accompanied by an increase in oxygenated hemoglobin relative to deoxygenated hemoglobin. Brain tissue with more oxygenated hemoglobin produces stronger BOLD fMRI signal because of increased brain activity in surrounding tissues.

The MRI scanner reads one BOLD signal value per voxel (volume element, approximately 3 × 4 × 3 mm³) at each sampled time point. This BOLD response for a voxel is often modeled using a hemodynamic response function (HRF). The HRF is an idealized BOLD signal change over time in response to a single stimulus. The general characterization of the hemodynamic response, which can differ across individuals and voxels within an individual [4, Ch. 3; 5–7], often begins with a delay of approximately 3 s after a given stimulus and then an increase to a peak at about 6 s [1, p. 251]. It then decays back to baseline, often dipping below the baseline before returning. In response to multiple stimuli, the BOLD signal in active parts of the brain is generally assumed to consist of a linear combination of the multiple responses to the stimuli, with each response to a stimulus being independent of previous responses. These assumptions lead to a linear time-invariant (LTI) HRF system model [8, 9] and the selection of stimulus times should be chosen to ensure the appropriateness of this assumption.

The present article was motivated by a need to consider various design questions posed by researchers performing a large number of fMRI experiments on veterans from the 1991 Persian Gulf War. The goal of the study was to characterize the symptoms of Gulf War Syndrome and aid in characterizing differences among syndrome and control groups to give insight into diagnosis and treatment protocols. Some of these experiments were aimed at extracting and comparing specific features of the HRF using nonlinear models. One reason for this is because nonlinear models allow variation in the HRF shape other than scale (as in the case of linear models), such as the timing of the maximum and minimum responses, overall duration, etc. Because an HRF is estimated at each voxel in a subject, differences in HRF responses across the brain can be captured with more specificity if a nonlinear model is used for the HRF response. The investigators were interested in overall guidelines for fMRI experimental designs that would yield more accurate parameter estimates for the HRF. The use of nonlinear models with multiple parameters is not without difficulty. Models of this complexity can fail to converge at some voxels [10]. However, these models continue to be investigated and optimal use of these models in the fMRI setting is of interest [11, 12].

Various experiments in the study necessitated the use of event-related (ER) designs, which are designs where stimuli are presented with short durations, usually a few seconds apart. These designs are essentially block designs with extremely short block lengths [1, Ch. 9]. There are a number of factors that can necessitate the use of an ER design, including the nature of the stimuli presented and experiments that cannot adhere to the framework for which pre-experiment interstimulus intervals (ISIs, the time from the initiation of one stimulus to the initiation of the next stimulus) are specified [13–15]. In part, the analyses reported in this article are intended to inform investigators of key properties of better designs that can be used to select ISIs in real time as scanning proceeds.

The essential components of an ER design are the times at which stimuli are presented during the course of an experiment. Numerous articles have focused on the choice of stimulus times or ISIs in an effort to improve estimation of the HRF. When using a constant ISI, the optimal interval was found to be in the range of 12–14 s [16]. Shorter ISIs (on the order of 2 s) were found to be optimal for designs with variable ISIs [17, 18]. Dale [18] suggested that better HRF estimation would be obtained by alternating between task and rest states as quickly as possible. There are concerns, however, that using ISIs that are too short can violate the LTI system assumptions necessary for the convolution HRF model [19–21] that is commonly fit. Altering the timing of stimulus presentation in relation to the timing of the scanner’s acquisition phase has been recommended to sample the HRF at multiple points [22, 23] that ensure accurate estimates of the key features of the HRF (e.g., the time to peak activation). These studies all assume a linear model for the HRF and the optimality criteria used in these studies are model-dependent because the criteria are related to the variance structure of the HRF parameter estimators.

While questions of design for fMRI experiments have been considered extensively for general linear models (GLMs) in which only scale parameters (additive and multiplicative regression parameters) are to be estimated, design issues have not been substantially addressed in the context of nonlinear models. The present article investigates the use of the D-optimality design criterion for the selection of ISIs in event-related fMRI experiments. One challenge in the selection of ISIs is the vast size of the design space, because stimuli can potentially be presented at very short intervals. One way to select designs...
for consideration is to use a genetic algorithmic approach [21]. Genetic algorithms are search engines that identify new designs from previous designs, such as using two parent designs to create a potentially more optimal child design. The Kao et al. [24] version of Wager and Nichols’ [21] algorithm was further modified for the nonlinear application of genetic algorithms used in this work.

The characteristics of many locally D-optimal designs are used to propose a discrete probability distribution for the ISIs. This distribution reflects the rates of prevalence of various ISIs among the locally D-optimal designs. The use of probability distributions for ISI selection has been investigated in the GLM setting in various ways. Burock et al. [25] varied the ISI according to an exponential distribution. Dale [18] and Birn [17] varied ISIs and compared efficiencies based on mean ISI. Friston et al. [20] placed probabilities on each sampled time point and used a Bernoulli process to determine whether or not a stimulus was applied at that time point. Dale [18], Birn [17] and Friston [20] compared designs by using the trace of the estimator covariance matrix, known as the A-optimality criterion.

A recent article by Maus et al. [11] employs the D-optimality criterion for choosing an optimal design when estimating a similar nonlinear model as the one used in this work, but with tighter restrictions on the range of possible parameter values. In addition, Maus et al. [11] limited their investigations to a single stimulus type and used a maximin criterion to identify a single D-optimal design given a minimum ISI of 2.5 s and restricting other ISIs to be multiples of 2.5. Kao et al. [12] also consider a similar nonlinear model with fewer freely varying parameters and use the maximin criterion on a related criteria known as A-optimality.

The primary purpose of this paper is to provide guidelines to experimenters for designs with desirable statistical properties for ER-fMRI experiments when an aim is to estimate the HRF with a nonlinear model using BOLD data. The efficiency of the identified design characteristics (i.e., the proposed ISI distributions) for the nonlinear model is investigated for both Gaussian white noise and AR(1) error structures when the errors have been both correctly and incorrectly specified. Applicability to estimating a linear model containing a commonly used canonical HRF and its partial derivative as covariates is examined. Investigations of optimal design characteristics for one and two stimulus types are presented. In addition, a comparison to the maximin criterion utilized by Maus et al. [11] is made.

2. Theory

2.1. Model specification

The HRF portion of the model used in this paper was chosen to allow the flexibility to model and identify the potential variation of the HRF response from person to person and voxel to voxel because none of the model parameters are prespecified as fixed in advance of the modeling. Although the whole brain fMRI image is three-dimensional, the HRF form is estimated for each voxel separately using the time series model parameters are prespecified as fixed in advance of the modeling. Although the whole brain fMRI

...
waveform, $s(t)$, and an impulse response function, $h(t)$, the convolution of the two gives the output waveform. In the continuous case the output waveform, $f(t)$, is written as

$$f(t) = h(t) \otimes s(t) = \int_{-\infty}^{\infty} h(u)s(t-u)du,$$  

(2)

where $\otimes$ denotes the convolution operator. In the case of fMRI experiments where the BOLD signal is sampled at discrete time points, the convolution model for the time series at a time $t_i$ for a single voxel with signal can be written as

$$y(t_i) = f(t_i; \theta) + e_i = \sum_{j=1}^{J} h(t_i - g_j; \theta) + e_i,$$  

(3)

where $h$ is the HRF defined in (1) and the $e_i$ are the random errors, which are assumed to be normally distributed with mean 0 and covariance matrix equal to $\sigma^2 \Sigma$ [2, Ch. 14].

Autoregressive error processes are often used in fMRI data modeling, as much evidence has been provided to show that temporal correlation is generally present in the data, even after drift components have been removed [26–29]. In this work, independent errors are investigated and those with an autoregressive component. The sampled time points, $t_i$, are determined by the fMRI scanner’s time to repeat parameter ($TR$, the time between acquisition of successive brain volumes), usually in the range of 1–3 s. The vector $G = \{g_1, \ldots, g_J\}$ is defined as the stimulus vector where the $g_j$ are the times at which stimuli are applied during the experiment. The intervals between the elements of $G$ are the ISIs used in the experiment. In this article, $TR = 2$ s and the duration of the scanning session is 300 s.

2.2. Optimal design criteria and the parameter space

Results from nonlinear least-squares theory demonstrate that the D-optimality criterion minimizes the asymptotic confidence region volume for $\theta$ by minimizing the determinant of the asymptotic covariance matrix of the parameter estimators [30, Section 5.13; 31]. Using a first-order Taylor series approximation to $f(t_i; \theta)$ in (3), the asymptotic covariance matrix is proportional to $[F(\hat{\theta})^T \Sigma^{-1} F(\hat{\theta})]^{-1}$, where $F(\hat{\theta})$ is the matrix of first partial derivatives of the model function evaluated at $\hat{\theta}$:

$$F(\hat{\theta}) = \left[ \frac{\partial f(t_i; \theta)}{\partial \theta^c} \right]_{\theta=\hat{\theta}} = \left[ \frac{\partial}{\partial \theta^c} \left( \sum_{j=1}^{J} h(t_i - g_j; \theta) \right) \right]_{\theta=\hat{\theta}} = \left[ \sum_{j=1}^{J} \frac{\partial h(t_i - g_j; \theta)}{\partial \theta^c} \right]_{\theta=\hat{\theta}}$$  

(4)

for the model in (3) [31, Section 6.2]. Here, $r$ and $c$ refer to row and column, respectively. $F(\hat{\theta})$ is an $n \times p$ matrix where $n$ is the number of time points in the experiment and $p$ is the dimensionality of the parameter vector $\theta$. In this paper we consider $n = 300/2 = 150$ and $p = 5$ for a single stimulus type experiment.

Minimizing confidence region volume is more complicated in the nonlinear setting than the canonical HRF setting because the asymptotic covariance matrix is a function of the estimated parameter vector $\hat{\theta}$. At the experimental design stage, $\hat{\theta}$ has not yet been obtained. Box and Lucas [30] suggested using preliminary estimates for $\theta$ taken from either previous work or other available information. An optimal design in the nonlinear regression setting that is based on a prespecified value of $\theta$ is considered ‘locally optimal’ [32]. The D-optimality criterion for a stimulus vector $G$ at a particular $p$-dimensional $\theta$ is

$$Dopt(G; \theta) = \frac{p}{\sqrt{|F(\hat{\theta})^T \Sigma^{-1} F(\hat{\theta})|}}.$$  

(5)

Smaller values of $Dopt(G; \theta)$ correspond to smaller confidence ellipsoids. We use scaling by the $p^{th}$ root because comparison of design efficiencies (Section 3.3) results in efficiencies that can be interpreted similar to relative variances regardless of the number of parameters in the model [33, p. 151].

A set of 400 feasible parameter vectors for $\theta$ in (1) were chosen to represent a large selection of physically reasonable 2-Gamma HRF functions [8,22,34–37]. Ranges for $d_1$ have been observed to be as low
as 3 s and as high as 10 s. For \( d_2 \), values ranged from 10 s to 18–19 s. The parameter \( c_2 \) is a scale parameter representing the magnitude (depth) of the secondary lobe of the 2-Gamma HRF, as a proportion of the magnitude of the primary lobe \((c_1)\). Values for \( c_2 \) range from 0.14 to 0.55. Glover’s [38] values for \( \alpha_1 \) and \( \alpha_2 \) are used with a few additional combinations. We let the set of vectors be denoted by \( \textbf{P} \), where

\[
\textbf{P} = \{ \theta = (\alpha_1, \alpha_2, d_1, d_2, c_2) \ |
\begin{align*}
\{ (\alpha_1, \alpha_2) \} &\in \{ (12, 20), (12, 28), (20, 20), (20, 28), (20, 35) \}, \\
\{ (d_1, d_2) \} &\in \{ (d_1 = 3, d_2 = 8), (d_1 = 5 \text{ or } 5.5, d_2 = (10, 11.5, 12, \text{ or } 13)), \\
& (d_1 = 6, d_2 = (11.5, 12, \text{ or } 13)), \\
& (d_1 = 8, d_2 = (13 \text{ or } 16)), \\
& (d_1 = 10, d_2 = (16, \text{ or } 18)) \}, \\
\text{and} &\, c_2 \in \{.1, .2, .3, .4, .5\}.
\}
\]

(6)

Figure 1 displays the theoretical HRF curves that are included in the parameter space (6). The broadness of the parameter space is desirable because it represents both normative HRFs and potentially non-normative HRFs because of illness as might characterize syndrome subjects included in the Gulf War Syndrome study.

3. Methods

3.1. Basic methodology

For a given value of \( \theta \in \textbf{P} \), a locally D-optimal design was found using the genetic algorithm of Kao et al. [24] assuming model (3). The genetic algorithm requires an initial collection of potential designs. The initial designs consisted of \( m \)-sequence-based designs and randomly generated designs [24, 39]. \( M \)-sequences are algorithmically created using linear shift registers to produce designs that have an approximately equal number of stimulus types (including no stimulus) and are almost orthogonal to time-shifted versions of themselves [39]. The created designs were modified to adhere to the minimum 4 s LTI requirement. The randomly generated designs were created by allowing a 50% probability of a stimulus being applied at each second, again with the restriction that stimuli must be placed at least 4 s apart. The maximum ISI allowed was 30 s [19, 20, 34, 40].

The genetic algorithm modifies designs in a multistage process by producing new designs that are, generally, more optimal than the designs at the previous stage (the optimality criterion is D-optimality in this implementation of the algorithm). Given a sufficient number of iterations, this algorithm converges to a locally optimal design. The algorithm has a few user-specified inputs: \( S \) (size of population) = 20, \( q \) (percentage of mutation) = 1%, \( I \) (number of immigrants) = 4 and \( M_g \) (number of generations; i.e., design iterations) = 10,000. Because of random mutation and a large number of iterations, this algorithm is fairly robust to the choice of starting designs. For all values of \( \theta \in \textbf{P} \), by the 7000th iteration the change in the D-optimal value was less than 0.5%. \( M_g = 10,000 \) was also considered sufficiently large to ensure convergence of the algorithm in both [24] and [11].
To assess which ISIs were most prevalent for a given locally D-optimal design, the proportion of time each ISI occurred (relative to the total number of ISIs for that design) was computed. Each locally D-optimal design produced 27 proportions corresponding to ISIs from 4 to 30 s. Thus, for a given ISI the mean of the proportions of that ISI across the 400 locally D-optimal designs was computed. Finally, the 26 mean proportions were normalized to sum to 1. This resulted in a discrete probability distribution of ISIs. This probability distribution (denoted \( PG \)) represents the optimal design characteristics (in terms of the distribution of ISIs) observed across \( P \) from the 400 locally D-optimal designs. The normalized medians were also calculated, but the resulting probability distributions were almost identical to those obtained when using the normalized means. The \( PG \) distribution represents an ‘average’ of all of the locally D-optimal design characteristics.

To create an actual fMRI stimulus vector, \( G \), the first stimulus is assumed to begin at \( g_1 = 0 \). An ISI is then randomly generated from the \( PG \) distribution and \( g_2 = g_1 + \text{ISI} \). This is repeated (with \( g_{j+1} = g_J + \text{ISI} \)) until the experiment time length is exceeded and the last stimulus time is discarded. It may be that the locally D-optimal design for \( \theta_1 \in P \) does not perform well for \( \theta_2 \in P \). Taking an aggregate of the optimal design characteristics across \( P \) (via the \( PG \) distribution) may produce designs that work reasonably well for both \( \theta_1 \) and \( \theta_2 \). This would be desirable because the true value of \( \theta \) is unknown at the experimental design stage. An investigation into the performance of randomly generated designs from \( PG \) distributions (henceforth referred to as \( PG \) designs) relative to the locally D-optimal designs is described in Section 3.3.

### 3.2. Three distributions

Model (3) assumes normally distributed additive error. In this work independent errors are considered and AR(1) errors with autocorrelation parameters of either \( \rho = 0.3 \) or \( \rho = 0.5 \). These values of \( \rho \) were informed by considering auditory fMRI data from active voxels taken from a pilot study from the Gulf War Syndrome project. These voxels had AR(1) estimated autocorrelation coefficients ranging from approximately 0.14 to 0.67. The methodology of Section 3.1 was used to create three separate ISI distributions, denoted \( PG^\rho \) for \( \rho \in \{ 0, 0.3, 0.5 \} \) (recall the value of \( \rho \) changes the D-optimality calculation in (5)). This was carried out to investigate the potential variability of optimal design characteristics based on error structure. In addition, because the true error structure of fMRI data is unknown at the design stage, multiple \( PG^\rho \) distributions allowed for investigation of the performance of the \( PG^\rho \) designs in the presence of error misspecification (Section 3.5).

### 3.3. Efficiencies relative to the locally D-optimal designs

The locally D-optimal design for each \( \theta \in P \) for a given value of \( \rho \in \{ 0, 0.3, 0.5 \} \) is denoted by \( G_{\theta, \rho}^* \). The D-optimal value for each \( G_{\theta, \rho}^* \) at \( \theta \) is denoted by \( Dopt(G_{\theta, \rho}^*) \) as defined in (5). For each \( \rho \), 100 designs were generated from the \( PG^\rho \) distribution. The D-optimal value was then calculated for each of the 100 \( PG^\rho \) designs for each of the 400 \( \theta \in P \). Denote the 100 D-optimal values for a particular \( \theta \) as \( Dopt_j(\rho; \theta) \), \( j \in \{ 1, 2, \ldots, 100 \} \).

The efficiency of the 100 \( PG^\rho \) designs relative to the locally D-optimal design for a particular \( \theta \) is computed using the efficiency ratio

\[
DE_{\theta, \rho}^j = \frac{Dopt(G_{\theta, \rho}^*)}{Dopt_j(PG^\rho; \theta)}.
\]  

This is a relative efficiency measure that takes on values between 0 and 1 and provides a measure of the performance of the \( PG^\rho \) designs relative to the locally D-optimal designs (\( G_{\theta, \rho}^* \)). Values close to 0 imply that the particular generated design does not perform well compared with the locally D-optimal design (for a particular combination of \( \rho \) and \( \theta \) ). Values close to 1 imply that the generated design performs almost as well as the locally D-optimal design. For comparison, the efficiencies (relative to the \( G_{\theta, \rho}^* \) designs) of the initially used \( m \)-sequence-based designs were also calculated. Other efficiency measures are defined below in a similar manner.

### 3.4. Interstimulus intervals that are multiples of the TR

There is concern that only allowing ISIs that are a multiple of the specified TR can decrease the efficiency of HRF estimation. This concern has been investigated for linear models [22, 23] but not for a
nonlinear model. The importance of having some ISIs in the design that are not a multiple of the TR was investigated by generating 100 designs from a slightly modified version of the PG$^0.3$ distribution in which only even ISIs were given positive probability. This distribution was formed using the methods of Section 3.1 with the exception that only ISIs equal to a multiple of TR=2 s were allowed (unlike the PG$^\rho$ distributions where both even and odd ISIs were allowed). The efficiencies of these designs relative to the locally D-optimal designs when $\rho = 0.3$ were computed in a manner similar to (7).

3.5. Error and model misspecification

An investigation of the robustness of the designs generated by the PG$^\rho$ distributions against misspecification of the error autocorrelation in the creation of these distributions was performed. For example, the robustness of the PG$^0$ designs against misspecification of $\rho$ was examined by calculating the D-optimal value of the PG$^0$ designs at the correct value of $\rho$ in (5) (e.g., $\rho = 0.3$). The efficiencies of the PG$^0$ designs were then computed relative to the correct locally D-optimal designs obtained assuming $\rho = 0.3$ in the genetic algorithm. These efficiencies provide information on how well the PG$^0$ designs perform relative to the locally D-optimal designs when the true errors are correlated (AR(1) with $\rho = 0.3$). Efficiencies for all possible misspecifications were obtained.

A separate investigation was conducted to assess the performance of the PG$^\rho$ designs when a canonical HRF (and hence linear model) is assumed. In this case, the canonical HRF and its partial derivative with respect to $d$ were included as basis functions (covariates) in the design matrix $X$ for the ordinary linear model [36, 40, 41]. The values of all the parameters in (1), $\theta = (6, 12, 5.4, 10.8, 1, .35)$, were obtained from Glover [38] and represent a commonly used HRF form. When a linear model is assumed, the D-optimality criterion is not dependent on $\theta$ and is defined as

$$Dopt(G) = \rho \sqrt{X^T \Sigma^{-1} X}^{-1},$$

where $X$ is the design matrix in the model $y = X\beta + e$. Here, $\beta$ is a two-dimensional vector of scale parameters to be estimated, and the members of the error vector, $e$, are normally distributed with zero mean and covariance matrix equal to $\sigma^2 \Sigma$ [42]. The matrix $X$ is created using the convolution of the stimulus vector $G$ and the two discretized basis functions (the canonical HRF and its partial derivative with respect to $d_1$) [2, Ch. 40]. In this case the D-optimal design does not depend on $\theta$, and is denoted by Lin$^*$, and the efficiencies of the PG$^\rho$ designs relative to Lin$^*$ are calculated using

$$LinDE_j = \frac{Dopt(Lin^*)}{Dopt_j(PG^\rho)}, j \in \{1, 2, \ldots, 100\}. \quad (9)$$

3.6. Comparison to the maximin procedure

A comparison was performed to assess the adequacy of the PG$^\rho$ designs relative to the maximin designs produced by the methods of Maus et al. [11]. Let the set of 400 locally D-optimal designs assuming model (3) for a given value of $\rho$ be denoted as $G_\rho = \{G^*_\rho | G^*_\rho = G^*_{\theta,\rho} \text{ for } \theta \in P\}$. Then the maximin design is

$$\text{MMD}_\rho = \arg \max_{G^*_\rho \in G_\rho} \min_{\theta \in P} \left( \frac{Dopt(G^*_\rho)}{Dopt_j(PG^\rho; \theta)} \right) \text{ for } \rho \in \{0, 0.3, 0.5\}. \quad (10)$$

For a given value of $\rho$, the relative efficiency of each set of 100 PG$^\rho$ designs relative to the MMD$\rho$ design was computed across $P$, for the various error correlations assumed in this work, via

$$\text{MMDE}_\rho^j (\rho, \rho') = \frac{Dopt(\text{MMD}_\rho; \theta)}{Dopt_j(PG^\rho; \theta)}, \theta \in P, j \in \{1, 2, \ldots, 100\}, \quad (11)$$

where $\Sigma$ is determined by the value of $\rho' \in \{0, 0.3, 0.5\}$. This results in nine comparisons that give insight into the performance of the PG$^\rho$ designs relative to the maximin design, MMD$\rho$, when the autocorrelation parameter has been both correctly (when $\rho = \rho'$) and incorrectly (when $\rho \neq \rho'$) specified in the determination of the maximin design and in the creation of the PG distribution.
3.7. A two-stimulus type investigation

One of the experiments performed in the Gulf War Syndrome study used a paradigm known as a Semantic Object Retrieval Test and is an ER paradigm by its nature [43]. It had two trial types with a minimum ISI of approximately 4 s between stimuli of either type. For experiments with two stimulus types it is assumed that there are two different HRF responses. In this case there are 10 parameters estimated by nonlinear least squares. There are two different stimulus vectors, each representing the stimulus types it is assumed that there are two different HRF responses. In this case there are 10 parameters estimated by nonlinear least squares. There are two different stimulus vectors, each representing the stimulus application times for their respective stimulus types. As before, the D-optimality criterion must be evaluated at specific parameter values.

The model specifics for the two-stimulus type model at a time, \( t \), are

\[
y(t_i) = f(t_i; \theta_1, \theta_2) + e_i = \sum_{j=1}^{J} h(t_i - f_j; \theta_1) + \sum_{k=1}^{K} h(t_i - m_k; \theta_2) + e_i, \tag{12}
\]

where \( \theta_1 \) and \( \theta_2 \) are both members of \( \mathbf{P} \) and the random errors are normally distributed with mean 0 and covariance matrix \( \sigma^2 \mathbf{\Sigma} \) (cf. (3)). Here, \( \mathbf{G} = \{g_1, g_2, \ldots, g_J\} \) is the stimulus vector for the first stimulus type, \( \mathbf{M} = \{m_1, m_2, \ldots, m_K\} \) is the stimulus vector for the second stimulus type, and \( \theta = (\theta_1, \theta_2) \).

For this model \( \mathbf{F}^i(\hat{\theta}) = \left[ \frac{\partial f(t_i; \theta)}{\partial \theta} \right]_{\theta=\hat{\theta}} \) is a 150 × 10 matrix (c.f. (4)) and the asymptotic covariance matrix is \( \left[ \mathbf{F}^i(\hat{\theta}) \mathbf{\Sigma}^{-1} \mathbf{F}^i(\hat{\theta}) \right]^{-1} \). Evaluating this matrix requires a pair of HRF parameter vectors to be defined (i.e., \( \theta_1 \) and \( \theta_2 \)).

For each possible pair of \( \theta_1 \) and \( \theta_2 \) in \( \mathbf{P} \), a measure of the difference between the two resulting HRF curves was evaluated by computing the area between the curves. Smaller areas signify HRF curves that are similar to each other while larger areas signify HRF curves that are disparate. A set of 100 pairs of \( \theta_1 \) and \( \theta_2 \) were created that represented the range of computed differences in areas. A genetic algorithm was then used to identify the locally D-optimal design at each pair of HRF parameter vectors. This was carried out for \( \rho = 0, 0.3, \) and 0.5 and the proportions of ISIs for the two stimulus vectors, \( \mathbf{G} \) and \( \mathbf{M} \), were computed within each of the 100 locally D-optimal designs to examine the optimal design characteristics.

4. Results

4.1. Efficiencies relative to the locally D-optimal designs

Figures 2(a) and (b) show boxplots of the proportions of each ISI in the 400 locally D-optimal designs for the error correlations of \( \rho = 0 \) and 0.3 (the figure for \( \rho = 0.5 \) is similar to that for \( \rho = 0.3 \) and is included in the online supplement). ISIs greater than 16 s are not plotted because they did not occur in any locally D-optimal designs. When \( \rho = 0 \), the 4 s ISI occurs the most often among the 400 locally D-optimal designs, with proportions of approximately 35%--65%. The ISIs of 6--9 s occur (approximately) in the range of 5%--20% of the ISIs in each of the D-optimal designs.

The normalized means of these proportions (red bullets overlaying each boxplot) are the probabilities obtained for the \( PG^\rho \) probability distributions (probabilities less than 0.05 were set to 0). Designs generated from the \( PG^\rho \) distributions have, on average, about half of the ISIs equal to 4 s (i.e., the minimum ISI allowed) and the other half from the range of 5--9 s, proportional to the probabilities shown in Figure 2. The notable difference between the \( PG^0 \) and \( PG^{0.3} \) (and \( PG^{0.5} \)) distributions is that the maximum ISI is 8 s for the latter distribution and it has slightly higher proportions of ISIs of 5–7 s than \( PG^0 \). Overall, however, the patterns in the probability distributions are quite similar for \( \rho = 0, 0.3, \) and 0.5.

Boxplots of the efficiencies (\( DE^i_{\theta, \rho} \) values) of the 100 designs generated by the \( PG^{0.3} \) probability distribution are shown in Figure 2(c). The efficiencies for the designs generated by the \( PG^0 \) and \( PG^{0.5} \) distributions (see online supplement ) are very similar to those of the \( PG^{0.3} \) distribution. Each of the 80 boxplots represents a value of \( \theta \in \mathbf{P} \) for \( c_2 = 0.3 \). The value of \( c_2 \) had very little impact on the efficiency values for all design comparisons, so only \( c_2 = 0.3 \) efficiencies are shown in Figure 2(c) for clarity. The red bullets show the efficiencies of the \( m \)-sequence-based designs used for the initial designs in the genetic algorithm at each \( \theta \in \mathbf{P} \). The \( PG^{0.3} \) designs have, overall, higher efficiency values than the \( m \)-sequence-based designs and are consistently in the range of approximately 0.80--0.95, regardless of the value of \( \theta \) or \( \rho \). Parameters with boxplot medians less than 0.80 are in two groups and are highlighted.
in yellow. The first are those parameters with the smallest times to the HRF maximum (time to peak) and minimum (time to peak undershoot) in (6), $d_1 = 3$ and $d_2 = 8$ s. For these parameters the locally D-optimal designs have fewer ISIs of 4 s than the PG$^0$ designs, on the order of 40% per design. The second group are those parameters with $d_1 = 10$ and $d_2 = 16$ or 18 s, the highest feasible times to maximum and minimum in (6). For these parameters the locally D-optimal designs have many ISIs of 10 s or more. The maximum ISI of any of the probability distributions generating the PG$^0$ designs is 9 s. The PG$^0$ designs perform less optimally at the aforementioned parameters because they are on the extreme edge of the parameter space.

4.2. Interstimulus intervals that are multiples of the TR

The efficiency values of the modified PG$^{0,3}$ designs (only even ISIs, multiples of TR=2 s, allowed) relative to the locally D-optimal designs are shown in Figure 3. Approximately 75% of the design and parameter combinations produced efficiency values less than 0.80, with 44% having values less than 0.70. Designs with only even ISIs do not perform and those with both odd and even ISIs.

4.3. Error and model misspecification

Figure 4 shows boxplots of the efficiencies of the PG$^0$ designs when the error component is misspecified (i.e., when the true value of $\rho$ is 0.3). Parameters with boxplot medians less than 0.80 are again highlighted in yellow. These occur at parameters with the same characteristics as those in yellow in Figure 2(c). Comparable figures of the efficiencies for all three sets of PG$^\rho$ designs when the error component is misspecified are included in the online supplement.

All of these efficiencies are very consistent with those in Figure 4. All these misspecification efficiencies are also very similar to the efficiencies when $\rho$ is correctly specified. Typically the loss in efficiency when an incorrect value of $\rho$ was used to construct the PG distributions was approximately 0.01 across $\rho$. The greatest average loss in efficiency when the correlation was misspecified was 0.04. It occurred for the PG$^{0.5}$ designs when the correct value of $\rho$ was 0.
Figure 3. Efficiencies of the modified $PG^{0.3}$ designs relative to the locally D-optimal designs.

Figure 4. Efficiencies ($DE_j$) of the $PG^0$ designs relative to the locally D-optimal designs when $\rho = 0.3$. Yellow boxplots have medians less than 0.80.

Using the $PG^{0.3}$ designs for estimation of the HRF using the canonical HRF and a partial derivative yielded efficiency ($LinDE_j$) values between 0.71 and 0.88 for all 100 $PG^{0.3}$ designs. The values for $PG$ designs with $\rho = 0$ and 0.5 are comparable.

4.4. Comparison to the maximin procedure

Figures 5(a)–(c) show boxplots of the efficiencies of $PG^{0.3}$ relative to $MMD_{0.3}$ when the D-optimal values are calculated assuming $\rho = 0, 0.3,$ and 0.5, respectively. In other words, these are $MMDE_\theta^{0.3}(0.3, 0), MMDE_\theta^{0.3}(0.3, 0.3),$ and $MMDE_\theta^{0.3}(0.3, 0.5).$ The efficiency values are high regardless of whether the autocorrelation parameter is correctly specified. Approximately 90% of the efficiency values are greater than 0.90, with approximately 25% greater than 1.0 (across $\rho$). The efficiency values greater than 1 indicate that at those parameter values the particular $PG^{0.3}$ design performs better than the maximin design. Results are similar for values of $\rho = 0$ and 0.5 (figures included in the online supplement).

4.5. A two-stimulus type investigation

Figures 6(a) and (b) show boxplots of the proportions of each ISI (ranging from 4–30 s) of the 100 locally D-optimal designs for designs with two stimulus types when $\rho = 0.3$. The characteristics of the two figures are similar to one another across the 100 pairs of parameter vectors for the two stimuli. ISIs of 4 s occur with approximate proportion 0.35 relative to the total number of ISIs in a design, for both stimulus types. In addition, there are many instances of long ISIs in between stimuli of the same type, in the range of approximately 6–20 s. This is because a stimulus type can be repeated, resulting in a
Figure 5. Efficiencies ($MMDE_\theta(0.3, \rho')$) of the $PG^{0.3}$ designs relative to the maximin design, $MMD_{0.3}$ when (a) $\rho' = 0$, (b) $\rho' = 0.3$, and (c) $\rho' = 0.5$.

Figure 6. Boxplots of the prevalence of ISIs among the locally D-optimal designs when considering 2 stimulus types ($\rho = 0.3$), (a) and (b), respectively.

long ISI for the other stimulus type. This resulted in approximately 70% of the ISIs between stimulus applications (irrespective of type) being 4 s in length with the other 30% in the range of 5–10 s. This is similar to the recommendations for a single trial type.
Nonparametric runs tests performed on the locally D-optimal designs suggested that the two stimuli were applied randomly during the course of the experiment. In addition, the average number of occurrences per trial type were approximately equal (31.63 and 30.85) indicating that good designs will have approximately half of the stimuli from each type. Almost all of the 100 locally D-optimal designs contained both odd and even ISIs, showing the importance of having some ISIs that are not multiples of the TR. These results suggest that for more than two stimulus types, good designs should have an equal distribution of each type in the experiment, their sequence should be randomly assigned, and the majority of intervals between stimuli (irrespective of type) should be at the minimum allowed with a smaller percentage of ISIs in the range of 5–10 s. The boxplots for \( \rho = 0 \) and 0.5 are similar and are included in the online supplement.

5. Discussion

A number of recent articles have proposed and implemented new statistically based algorithms using optimal design theory to identify ISIs for fMRI experiments. Although computer software is available to implement these algorithms for specific purposes, the complexity of the background theory and algorithmic components of the software can inhibit investigators from applying these methods to their specific needs, especially for smaller fMRI studies where statistical and technical expertise is not easily accessible. In this article, general properties of the most effective designs for fitting 2-Gamma hemodynamic response functions are identified to provide guidelines for experimenters who (i) have special goals that preclude the execution of these algorithms without substantial programming code modifications or (ii) want to select from among many designs that possess good statistical properties.

The D-optimality criterion was used in conjunction with the genetic algorithm of Kao et al. [24] to investigate optimal design characteristics that are applicable to event-related fMRI studies when the goal is to fit a nonlinear HRF. ISI distributions were created for various error structures that generated fMRI designs possessing these properties. These distributions are very similar for different error structures and create designs that perform very well (i.e., high relative efficiencies) relative to the locally D-optimal designs.

The results in this article lead to general design characteristics that are not critically dependent on correct identification of the AR(1) error structure or knowledge of the particular HRF parameter vector. ER-fMRI experiments should contain a high percentage of ISIs at the minimum value permitted by the nature of the experiment; in this work, ISIs of 4 s with frequency of approximately 50%. The second most prevalent ISI should be approximately twice the minimum allowable without violating the LTI assumption, here 8 s (on the order of 20% of the total ISIs in the experiment), with the rest falling in the range of 5–9 s. The vast majority of locally D-optimal designs (c.f. Figures 2(a) and (b)) contained ISIs of both 7 and 8 s, which indicates that the presence of a reasonable range (4–9 s) of both odd and even ISIs in the design is an important factor for optimality.

In addition, it was shown that designs that only include ISIs that are multiples of the TR do not perform well relative to locally D-optimal designs (c.f. Figure 3). This is in agreement with Josephs et al. [22] and Price et al. [23]. It has also been recommended in the case of ER designs and GLMs that shorter average ISIs are optimal [17, 18, 20, 44], and our results are in agreement with this conclusion to the degree that most of the ISIs should occur at the minimum allowable without violation of the LTI assumption. However, our results imply that to model the entire shape of the HRF it is useful to have at least some longer ISIs. This is a difference because of the use of a nonlinear HRF model and the D-optimality criterion and is in agreement with Maus et al. [11]. This provides the ability for consecutive HRFs to be observed without as much overlap as is present with all ISIs very short.

The maximin design discussed by Maus et al. [11] shares the characteristic with the \( PG^k \) designs that approximately half of the ISIs occur at the minimum. However, approximately 20% of the ISIs in the maximin design are greater than 7.5 s. The reason for this is that only ISIs that are a multiple of the minimum of 2.5 s are used in their algorithm. In this work ISIs are allowed to have values at every integer from 4–30 s and it was shown that by using all possible integer ISIs it is not necessary to have ISIs longer than 9 s. This in turn allows more stimuli to be presented.

It was also shown that the \( PG \) designs perform very well across the parameter space when compared with the maximin design obtained by the approach of Maus et al. [11], and for some elements of the parameter space they outperform the maximin design. This is because the maximin design is the locally D-optimal design at a particular element of the parameter space, but does not necessarily outperform the \( PG \) designs across the parameter space.

When using a GLM with a canonical HRF and its partial derivative, there is a loss of overall efficiency when using the PG designs. However, the efficiencies are surprisingly high considering that the PG designs were created using optimal characteristics based on estimating five parameters of a nonlinear HRF model. If two stimulus types are used, along with two different nonlinear HRF functions, it was shown that optimal design characteristics are very similar to those of the single trial models, with a few exceptions noted in Section 4.5.

Finally, this work should not be construed as advocating the general characteristics identified for desirable fMRI designs as applicable in all circumstances. Specific design goals require an assessment of alternative designs and design criteria. Nevertheless, the results reported above provide guidance to experimenters whose ultimate goals include fitting nonlinear models to data collected in fMRI experiments.

Acknowledgements

This study was supported by VA IDIQ contract number VA549-P-0027 awarded and administered by the Dallas, TX VA Medical Center. The content of this paper does not necessarily reflect the position or the policy of the US government, and no official endorsement should be inferred. The authors would also like to thank Dr. Sergey Cheshkov for insightful comments on this work and Neile Havens and Dr. Cary Gray for their valuable computational assistance. The authors also wish to express their sincere appreciation to the reviewers for their detailed and helpful comments on earlier drafts of this article.

References


