COMPUTATIONAL SCIENCE | RESEARCH ARTICLE

On prediction error variance to determining optimal design for two variable quadratic logistic model

F. B. Adebola¹, O. A. Fasoranbaku¹ and J. A. Kupolusi¹*

Abstract: Optimal design of experiment for logistic models has been examined and applied in a wide range of applications. The optimality of the designs is mostly determined by using general equivalence theorem with no attention paid to the extent at which the design can be useful for determining the predictive capability of the model. This paper addressed the predictive capability of optimal design model for two variable quadratic logistic regression model through prediction error variance(PEV). The PEV is a useful way to determining the predictive capability of a model in optimal design. The study used some initial guess parameters to represent any position of parameter in the design space through a simulation study of 10000 experimental runs. The design was optimal when the PEV value is less than one at nine equally weighted support points. The result of the analysis was able to identify the design that is good for prediction among all the designs obtained and conclude that prediction error variance should be used to test the stability of optimal design of experiment for two variable quadratic logistic models.

Subjects: Applied Mathematics; Statistics & Probability; Computation; Databases

Keywords: prediction error variance (PEV); general equivalence theorem; generalized linear model; optimal design and quadratic logistic models

ABOUT THE AUTHORS

O. A. Fasoranbaku is a Professor of Statistics and faculty member of the Department of Statistics, School of Sciences at Federal University of Technology Akure, Nigeria. 
F. B. Adebola is an Associate professor of Statistics and also a faculty member of the Department of Statistics, School of Sciences at Federal University of Technology Akure, Nigeria. J. A. Kupolusi is a Ph.D. scholar with a promising future in the field of Statistics. He is also a faculty member of the Department of Statistics, School of Sciences at Federal University of Technology Akure, Nigeria.

PUBLIC INTEREST STATEMENT

Experiment is a controlled study in which observations are made and data collected forms the basis for analysis and subsequent conclusions. The quality of the analysis depends directly on the experimental design. Research in optimal design of experiment focused majorly on linear models in the past but has been extended to non-linear models including Generalized Linear Models (GLM). One major problem of finding optimal designs for GLM is its dependency on unknown model parameters. Prediction Error Variance (PEV) is a useful way to investigate the predictive capability of the model in optimal design techniques. It gives a measure of the precision of a model’s predictions. This can be used in any area of research to ascertain its optimality. A low PEV (close to zero) means that good predictions are obtained at that point. If the design PEV < 1, then the errors are reduced by the model fitting process and a good prediction is obtained at that point.
1. Introduction

Experimenters carry out experiments in almost all bodies of research and investigations for the purpose of discovering certain things about a particular process or system. Each of the experimental run is a test. Therefore, an experiment can be defined as a test or series of runs in which a reasonable adjustment is made to the input factor of a process or system to determine the purpose of changes that may be observed in the output response (Douglas, 2001). The nature of the experiment determines the analysis that will be carried out and the conclusion that will be made. Previous works on design of experiment are concentrated majorly on continuous response for linear model. The assumption of ordinary least square for classical model that the error term is normally distributed with a constant variance over the design space is widely used in this regard (Jie et al., 2016). The assumption is not satisfied even asymptotically with logistic models for binary response data where response Y can only take one out of the two possible outcomes denoted for convenience as 0 and 1. This kind of response observation comes up in medical trials where at the end of a trial period, the patient has either recovered denoted as (Y = 1) or has not recovered denoted as (Y = 0) from a particular illness, Yang et al. (2015). Some other examples include but not limited to: student pass or fail an examination; a lecturer is an indigene of a state or not an indigene. Response variables in these forms are logistic in nature that followed binomial distribution which is a family of generalized linear model that are characterized by three components namely:

(i) The distribution of each of the independent response variables Y₁, Yₙ belongs to the exponential family
(ii) The linear predictor η = Xᵀβ is linear (in β) combination of k control variables x₁, x₂, . . . , xₖ and p parameters
(iii) The link function g(μ) specifies the relationship between expected value of the response variable E(Y) = μ and the linear predictor. This is a monotonic and differentiable function

Designing an optimal experiment for this kind of response model is not easy because it solely depends on the initial guess parameter. The optimality of the design is determined by using a general equivalence theorem. This may not be actually optimal if further investigation to obtain the design predictive capability is carried out. This paper hereby used prediction error variance (PEV) as a means to confirm the design optimality.

1.1. Aim and objectives

This study aims to design experiments that are optimal for two-variable quadratic logistic regression models with the purpose to obtaining optimal designs for estimation of the associated quadratic response curve. The objectives of the study are to:

(1) obtain the performance of optimal as well as some non-optimal designs in small samples under two-factor quadratic logistic models.
(2) determine the predictive capability of two variable quadratic logistic model by using prediction error variance.

1.2. Motivation

Most of the researches that are carried out in the literature on optimal design of experiment for quadratic logistic models are based on one variable for linear models, one variable for quadratic models and multiple linear models. Hence, this study is motivated to extend the work of Fornius (2008) of one variable quadratic logistic model to two variable quadratic regression models to determine a design that can be useful for prediction.

2. Review of related work

Non-linear models are very important due to their various areas of application in different fields of study. Jafari et al. (2014) discussed D-optimal criterion where an appropriate models based on
a logistic regression model with three covariates and the information matrix rely on less active parameters was introduced that resulted in a locally optimal designs for various specific states. It is noteworthy that certain designs with different points were presented and their optimality were obtained subject to the parameters space.

Silvey (1980) discussed the support points for D-optimal design in general and the Caratheodory's theorem. The work also explained that in most linear problems, there is an upper bound on the number of support points for an optimal design. This was later presented by Chaloner & Larntz (1989) who gave examples of D-optimal design for one-variable logistic regression models that is based on a finite number of support points and more specifically on at most \( p(p + 1)/2 \) points, where \( p \) is the number of model parameters. Also, the D-optimal design for the linear models typically has an optimal number of support points that is the same as the number of parameters \( p \) and the weights associated with the support points are equal and thus equal to \( 1/p \) Gwenda (2010).

Li and Deng (2018) proposed a prediction-oriented design criterion, I-optimality and developed an efficient sequential algorithm of constructing I-optimal designs for generalized linear models. The General Equivalence Theorem was established for the I-optimality and the proposed algorithm was used to choose the support points sequentially and also to update the weights of support points of the design. The proposed algorithm is computationally efficient with guaranteed convergence property. Illustrative numerical examples were conducted on the proposed algorithm with a view to evaluate how feasible and efficient it is computationally.

3. Research and methods
Optimal design for two variable quadratic logistic models are described in this section with overall predictive power of the model. Method to obtaining optimal design was also presented that will be used for the simulation in conformity with the objectives of this study.

3.1. General equivalence theorem
The General Equivalent Theorem is fundamental to the theory of optimal design. There are many methods in practice to determine the optimal design. These include analytical, numerical, algorithms and graphical methods, used separately or in combination of two or more methods. Ordinarily, no method is preferred favorable but the optimality design determination depends majorly on the problem under consideration. The selected method to derive D-Optimal designs is described below;

According to Li & Deng (2018), D-optimality is achieved by minimizing \( |M^{-1}(\varepsilon, \alpha)| \), or equivalently maximizing \( |M(\varepsilon, \alpha)| \). To start with, the number of design point \( k \) is not known. Though it is known that there exists a D-optimal design with \( p \leq k \leq \frac{p(p+1)}{2} \) design points. A plot of the standardized predictor variance, \( d(x, \varepsilon, \alpha) \) serves as a useful tool, it helps to show whether a suggested design is optimal or not. If the design is optimal, the maximum peak of \( d(x, \varepsilon, \alpha) \) be equal to the number of parameters in the model. The maxima will also appear at the design points. In the non-optimal case the plot can give a hint of the optimal number of design points by looking at the number of peaks of the function \( d(x, \varepsilon, \alpha) \). The D-optimal design is essentially obtained according to the following steps.

(i) Begin with a k-point design
(ii) Minimize \( |M^{-1}(\varepsilon, \alpha)| \) yielding the best possible k-point design.
(iii) Plot the standardized predictor variance, \( d(x, \varepsilon, \alpha) \). If the visual inspection indicates that the design clearly is non-optimal, return to step ii and try a design with \( k = k + 1 \) point.
(iv) Verify optimality of the suggested design using the General Equivalent Theorem by assuming that the maximum of \( d(x, \varepsilon, \alpha) \) are attained at the candidate design points (either analytically or numerically). If the design cannot be verified to be optimal, go back to step ii.

The general equivalence theorem is used to construct and check designs and this is applied to a wide variety of the alphabet optimality criteria. Many of which are based on the information matrix \( M(\xi) \), which is a function of the design \( \xi \). Atkinson and Donev (2007) state that “The General Equivalence Theorem can be viewed as a consequence of the result that the derivatives are zero at a minimum of a smooth function over an unconstrained region.”

Let \( \psi(M(\xi)) \) represent a general measure of impression for the design \( \xi \). In the context of D-optimal design, for example, this measure is the determinant of the inverse of the information matrix \( \psi(M(\xi)) = \det[M(\xi)^{-1}] = -\det[M(\xi)] \), which is minimized.

Atkinson and Donev (2007) and Fisher (1989) showed that the function we are minimizing, \( \psi \), depends on \( \xi \) through the information matrix \( M(\xi) \). Using the notation adopted by the authors, let \( \xi \) be a one point design, denoted by \( x \), and let the design \( \xi' \) be given by \( \xi' = (1 - \alpha)\xi + \alpha \xi' \). Then the information matrix of \( \xi' \) can be written as

\[
M(\xi') = (1 - \alpha)M(\xi) + aM(\xi).
\]

and the derivative of \( \psi \) in the direction of \( \xi \) is

\[
\psi(x, \xi) = \lim_{\alpha \to 0} \frac{1}{\alpha} \left[ \psi \left( (1 - \alpha)M(\xi) + aM(\xi) \right) - \psi(M(\xi)) \right]
\]

The General Equivalence Theorem states the equivalence of the following three conditions on \( \xi^* \) according to Ford et al (1992) and Li et al (2018):

1. The design \( \xi^* \) minimizes \( \psi(\xi) \).
2. The design \( \xi^* \) maximizes the minimum over \( X \) of \( \psi(x, \xi) \).
3. The maximum over \( X \) of \( \psi(x, \xi^*) \) is equal to \( p \), the number of parameters in the model, and the maximum occurs at the points of support of the design.

### 3.2. Standardized predictor variance or prediction error variance (PEV)

Prediction Error Variance (PEV) is a useful way to investigate the predictive capability of the model. It gives a measure of the precision of a model’s predictions. PEV can be examined for designs and for models. It is useful to remember that: PEV (model) = PEV (design) * MSE. So the accuracy of the model’s predictions is dependent on the design PEV and the mean square errors in the data. For this, PEV for the design should be made as low as possible, as it is multiplied by the error on the model to give overall PEV for the model. A low PEV (close to zero) means that good predictions are obtained at that point. If the design PEV < 1, then the errors are reduced by the model fitting process and a good prediction is obtained at that point. If design PEV > 1, then any errors in the data measurements are multiplied. Overall predictive power of the model will be more accurate if PEV is closer to zero.

This study considered the following design matrix for the quadratic logistic model,
\[
X'X = \begin{bmatrix}
  1 & x_{1i} & x_{2i} & x_{1i}^2 & x_{2i}^2 \\
  x_{1i} & x_{1i}^2 & x_{1i}x_{2i} & x_{1i}^3 & x_{1i}x_{2i}^2 \\
  x_{2i} & x_{1i}x_{2i} & x_{2i}^2 & x_{1i}^2x_{2i} & x_{2i}^3 \\
  x_{1i}^2 & x_{1i}^3 & x_{1i}^2x_{2i} & x_{1i}^3 & x_{1i}^2x_{2i}^2 \\
  x_{2i}^2 & x_{1i}x_{2i}^2 & x_{2i}^3 & x_{1i}^2x_{2i}^2 & x_{2i}^4 
\end{bmatrix}
\]

Note: the only dependency of the design on observed values is in the variance of the measurement error. We can look at the PEV for a design (without MSE, as we don't yet have any observations) and see what effect it will have on the measurement error, if it is greater than 1 it will magnify the error, and the closer it is to 0 the more it will reduce the error. PEV is examined for designs or for the global models and can also be used to know how the underlying model predicts over the design region.

### 3.3. Two variables quadratic logistic model for generalized linear model

This study expressed two control variables quadratic logistic model which is an extension of logistic regression model with one control variable examined by Fornius (2008) & Yang et al (2012). The response variable is binary in nature like: Yes, or No, defective or non-defective, success or failure, survive a sickness or not survive a sickness. The control variable can either be continuous, temperature or categorical, (treatment group) which is represented by the square matrix in Equation (1). The responses are assumed to be Bernoulli and independently distributed \( Y \sim \text{bern}(\omega_i) = \text{bin}(1, \omega_i) \) with the logit link function according to Rursell (2019) and McCullagh & Nelder (1989).

\[
g(\omega_i) = \ln \left( \frac{\omega_i}{1 - \omega_i} \right) = \lambda(x_i)
\]

and response probability \( E(Y) = P(Y = 1) = \omega(x) \) given by

\[
\omega(x_i) = \frac{\exp(\lambda(x_i))}{1 + \exp(\lambda(x_i))}
\]

where the linear predictor for the logistic model with two control variables and quadratic terms is

\[
\lambda(x_i) = x_i'\alpha = \begin{bmatrix} 1 & x_{1i} & x_{2i} & x_{1i}^2 & x_{2i}^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}
\]

\[
\lambda(x_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{1i}^2 + \alpha_4 x_{2i}^2 \quad i = 1, 2, 3, \ldots, n
\]

The logit link specifies the connection between probability of positive response \( \omega(x) \) and linear predictor \( \lambda(x_i) \).

From Equation (5),

\[
\mu_i = \omega(x_i) = \frac{\exp(\lambda(x_i))}{1 + \exp(\lambda(x_i))} = \frac{\exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{1i}^2 + \alpha_4 x_{2i}^2)}{1 + \exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{1i}^2 + \alpha_4 x_{2i}^2)}
\]

The variance of response probability in generalized linear model setting is given as

\[
V(X) = \omega_i(1 - \omega_i)
\]

Recall from Equation (37) that \( \omega(x_i) = \frac{\exp(\lambda(x_i))}{1 + \exp(\lambda(x_i))} \)

Then,
\[ V(X) = \frac{\exp(\lambda(x_i))}{1+\exp(\lambda(x_i))} \left( 1 - \frac{\exp(\lambda(x_i))}{1+\exp(\lambda(x_i))} \right) = \frac{\exp(\lambda(x_i))}{1+\exp(\lambda(x_i))} \left( \frac{1 + \exp(\lambda(x_i)) - \exp(\lambda(x_i))}{1+\exp(\lambda(x_i))} \right) \]

\[ = \frac{\exp(\lambda(x_i))}{1+\exp(\lambda(x_i))} \left( 1 + \exp(\lambda(x_i)) \right) \left( \frac{1}{1+\exp(\lambda(x_i))} \right) \left( \frac{1 + \exp(\lambda(x_i))}{1+\exp(\lambda(x_i))} \right) \]

\[ \lambda(x_i) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3^2 + a_4 x_4^2 \]

hence, the variance becomes

\[ \text{var}(x_i) = \omega(x_i) \cdot (1 - \omega(x_i)) = \frac{\exp(a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3}^2 + a_4 x_{i4}^2)}{(1 + \exp(a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3}^2 + a_4 x_{i4}^2))^2} \]

(9)

Where \(a_0, a_1, a_2, a_3, a_4\) represent the model parameters.

Because of the fact that the link function is monotonic optimizing \(\omega\) implies optimizing the linear predictor \(\lambda\), thus the optimum points are obtained by using gradient vector \(\nu_i\) and equate the resulting solution to zero by Nelder & Wedderburn (1972). That is \(\nu_i = \left[ \frac{\partial \lambda}{\partial a_0}, \frac{\partial \lambda}{\partial a_1}, \ldots, \frac{\partial \lambda}{\partial a_4} \right]^T\)

(10)

Equating (10) to zero and solve for \(x_1\) and \(x_2\)

\[ a_1 + 2a_3 x_1 = 0 \]

\[ a_2 + 2a_4 x_2 = 0 \]

Thus, the solution for \(x_1\) and \(x_2\) at the optimum are respectively

\[ x_{1m} = \frac{-a_1}{2a_3} \text{ and } x_{2m} = \frac{-a_2}{2a_4} \]

(11)

The response curve that describes \(\omega_3\) as a function of \(x\) is symmetric around these points. Computing the Hessian matrix of the linear predictor \(\lambda_i\),

\[ H_i = \begin{bmatrix} \frac{\partial^2 \lambda}{\partial a_0^2} & \frac{\partial^2 \lambda}{\partial a_0 \partial a_1} & \frac{\partial^2 \lambda}{\partial a_0 \partial a_2} \\ \frac{\partial^2 \lambda}{\partial a_1 \partial a_2} & \frac{\partial^2 \lambda}{\partial a_1^2} & \frac{\partial^2 \lambda}{\partial a_1 \partial a_3} \\ \frac{\partial^2 \lambda}{\partial a_2 \partial a_3} & \frac{\partial^2 \lambda}{\partial a_2 \partial a_4} & \frac{\partial^2 \lambda}{\partial a_2^2} \end{bmatrix} \]

(12)

Equation (12) shows whether the probability of positive response \(\omega(x)\) has a maximum or minimum which is determined by the sign of the parameters \(a_3\) and \(a_4\). The parameter \(a_0\) determines the height of the curve in the optimum point. The linear predictor at the optimum point is obtained as

\[ \lambda(x_m) = a_0 - \frac{a_1^2}{4a_3} - \frac{a_2^2}{4a_4} \]

(13)

If the response curve has a maximum \((a_3 < 0, a_4 < 0)\) a greater of \(a_0\) means that the maximum of \(\omega\) is closer to 1 and if there is minimum \((a_3 < 0, a_4 < 0)\) a greater of \(a_0\) means that the minimum of \(\omega\) is closer to 1. Therefore, the size of \(a_3\) and \(a_4\) determines the relative width of the response curve for
Table 1. Four types of response curves with the parameter set

| Type of response curve | Parameter set          |
|------------------------|------------------------|
| High-wide              | $\alpha = (200 - 0.1 - 0.1)^T$ |
| High-narrow            | $\alpha = (200 - 4 - 4)^T$ |
| Low-wide               | $\alpha = (-200 - 0.1 - 0.1)^T$ |
| Low-narrow             | $\alpha = (-200 - 4 - 4)^T$ |

a given height and a larger absolute value of $\alpha_3$ and $\alpha_4$ means a more narrow curve. Thus, the parameters determine the shape of the function $\omega(x)$.

Four sets of parameters guess that are considered for the simulation study in this study are presented in Table 1. The parameter sets are chosen to represent different variations of the shape of the response curve, it can be “high” or “low” and given the scale on x-axis, it can be “wide” or “narrow”.

For the logistic regression with linear predictor in Equation (5), the standardized information matrix for the design obtained as

$$M(\xi, \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_i M(\alpha, x_{ij}) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_i \text{var}(x_{ij}) \cdot x_{ij}^T x_{ij}$$

(14)

Where $\text{var}(x_{ij})$ is the variance of the probability of response as defined in Equation (10), and

$$m(\alpha, x_{ij}) = \text{Var}(x_{ij}) \cdot x_{ij} \cdot x_{ij}^T$$

(15)

is the information matrix for the design and $P_i$ is design weight.

Equation (14) becomes

$$M(\xi, \alpha) = \sum_{i=1}^{n} P_i \omega(x_i) \cdot (1 - \omega(x_i)) \begin{bmatrix} 1 & x_1 & x_2 & x_3^2 & x_4^2 \\ x_1 & x_1^2 & x_1x_2 & x_1^3 & x_1x_4^2 \\ x_2 & x_1x_2 & x_2^2 & x_1^2x_2 & x_2^3 \\ x_3^2 & x_1^2x_2 & x_2^3 & x_1^2x_4 & x_3^3 \\ x_4^2 & x_1^3 & x_2^3 & x_1x_4^2 & x_4^4 \end{bmatrix}$$

(16)

$x_i$ is a $p \times 1$ design vector of the $i$th design points.

The prediction error variance is obtained from Equation (16) as

$$d(x, \xi_5) = \text{var}(x_i)^T M^{-1}(\xi_5, \alpha) X = \text{tr}[m(\alpha, x) M^{-1}(\xi_5, \alpha)]$$

(17)

4. Results and discussion

The initial design for the four set of true parameters are presented in Table 2. This was used to obtain the values of the covariates $X_2$ and $X_3$ that maximizes $|M(\xi, \alpha)|$ and the results of the investigation of optimal design of experiment for small and large samples in simulation study were discussed and presented alongside with the PEV value.

4.1. Simulation study

Optimal design of experiment for generalized linear models are obtained in many situations by using best parameter guess. This is achieved either by a conducted experiment or a pilot
The optimal model was obtained through a simulation study of 10000 experimental runs by using some parameter guess for the initial design. The purpose of the simulation study is to juxtapose optimal designs to non-optimal designs and determining the existence of optimal design in small samples. Therefore, grid search method was used in the simulation and the result of the four types of response curve are presented in Tables 3–10. The expected optimum value and prediction error variances were obtained for both small and large samples.

Table 3 presents predictive power of the model with initial parameter $\alpha = (2.0, 0.0, -0.1, -0.1)$ for high-wide type of response curve when the sample sizes are small. Here, PEV values obtained for 5point, 6point, 7point and 8point designs cannot be used to predict the model performance since they are greater than one (the higher the PEV, the lower the predictive capability of the model). The value of PEV for the four non-optimal designs; 5points design, 6points design, 7points design and 8points design, respectively, 2.012, 1.454, 1.219 and 1.161 when the sample size is 10 at a very high response on the vertical axis spread widely over the design region of the two X coordinates. This result does not indicate any significant change when the sample size is increased to 25 for all the non-optimal designs. Only the D-optimal design can be used for the model prediction since its PEV value is approximately 0.6 for the two small sample sizes. Table 4
| Type of response curve | Design  | Sample | PEV | Model parameters                  |
|------------------------|---------|--------|-----|-----------------------------------|
|                         |         |        |     |                                   |
|                         | n = 50  |        |     |                                   |
| “High-Wide”            | D-Optimal | 0.556  |     |                                   |
| 5points                |         | 2.839  |     |                                   |
| 6points                |         | 1.456  |     | $\alpha = (2, 0, 0, -0.1, -0.1)$ |
| 7points                |         | 1.117  |     |                                   |
| 8points                |         | 1.082  |     |                                   |
| n = 100                |         |        |     |                                   |
| “High-Wide”            | D-Optimal | 0.556  |     |                                   |
| 5points                |         | 3.760  |     |                                   |
| 6points                |         | 1.420  |     | $\alpha = (2, 0, 0, -0.1, -0.1)$ |
| 7points                |         | 1.131  |     |                                   |
| 8points                |         | 1.082  |     |                                   |

| Type of response curve | Design  | Sample | PEV | Model parameters                  |
|------------------------|---------|--------|-----|-----------------------------------|
|                         |         |        |     |                                   |
|                         |         |        |     |                                   |
|                         | n = 10  |        |     |                                   |
| “High-Narrow”           | D-Optimal | 0.591  |     |                                   |
| 5points                |         | 2.188  |     |                                   |
| 6points                |         | 1.521  |     | $\alpha = (2, 0, 0, -4, -4)$     |
| 7points                |         | 1.197  |     |                                   |
| 8points                |         | 1.091  |     |                                   |
| n = 25                 |         |        |     |                                   |
| “High-Narrow”           | D-Optimal | 0.562  |     |                                   |
| 5points                |         | 2.238  |     |                                   |
| 6points                |         | 1.480  |     | $\alpha = (2, 0, 0, -4, -4)$     |
| 7points                |         | 1.179  |     |                                   |
| 8points                |         | 1.095  |     |                                   |

| Type of response curve | Design  | Sample | PEV | Model parameters                  |
|------------------------|---------|--------|-----|-----------------------------------|
|                         |         |        |     |                                   |
|                         |         |        |     |                                   |
|                         | n = 50  |        |     |                                   |
| “High-Narrow”           | D-Optimal | 0.556  |     |                                   |
| 5points                |         | 2.166  |     |                                   |
| 6points                |         | 1.484  |     | $\alpha = (2, 0, 0, -4, -4)$     |
| 7points                |         | 1.145  |     |                                   |
| 8points                |         | 1.141  |     |                                   |
| n = 100                |         |        |     |                                   |
| “High-Narrow”           | D-Optimal | 0.556  |     |                                   |
| 5points                |         | 2.383  |     |                                   |
| 6points                |         | 1.468  |     | $\alpha = (2, 0, 0, -4, -4)$     |
| 7points                |         | 1.215  |     |                                   |
| 8points                |         | 1.175  |     |                                   |
### Table 7. Two small sample sizes $n = 10$ and $n = 25$ for Low-Wide response curve

| Type of response curve | Design   | Sample | PEV  | Model parameters          |
|------------------------|----------|--------|------|---------------------------|
|                        |          | $n = 10$ |      |                           |
| “Low-Wide”             | D-Optimal|        | 0.606|                           |
|                        | 5points  |        | 1.887|                           |
|                        | 6points  |        | 1.486| $\alpha = (-2, 0, 0, -0.1, -0.1)$ |
|                        | 7points  |        | 1.112|                           |
|                        | 8points  |        | 1.043|                           |
|                        | $n = 25$ |        |      |                           |
| “Low-Wide”             | D-Optimal|        | 0.560|                           |
|                        | 5points  |        | 3.029|                           |
|                        | 6points  |        | 1.422| $\alpha = (-2, 0, 0, -0.1, -0.1)$ |
|                        | 7points  |        | 1.215|                           |
|                        | 8points  |        | 1.097|                           |

### Table 8. Two large sample sizes $n = 50$ and $n = 100$ for Low-Wide response curve

| Type of response curve | Design   | Sample | PEV  | Model parameters          |
|------------------------|----------|--------|------|---------------------------|
|                        |          | $n = 50$ |      |                           |
| “Low-Wide”             | D-Optimal|        | 0.556|                           |
|                        | 5points  |        | 2.418|                           |
|                        | 6points  |        | 1.427| $\alpha = (-2, 0, 0, -0.1, -0.1)$ |
|                        | 7points  |        | 1.209|                           |
|                        | 8points  |        | 1.110|                           |
|                        | $n = 100$|        |      |                           |
| “Low-Wide”             | D-Optimal|        | 0.556|                           |
|                        | 5points  |        | 3.349|                           |
|                        | 6points  |        | 1.528| $\alpha = (-2, 0, 0, -0.1, -0.1)$ |
|                        | 7points  |        | 1.133|                           |
|                        | 8points  |        | 1.083|                           |

### Table 9. Two small sample sizes $n = 10$ and $n = 25$ for Low-Narrow response curve

| Type of response curve | Design   | Sample | PEV  | Model parameters          |
|------------------------|----------|--------|------|---------------------------|
|                        |          | $n = 10$ |      |                           |
| “Low-Narrow”           | D-Optimal|        | 0.586|                           |
|                        | 5points  |        | 2.921|                           |
|                        | 6points  |        | 1.450| $\alpha = (-2, 0, 0, -4, -4)$ |
|                        | 7points  |        | 1.151|                           |
|                        | 8points  |        | 1.157|                           |
|                        | $n = 25$ |        |      |                           |
| “Low-Narrow”           | D-Optimal|        | 0.556|                           |
|                        | 5points  |        | 2.746|                           |
|                        | 6points  |        | 1.606| $\alpha = (-2, 0, 0, -4, -4)$ |
|                        | 7points  |        | 1.490|                           |
|                        | 8points  |        | 1.097|                           |
through Table 10 present the predictive power of the model with various initial parameter guess at different sample sizes. It was demonstrated for the large sample sizes that there is no significant different in the performance of the model performance for the PEV values at different design points considered.

Figures 1–5 below present the surface plot of optimal design and some non-optimal designs. The optimality of the design is achieved by updating the initial five design points to estimate the probability of the response curves. A common feature of the design is that the design points are placed symmetrically around the point of maximum response. An examination of the surface plots for optimal design indicate that the number of peaks of the surfaces should be equal to the number of parameters in the model i.e. $d(x, \xi) \leq p = 5$ and the maxima are attained at nine design points.

| Type of response curve | Design      | Sample | PEV   | Model parameters |
|------------------------|-------------|--------|-------|------------------|
| "Low-Narrow"           | D-Optimal   | 5      | 0.556 |                  |
|                        | 6points     | 3.358  |       |                  |
|                        | 7points     | 1.454  | \(\alpha = (-2, 0, 0, -4, -4)\) |
|                        | 8points     | 1.089  |       |                  |

| Type of response curve | Design      | Sample | PEV   | Model parameters |
|------------------------|-------------|--------|-------|------------------|
| "Low-Narrow"           | D-Optimal   | 5      | 0.556 |                  |
|                        | 6points     | 3.353  |       |                  |
|                        | 7points     | 1.454  | \(\alpha = (-2, 0, 0, -4, -4)\) |
|                        | 8points     | 1.184  |       |                  |

Figure 1. Surface plot of 5point design.
The visual inspection of the surface plot is a necessary but not a sufficient condition to determine the optimality of a design. The optimality of the design is confirmed through predictive error variance (PEV) as presented in Table 3 through Table 10.
4.2. Discussion
Tables 3–10 had shown the predictive power of the model considered in this study. The results had shown that model performance in optimal design setting is dependent on the initial best guess parameters but not on the sample size. The type of response curve considered for two small sample sizes $n = 10$ and $n = 25$ and two large sample sizes $n = 50$ and $n = 100$ did not indicate any significant variation in the values of PEV obtained from the simulation. This is a confirmation of existing argument in the literature that optimal design is not sample size-dependent irrespective of the sample size used. D-optimal design shows a noticeable predictive capability of the model.
since its prediction error variance (PEV) is less than one. Any design with PEV greater than one, such a design may be optimal locally but not good for prediction.

4.3. Conclusion and recommendation
The study has presented optimal experimental design for two variables quadratic logistic regression model. The optimality of the design was confirmed through the prediction error variance which is less than one and the optimal design according to caratheodory’s theorem is within the bounds on the support points p where 5≤p≤15. Four types of response curves are examined and the number of support points ranges from 5 to 9 on each with the sample size. For quadratic logistic model with number of parameters P = 5, caratheodory’s theorem gives the bounds on the support points as 5≤p≤15. However, these designs satisfied the bound as the optimal design for the model considered was obtained at 9 support points. Locally optimal designs for the model under consideration were derived for four different parameter sets. The result of the experiment indicates that optimal design exists by observing the visual inspection of the surface plot and contour plot. It is therefore recommended that determination of optimal design for quadratic logistic model should be verified through prediction error variance to ascertain the stability of the model used for the design.

Funding
The authors received no direct funding for this research.

Author details
F. B. Adebola1
E-mail: fbadobola@futa.edu.ng
O. A. Fasoranbaku1
E-mail: jakupolusi@futa.edu.ng
J. A. Kupolusi1
E-mail: emeritus.ay@gmail.com

1 Department of Statistics, Federal University of Technology, Akure, Nigeria.

Authors’ statement
The research of the authors span across various areas of statistics like Design and Analysis of Experiment, Econometrics, Survey Design and Operation Research. Much of their work has been centered on optimal design of experiment for GLM models, Design of Experiment for two variable quadratic logistic models and performance of Prediction Error Variance for design optimality.

Citation information
Cite this article as: On prediction error variance to determine optimal design for two variable quadratic logistic model, F. B. Adebola, O. A. Fasoranbaku & J. A. Kupolusi, Cogent Mathematics & Statistics (2020), 7: 1853888.

References
Atkinson, A. C., & Done, A. N. 2007. Optimum experimental designs. Oxford Statistical Science Series-8. Oxford University Press.
Chaloner, K., & Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. Journal of Statistical Planning and Inference, 21, 191–208. https://www.sciencedirect.com/science/article/pii/0378375889900049
Fisher, R. A. (1935). The design of experiments. Oliver & Boyd Ltd.
Ford, I., Torsney, B., & Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for nonlinear problems. Journal of Royal Statistical Society, Series B, 54, 569–583.
Fornius, E. F. (2008). D-optimal designs for quadratic logistic regression models. International Journal of Statistical Sciences, 6(Special Issue), 269–303. http://gauss.stat.su.se/site/pdf/Effdoc.pdf
Gwenda, T. (2010). Optimal design for generalized linear models with a multinomial response [Ph.D. Thesis]. University of Wollongong, Research Online, 10–130.
Jafari, H., Khazai, S., Khaki, Y., & Jafari, T. (2014). D-optimal design for logistic regression model with three independent variables. Asian Journal of Scientific Research, 4(3), 120–124.
Jie, Y., Abhyuday, M., & Dibyen, M. (2016). Optimal designs for 2k factorial experiments with binaryresponse. Statistica Sinica, 385–411. http://dx.doi.org/10.5705/ss.2013.265
Li, Y., & Deng, X. (2018). On I-optimal designs for generalized linear models: An efficient algorithm via general equivalence theory. http://arxiv.org/abs/1801.05861v1.
McCullagh, P., & Nelder, J. A. (1989). Generalized linear models (2nd ed.). Chapman and Hall.
Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. Journal of the Royal Statistical Society. Series A, 135(3), 370–384. https://doi.org/10.2307/2344614
Russell, K. G. (2019). Design of experiment for generalized linear models. Taylor and Francis group.
Silvey, D. (1980). Optimal design. Chapman & Hall.
Yang, J., Tong, L., & Mandal, A. (2015). D-optimal designs with ordered categorical data. Communication In Statistics-Simulation And Computation, 1879–1902.
Yang, M., & Stufken, J. (2012). On locally optimal designs for generalized linear models with group effects.
