
Model Adequacy in Assessing the Predictive Performance of Regression Models in Pharmaceutical Product Optimization: The Bedaquiline Solid Lipid Nanoparticle Example

[Chidi U Uche](#)*, [Mercy Amaka Okezue](#)*, [Ibrahim Amidu](#), [Stephen R Byrn](#)

Posted Date: 31 October 2024

doi: 10.20944/preprints202410.2593.v1

Keywords: Optimization; Response surface; Central composite design; First-order model; Second-order model; Contour plots; Model adequacy; Fitness; Sonicated; Unsonicated



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Model Adequacy in Assessing the Predictive Performance of Regression Models in Pharmaceutical Product Optimization: The Bedaquiline Solid Lipid Nanoparticle Example

Chidi U. Uche ¹, Mercy A. Okezue ², Ibrahim Amidu ³, and Stephen Byrn ⁴

¹ Biotechnology Innovation and Regulatory Science (BIRS) Center, Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907, USA; uche0@purdue.edu

² School of Packaging, Michigan State University, East Lansing, MI 48824, USA; okezueme@msu.edu

³ Nigerian Airforce Hospital, Headquarters 661, Ikeja Lagos, Nigeria; pharmibrahim89@gmail.com

⁴ Industrial and Molecular Pharmaceutics Department, Purdue University, West Lafayette, IN 47907, USA; sbyrn@purdue.edu

* Correspondence: okezueme@msu.edu; uche0@purdue.edu

Abstract: This study aimed to assess the predictive performance of first and second-order regression models in optimizing bedaquiline (BQ) solid lipid nanoparticles (SLNs) formulations. Employed a central composite design and graphical optimization process using three steps. A design of experiments to evaluate the impact of BQ, Tween 80 (T80), polyethylene glycol (PEG), and lecithin on the formulations' response variables: Z-average (PSD), polydispersibility index (PDI), and Zeta potential (ZP). Secondly, quantified the relationship between experimental variables using the regression model coefficients. Lastly, predicted the responses and verified the models' adequacies to ensure accurate representation and effective optimization. First-order polynomial showed poor model adequacy that required further refinement. The second-order models provided superior fitness, sensitivity to variability, complexity, and prediction consistency. The optimized formulation achieved a desirability value of 0.9998 indicating alignment with the desired criteria. Specifically, the levels of BQ (19.4mg), T80 (25.2mg), PEG (39.2mg), and Lecithin (200mg) corresponded to PDI (0.41), PSD (250.99 nm), and ZP (-25.95 mV). Maintaining BQ concentration between 10-20% and T80 between 15-18% is vital for maximizing ZP and minimizing PDI and PSD, ensuring stable SLN formulations. The study underscores significance of precise model selection and statistical analysis in pharmaceutical formulation optimization for enhanced drug delivery systems.

Keywords: optimization, response surface, central composite design, first-order model, second-order model, contour plots, model adequacy, fitness, sonicated, unsonicated

Introduction

Developing efficient drug delivery systems in the pharmaceutical industry is crucial for enhancing therapeutic outcomes and patient compliance. Solid lipid nanoparticles (SLNs) have emerged as promising carriers for drug delivery due to their ability to improve drug solubility, stability, and bioavailability. One such drug of interest is bedaquiline (BQ), which is used in combination with other drugs in treating multi-drug-resistant tuberculosis (MDR-TB). In earlier research, this group communicated formulating BQ as SLNs for oral delivery presenting an opportunity to enhance the drug product performance [1]. The various regimens for treating tuberculosis (TB) require using multiple drug combinations for at least six months [2], therefore, the prior research aimed to enhance patient adherence by formulating BQ as SLNs over an existing parenteral approach [3]. This current effort provides an in-depth analysis of how to use first-order linear and second-order quadratic models to assess the adequacy of predicting the performance of a new formulation. The regression equations were methodically employed to investigate how levels of the drug and excipients impact critical quality attributes (CQAs) such as Z-average (PSD),

polydispersibility index (PdI), and Zeta potential (ZP). Utilizing a full factorial design, the group systematically investigated the impact of these key formulation variables on the CQAs of the BQ SLN formulation.

After identifying a potential nanocarrier system, a critical phase in product development involves designing and optimizing the formulation [4, 5]. Various statistical experimental design methods have been employed in pharmaceutical manufacturing to optimize process parameters. Traditional optimization techniques typically involve altering one factor at a time, necessitating numerous experiments and failing to capture the combined effects of multiple variables. These methods also demand extensive data and time to identify optimal levels, making them unreliable [6]. In contrast, the design of the experiment (DoE) approach aims to understand interactions among variables, facilitating the optimization process and providing relevant statistical models [7, 8]. In line with this, multiple techniques and approaches can be utilized to achieve this optimal nanoparticle formulation; one such method is the central composite design (CCD) method [4, 8]. Therefore, this study examined the combined effects of input factors such as BQ, polysorbate or tween 80 (T80), polyethylene glycol (PEG), and lecithin on sonicated and unsonicated SLN formulations. Furthermore, the research optimized the formulation process parameters using CCD and the response surface methodology (RSM). This approach seeks to provide contour and response surface plots to further evaluate the effects of input factors on response variables, allowing for a more comprehensive understanding of the formulation optimization process. Therefore, in product optimization, a typical first step is to use a first-order statistical analysis to screen the variables that will be further investigated under a multifactorial experimental design.

The first-order model offers an initial exploration of the response surface and identifies regions of interest for further investigation. The results obtained can guide the design of experiments for a second-order model or other multifactorial experimental design. For instance, if certain factors are identified as significant in the first-order model, they can be prioritized in the design of the second-order experiments [9]. These steps were used to ensure a more efficient and targeted approach to experimentation used in the current study.

1.1. First-order statistical analysis

Generating a first-order model using statistical software like JMP involves several critical steps. Initially, data collection is paramount, requiring the identification of input variables (factors) and the response variable, collected systematically for accuracy. Various metrics such as R-squared (R^2), adjusted R^2 , p -values, and residual plots are provided to evaluate the model's fit and explain the variability in the response variable [10].

1.2. Multifactorial experimental design

The parameters were analyzed using an RSM design called CCD. RSM is a systematic approach that uses statistical and mathematical tools to model and analyze the relationships between several variables and the responses they produce in a process [9]. It is useful for fitting a quadratic surface to the data, which helps to identify the optimal settings for process parameters. By minimizing the number of required experiments, RSM efficiently explores the relationships and interactions between multiple variables, enabling a thorough analysis and optimization of the process [11, 12]. It employs quantitative data from the experiments to construct regression models and optimize an outcome (dependent variable) affected by the independent factors (input variables).

Equations (1) and (2) demonstrate linear and quadratic functions, respectively.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q + \epsilon \quad (i = 1, 2, \dots, N) \quad (1)$$

$$Y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \sum_{i < j=2}^k \beta_{ij} x_i x_j + \epsilon \dots \dots \dots \quad (2)$$

[9]

In these equations, Y represents the response, β_0 is a constant, and β_i , β_{ii} , and β_{ij} are coefficients representing linear effects and interaction terms. The terms x_i and x_j are the independent variables, and ϵ denotes the error.

1.3. Central Composite Design (CCD)

CCD is an experimental approach commonly used in RSM [10]. It offers an efficient way to explore the influence of multiple variables (n) on a response. CCD combines several elements:

- Factorial runs ($2n$): These experiments investigate each variable at two pre-defined levels.
- Axial points ($2n$): These additional runs explore values beyond the initial two levels for each variable, often denoted as $\pm\alpha$ from the center point.
- Center points (nc): Replicates are performed at a central point where all variables are set at their mid-point values.

This design allows researchers to assess both linear and quadratic effects of the variables on the response. With a growing number of variables (n), the total number of runs in a complete CCD design increases slower than in a full factorial design, making it more efficient. CCD is useful for modeling quadratic relationships, where the individual second-order effects cannot be isolated using only factorial runs. This makes it an invaluable tool for process optimization, as it can identify the combination of variables that achieve the desired outcome, particularly in exploring the potential of BQ SLNs for oral delivery. The combined effects of these factors are optimized to achieve the desired outcomes, either by minimizing or maximizing specific conditions [8, 10, 13]. This study aimed to assess the predictive performance of first and second-order regression models in optimizing the BQ SLN formulations.

2. Materials and Methods

2.1. Experimental Design

The method for preparation of the BQ SLN was comprehensively described in the first paper [1]. Summarily, 10- 20mg of BQ and a target mass of 212mg lecithin were used to form the organic phase, while the aqueous phase was prepared using a 1:2 ratio of T80: PEG. The optimization process for the BQ SLN formulation used RSM involving three fundamental steps. The first was to design statistically based experiments (DOE) using a selection and control of four independent variables (BQ, T80, PEG, and lecithin) and systematically studying their effects on the CQAs of the BQ SLN formulations. Secondly, estimating the coefficients within a mathematical model to quantify the relationship between the independent variables and the formulation's CQAs (PSD, PdI, and ZP). Lastly, predicting the response and verifying the model's adequacy, to ensure that the model accurately represents the experimental data and can be used to optimize the process effectively. Four independent variables were chosen for the DOE: BQ (X_1), T80 (X_2), PEG (X_3), and lecithin (X_4), see Equation 3. The range and levels of these factors were varied according to the experimental design. These four variables, along with their respective ranges, were identified as critical parameters for improving the drug release profile of the solid lipid nanoparticle formulation [1].

$$Y = f(X_1, X_2, X_3, \dots, X_n) \quad (3)$$

where Y is the response of the system and X_n is the independent variables (input factors).

It is assumed in the design, that the independent variables in the experiments are continuous and regulated by experiments with negligible errors. The CQAs were identified as PdI (Y_1), PSD (Y_2), and ZP (Y_3).

2.2. Generating experimental data from DOE trial runs

The various ranges and levels of the input variables were used to generate twenty-four trial samples of BQ SLNs. These samples were then evaluated to obtain the response values (observed data) needed to generate predicted data to optimize the formulation. A full factorial design was

created to consider all factor-level combinations before fitting the model. The collected data was formatted and entered into JMP Pro version 16 statistical software [14] to fit the model. The software employs the least squares method for fitting linear regressions such as the first-order model.

2.3. Model development and analyses.

This section was divided into two parts. The first was model fitting using a first-order polynomial, and then a statistical analysis and optimization using a second-order model.

2.3.1. Model fitting using a first order polynomial.

The first-order model was fitted using the JMP software, followed by regression analysis to estimate coefficients that minimize the difference between observed and predicted response values. This first-order regression model was applied to the data using ordinary least squares (OLS) regression methods. The resulting coefficients (β_0 , β_1 , β_2 , β_3 , etc.) were analyzed to understudy the linear relationships between the input variables and the responses. The first-order regression model is expressed as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \dots\dots\dots (4)$$

Y_{PdI} , Y_{PSD} , and Y_{ZP} are the dependent variables for unsonicated and sonicated BQ SLN formulation, while X_1 , X_2 , X_3 , and X_4 are the independent variables. β_0 , β_1 , β_2 β_p are the coefficients and ϵ is the error term (residual).

2.3.2. Statistical analysis and optimization using a second-order model

The JMP software also performed parameter combinations for the CCD. The design evaluated the effect of BQ, T80, PEG, and lecithin as independent variables, on PSD, PdI, and ZP as response variables. Polynomial coefficients were computed from experimental data to predict these response variables. The model adequacy was tested using the Analysis of Variance (ANOVA) test. By applying a multiple regression analysis, the results were adjusted to a second-order polynomial equation. The equation for responses in both the unsonicated BQ formulation and the sonicated versions were obtained from the group's prior study, shown in Table 1

Table 1. Ranges and levels for independent variables used in the DOE to target achieving CQAs for the BQ SLN formulation.

Input Variables (mg)		Range and Levels				
		$-\beta$	-1	0	+1	$+\beta$
X_1	Bedaquiline (BQ)	11.1	13.5	15.8	18.2	20.5
X_2	Polysorbate 80 (T80)	11.3	14.6	17.9	21.1	24.4
X_3	Polyethylene Glycol (PEG)	24.1	30.8	37.5	44.2	50.9
X_4	Lecithin (Lec)	202.6	215.5	228.4	241.2	254.1
Response Variables		Goals		Acceptable ranges		
Y_1	Polydispersity Index (PdI)	Minimize		≤ 0.4		
Y_2	Z Average (PSD)	Minimize		$\leq 500\text{nm}$		
Y_3	Zeta Potential (ZP)	Match target		$-30\text{mV} < X < +30\text{mV}$		

Source [1]]

The experimental runs were carried out in a randomized sequence to minimize errors and the impact of uncontrolled factors. The research optimized the response variables (Y) by closely approximating the relationship between the independent variables and the response surfaces. The

response variables were utilized to develop an empirical model that correlates experimental variables through a second-degree polynomial equation, as represented in Equation 5 [9, 15].

$$Y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \sum_{i < j=2}^k \beta_{ij} x_i x_j + \epsilon \dots \dots \dots \quad (5)$$

In this equation, Y represents the response predicted by the model. β_0 is the constant coefficient, β_j is the linear coefficient, β_{jj} is the quadratic coefficient, and β_{ij} is the interaction coefficient. The variable parameters for the BQ SLN formulation process are denoted by x_i and x_j , with n representing the number of factors being studied and optimized. The term ϵ signifies the random error component in the model.

Table 2. Relationship between the coded and actual value of the variables.

Code	Actual values of variables
$-\beta$	X_{\min}
-1	$[(X_{\max} + X_{\min})/2] - [(X_{\max} - X_{\min})/2] \alpha$
0	$[(X_{\max} + X_{\min})/2]$
+1	$[(X_{\max} + X_{\min})/2] + [(X_{\max} - X_{\min})/2] \alpha$
$+\beta$	X_{\max}

Note: $\alpha = 2^{k/4}$; K is the number of independent variables = 4. Contents of Table adapted from [16].

The codes used in Table 2 were used to calculate the function of the range of interest of each input factor shown in Table 1. This study utilized a face-centered (FC) CCD, where the number of tests (N) includes the standard 2^k factorial points with the origin at the center. It placed axial points at the face centers of the factorial cube to simplify the design and avoid issues with extreme points. In this design, the quadratic terms were generated from the center at a distance of α from $2k$ axial points with no replicates or center points. Here, the number of independent variables, $K=4$, therefore $\alpha = 2.00$, obtained from $2^{4/4}$. The axial points ($2k$) are used for screening analysis and maintaining readability, ensuring constant variance of model predictions at equidistant points from the design center [16].

The total number of experimental trials (N) used in the DOE is given by Equation 5, where C is the number of center points. The study design has four independent variables ($k=4$), and the number of center points is assumed to be 0, therefore, the total number of trial experiments conducted was $N=16+8+0=24$

$$N=2^k+2k+C \quad (6)$$

Thus, for these experiments with four variables, the design includes sixteen factorial points, eight axial points, and zero replicated center points, totaling 24 runs. The variables are coded to ± 1 for factorial points, 0 for center points, and $\pm \beta$ for axial points, facilitating a straightforward and efficient experimental setup.

2.3.3. Model fitting parameters.

In this study, various fit statistics were used to assess the adequacy and performance of the regression models. The primary parameters included the coefficient of determination (R^2), adjusted R^2 , and predicted R^2 , which indicate the proportion of variance in the response variable explained by the model. A higher R^2 value signifies better model fit. Adjusted R^2 accounts for the number of predictors in the model, offering a more accurate measure of fit when multiple variables are involved. Predicted R^2 evaluates the model's ability to predict new data, with higher values indicating strong predictive performance. Additional metrics such as the Akaike Information Criterion (AIC) and

Bayesian Information Criterion (BIC) were used to balance model fit with complexity, where lower values suggest a more parsimonious model. Finally, the Root Mean Square Error (RMSE) and Coefficient of Variation (CV) were calculated to assess the model's precision and variability, with lower values reflecting higher accuracy and consistency [9].

2.4. Stepwise evaluation of the regression model's adequacy

To evaluate the adequacy of regression models applied to the BQ formulations, the model adequacy check compared the performance of first-order regression models between the unsonicated and sonicated BQ formulations. The analysis involved fitting first-order polynomial equations to the data to reveal which formulations demonstrate acceptable model adequacy. If initial analyses suggest that the models need further refinement to enhance their predictive accuracy and overall reliability, second-order models will be utilized.

2.5. Contour Plots

The JMP statistical software was employed to perform second-order polynomial regression analyses on experimental data generated from the BQ SLN sonicated formulations. The process involved generating critical diagnostic plots such as the actual vs. predicted plots, contour plots, and response surface plots, which are essential for evaluating model performance and understanding the relationships between variables.

3. Results

3.1. Model fitting using a first order polynomial.

The following first-order regression models were obtained for the sonicated and unsonicated BQ SLN formulations. The model equations listed below were derived by executing parameter estimations in the JMP software. The linear equations describe the overall relationship between predictors and expression levels (response variables) within the chosen statistical model. The resulting coefficients illustrate the linear relationships between the input variables and the responses.

For unsonicated bedaqualine formulations:

$$Y_{PdI} = 0.2854 + 0.0131(X_1) - 0.0004(X_2) + 0.0011(X_3) \quad (7)$$

$$Y_{PSD} = 339.979 + 8.354(X_1) - 0.4992(X_2) - 1.2104(X_3) - .0107(X_4) \quad (8)$$

For Sonicated Bedaqualine Formulation:

$$Y_{PdI} = 0.0290 + 0.0213(X_1) - 0.0004(X_2) + 0.0004(X_3) \quad (9)$$

$$Y_{PSD} = 70.8680 + 7.1795(X_1) + 0.7935(X_2) - 0.4162(X_3) - 0.0113(X_4) \quad (10)$$

$$Y_{ZP} = -42.0184 + 0.1594(X_1) + 0.0907(X_2) - 0.1958(X_3) - 0.0009(X_4) \quad (11)$$

3.1.1. ANOVA test results for the BQ SLN formulations

Full details of the ANOVA results for the sonicated and unsonicated BQ SLN formulations were communicated in an earlier paper [1]. Summarily, the ANOVA results from the DOE data indicated that the model lacked significance for all dependent variables in the unsonicated formulations, the p -values for PdI and PSD were estimated as 0.5877 and 0.0814, respectively. However, in the sonicated formulations, significant effects were more pronounced, particularly with PdI, compared to other response variables such as PSD and ZP. The p -value of PdI was found to be statistically significant ($p = 0.0003$) for a level of significance, $\alpha = 0.05$.

3.1.2. Summary of the first-order regression models

The optimization of pharmaceutical formulations, such as BQ, requires precise and reliable models to predict key response variables. The adequacy of first-order models for predicting response variables in the sonicated and unsonicated BQ formulations is summarized in Table 3.

Table 3. Model summary for unsonicated and sonicated bedaqualine formulation using first-order linear regression models.

Model	R ²	Adj R ²	Predicted R ²	Precision Ratio	Std Dev.	CV(%)	AIC	BIC
Unsonicated Formulation								
PdI	0.1319	-0.0584	-0.2568	0.8809	0.16	31.77	-5.97	-3.85
PSD	0.3403	0.2014	-8.03	0.5158	67.68	14.20	276.4	278.5
Sonicated Formulation								
PdI	0.6537	0.5808	-12.2718	0.9748	0.01691	32.60	-45.68	-43.55
PSD	0.3403	0.2014	-15.8241	0.00008	11.4569	28.04	267.16	269.29
ZP	0.07806	-0.1160	-11.3790	0.0025	2.0795	28.83	185.25	187.38

The study compared the optimization of BQ formulations using first-order models for sonicated and unsonicated variants. The unsonicated formulation's PdI model showed poor fit and prediction capabilities ($R^2 = 0.1319$, Adjusted $R^2 = -0.0584$, Predicted $R^2 = -0.2568$), with a Precision Ratio of 0.8809, a standard deviation of 0.0331, and a Coefficient of Variation (CV) of 20.41%. The Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were -5.97 and -3.85, respectively. The PSD model for unsonicated formulations had moderate fit ($R^2 = 0.3403$, Adjusted $R^2 = 0.2014$, Predicted $R^2 = -15.8241$), but poor precision (Precision Ratio = 0.00008, CV = 28.04%), with AIC and BIC values of 267.16 and 269.29. The ZP for unsonicated formulation was not included because ZP values in sonicated formulations provide more meaningful insights into the formulation's stability and performance. Sonication unlike the unsonicated counterpart can alter particle size and surface charge, making ZP a critical factor in evaluating and optimizing the formulation [17].

For the sonicated formulation, the PdI model exhibited better fit ($R^2 = 0.6537$, Adjusted $R^2 = 0.5808$), but poor prediction (Predicted $R^2 = -12.2718$), with a Precision Ratio of 0.9748, a standard deviation of 0.01691, and a CV of 32.60%. It also had lower AIC and BIC values, -45.68 and -43.55. The PSD model for sonicated formulations had similar fit and prediction issues as unsonicated, but the ZP model showed poor fit and prediction ($R^2 = 0.07806$, Adjusted $R^2 = -0.1160$, Predicted $R^2 = -11.3790$), with AIC and BIC values of 185.25 and 187.38. A lower AIC or BIC value indicates a better fit [18]. Overall, the sonicated formulation's PdI model demonstrated better fit and precision, making it more adequate despite prediction challenges, highlighting areas for improvement in both formulations. Considering the poor indices obtained using the first-order modeling, there was a need to employ a second-order model approach.

3.2. Model fitting using a second-order polynomial (CCD)

The software was also used to fit the second-order model equations that evaluated the effect of BQ, lecithin, T80, and PEG as independent variables, on PSD, PdI, and ZP as response variables. Second-order polynomial coefficients were computed from experimental data to predict these response variables. The model adequacy was tested using the ANOVA test. Multiple regression techniques were used to analyze the data and fit a model, which unlike the first order, includes linear terms, quadratic, and interaction terms. By this, second-order polynomial equations for the sonicated BQ formulations, that captured more complex relationships and curvature in the data were generated as shown below:

$$Y_{PdI} = 0.411 + 0.002X_1 + 0.028X_2 + 0.001X_3 + 0.001(X_2)^2 \quad (12)$$

$$\begin{aligned}
 Y_{\text{PSD}} = & -34.309 - 44.184X_1 + 38.489X_2 - 5.840X_3 + 2.637X_4 + \\
 & 0.927(X_1)^2 \\
 & +0.562X_1X_2 - 0.662(X_2)^2 + 0.226X_1X_3 + 0.234X_2X_3 + 0.063(X_3)^2 \quad (13) \\
 & - 0.012 X_1X_4 - 0.101X_2X_4 - 0.030 X_3X_4
 \end{aligned}$$

$$\begin{aligned}
 Y_{\text{ZP}} = & 53.184 - 5.253X_1 - 1.943X_2 + 2.593X_3 + 49.309X_4 - \\
 & 0.051(X_1)^2 \\
 & + 0.135X_1X_2 - 0.180(X_2)^2 + 0.027X_1X_3 + 0.024X_2X_3 \quad (14) \\
 & 0.031(X_3)^2 + 0.01 X_1X_4 - 0.003 X_2X_4 - 0.004 X_3X_4
 \end{aligned}$$

3.2.1. Model-adequacy checks for the first and second-order regression models used to evaluate sonicated BQ SLN formulations.

To evaluate the model adequacy of the first-order and second-order models in predicting the behavior of sonicated BQ formulations, the main effects from both models for sonicated BQ SLN formulation are compared in Table 4. To further compare fitness, variability in response factors, complexity, and prediction consistency in the two models, other model evaluation metrics were computed in Table 5.

Table 4. ANOVA table comparing main effects from first and second-order model sonicated BQ SLN formulation.

Variables	Sonicated Formulation (1 st order Model)			Sonicated Formulation (2 nd order model)		
	PdI	Z-Average (PSD)	ZP	PdI	Z-Average (PSD)	ZP
Main Effects	<i>p</i> -Value	<i>p</i> -Value	<i>p</i> -Value	<i>p</i> -Value	<i>p</i> -Value	<i>p</i> -Value
Model	0.0003*	0.0814	0.8046	0.0073*	0.0918	0.6389
BQ	0.0001*	0.0091*	0.7264	0.00006*	0.0064*	0.5856
T80	0.6259	0.6803	0.7950	0.5000	0.2728	0.4226
PEG	0.7577	0.6689	0.2744	0.9695	0.5925	0.4344
Lecithin	0.6971	0.6434	0.8293	0.5088	0.3806	0.3697

*Significant *p*-values.

The ANOVA Table 4 for the sonicated BQ formulations provides crucial insights into the statistical significance of the main effects in both the first-order and second-order models. The primary (input) variables under analysis were BQ concentration, T80, PEG, and lecithin; while the response variables examined included PdI, PSD, and ZP. The *p*-values for each variable are highlighted, indicating the statistical significance of their effects. In the first-order model, the analysis reveals several key findings. For the PdI, the model is highly significant (*p* = 0.0003), indicating that the main effects significantly explain the variability in PdI. BQ concentration is highly significant (*p* = 0.0001), suggesting that BQ is a major factor influencing PdI. In contrast, T80 (*p* = 0.6259), PEG (*p* = 0.7577), and lecithin (*p* = 0.6971) are not significant, indicating these variables do not significantly affect PdI in the first-order model.

For the PSD, the model is marginally significant (*p* = 0.0814), implying the main effects explain some of the variability in particle size. BQ concentration is significant (*p* = 0.0091), indicating it has a notable effect on particle size. However, T80 (*p* = 0.6803), PEG (*p* = 0.6689), and lecithin (*p* = 0.6434) are not significant, showing these variables do not significantly impact particle size. For ZP, the

model is not significant ($p = 0.8046$), indicating the main effects do not significantly explain the variability in the ZP. BQ concentration ($p = 0.7264$), T80 ($p = 0.7950$), PEG ($p = 0.2744$), and lecithin ($p = 0.8293$) are not significant, suggesting these variables do not significantly affect ZP.

In the second-order model, the analysis reveals more nuanced findings. For the PdI, the model is significant ($p = 0.0073$), indicating the second-order model significantly explains the variability in PdI. BQ concentration is highly significant ($p = 0.00006$), showing it is a major factor influencing PdI. T80 ($p = 0.5000$), PEG ($p = 0.9695$), and lecithin ($p = 0.5088$) are not significant, indicating these variables do not significantly affect PdI. For the PSD, the model is marginally significant ($p = 0.0918$), implying the second-order model explains some of the variability in PSD. BQ concentration is significant ($p = 0.0064$), indicating BQ has a notable effect on particle size. However, T80 ($p = 0.2728$), PEG ($p = 0.5925$), and lecithin ($p = 0.3806$) are not significant, showing these variables do not significantly impact particle size. The ZP model is not significant ($p = 0.6389$), indicating the second-order model does not significantly explain the variability in ZP. BQ concentration ($p = 0.5856$), T80 ($p = 0.4226$), PEG ($p = 0.4344$), and lecithin ($p = 0.3697$) are not significant, suggesting these variables do not significantly affect ZP.

Table 5. Comparative Analysis of Sonicated Bedaquiline formulation parameters in the linear and quadratic model.

Model	R ²	Adj R ²	Predicted R ²	Precision Ratio	Std Dev.	CV(%)	AIC	BIC
<i>Sonicated BQ Formulation for first-order Model</i>								
PdI	0.6537	0.5808	-12.2718	0.9748	0.01691	32.60	-45.68	-43.55
PSD	0.3403	0.2014	-15.8241	0.00008	11.4569	28.04	267.16	269.29
ZP	0.07806	-0.1160	-11.3790	0.0025	2.0795	28.83	185.25	187.38
<i>Sonicated BQ Formulation for Second-order Model</i>								
PdI	0.8950	0.7317	0.0111	8.5101	0.0591	16.36	18.44	-40.42
PSD	0.7907	0.4652	-0.5806	3.7789	40.8672	22.94	332.38	273.51
ZP	0.5623	-0.1184	-2.0822	1.2850	9.0742	28.87	260.14	201.28

3.3. Model summary for first-order regression for the sonicated samples.

The sonicated BQ formulation parameters analysis derived from linear and quadratic models reveals significant differences in model performance and prediction accuracy. The linear model for PdI demonstrates moderate explanatory power with an R² of 0.6537 and an Adjusted R² of 0.5808. However, the Predicted R² is highly negative (-12.2718), indicating poor predictive capability. The Precision Ratio is close to 1, suggesting low precision. The model's standard deviation is low (0.01691), but the high CV of 32.60% indicates substantial variability relative to the mean. For PSD, the linear model shows low explanatory power with an R² of 0.3403 and an Adjusted R² of 0.2014.

3.4. Model summary for second-order regression for the sonicated samples.

The quadratic model for PdI exhibits high explanatory power with an R² of 0.8950 and an Adjusted R² of 0.7317. The Predicted R² is positive (0.0111), indicating reasonable predictive capability, compared to the first-order model approach. Similarly, the Precision Ratio is higher (8.5101), reflecting good precision. Furthermore, the standard deviation (0.0591) and CV (16.36%) are both lower than those in the first-order model, indicating less variability. These results suggest the second-order model gives better predictions for the formulations.

3.5. Optimization for BQ SLN Formulations

3.5.1. Contour and response surface plots (2D and 3D RSM)

These visual tools are essential for diagnosing model fit, understanding variable interactions, and optimizing conditions to achieve the best possible outcomes. Statistics obtained from the second-order modeling were used to generate 2D and 3D RSM.

3.5.2. Effect of BQ and T80 on PdI

In Figure 1(a), the actual versus predicted residual plot shows that the observed relationship appears reasonable; supported by the high significance of the quadratic model ($p < 0.0073$). The dispersion of data (dots) around the trend line indicates a good model fit, evidenced by an R^2 value of 0.89. This suggests that changes in the input factors explain 89% of the variability in the PdI. Additionally, only two actual observations fall outside the shaded confidence interval region.

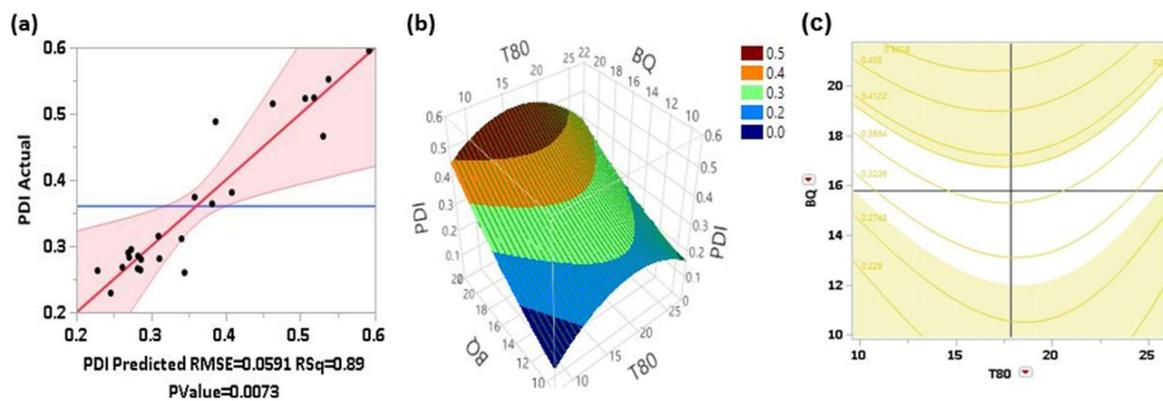


Figure 1. (a) Observed versus predicted plot for values of PdI (Y1) variable in the sonicated formulation. (b) A three-dimensional response surface plot between BQ and T80 concentrations and their combined effect on PdI of the sonicated formulation, at fixed levels of lecithin and PEG. (c) A two-dimensional contour plot depicting the effects of BQ and T80 concentrations on PdI of the sonicated formulation at constant levels of lecithin and PEG.

The response surface plot Fig 1(b) appears to be elliptical, suggesting a quadratic relationship between the input factors (BQ and T80) and the response variable (PdI). This implies the presence of an optimal region within the design space. As BQ increases, PdI initially increases to a maximum value and then decreases, indicating a quadratic effect. As T80 increases, PdI shows a similar quadratic trend, first increasing to a peak and then decreasing. The gradients show that the system is sensitive to changes in both factors, especially in the central region. This plot is useful for identifying the precise levels of BQ and T80 that maximize PdI; guiding the optimization process in formulation development. The legends in the response surface plots are critical components that provide a color-coded guide to the values of the response variable across the surface. It also visually maps the response values and aids in trend identification; optimal region detection; and data comparison.

The contour plot Fig 1(c) shows a range of PdI values, from approximately 0.229 to 0.458. There is a curvature in the contour lines, suggesting a nonlinear relationship between the variables and the response. This indicates that the effect of one variable on PdI depends on the level of the other variable. The central region of the contour plot, where the PdI values are lower (around 0.229 to 0.3206), appears to be an optimal range for both BQ and T80 concentrations. This region represents conditions where the PdI is minimized, indicating more uniform particle sizes. As one moves away from this central region, the PdI values increase, indicating less optimal conditions for the formulation. Increasing the concentration of BQ tends to increase the PdI initially, as indicated by the rising contour lines moving from left to right. Similarly, increasing the concentration of T80 also tends to increase the PdI initially, as indicated by the rising contour lines moving from bottom to top. Furthermore, the contour lines are closely packed, indicating a more sensitive relationship between

levels of BQ and T80, and the PDI. This suggests that small changes in the level of these input factors in the sonicated formulation have a substantial impact on PDI.

3.5.3. Effect of BQ and T80 on Z Average (PSD)

The plot in Fig 2(a) demonstrates that the model provides a moderate fit ($R^2 = 0.79$) for predicting the PSD of the sonicated BQ formulation, explaining a substantial portion of the variability. However, the prediction error (RMSE = 40.867) and the p-value (0.0918) suggest that while the model is fairly accurate, there is room for improvement in precision and statistical significance. The 3D and contour plots for the PSD model are shown in Figures 2 (b) and (c).

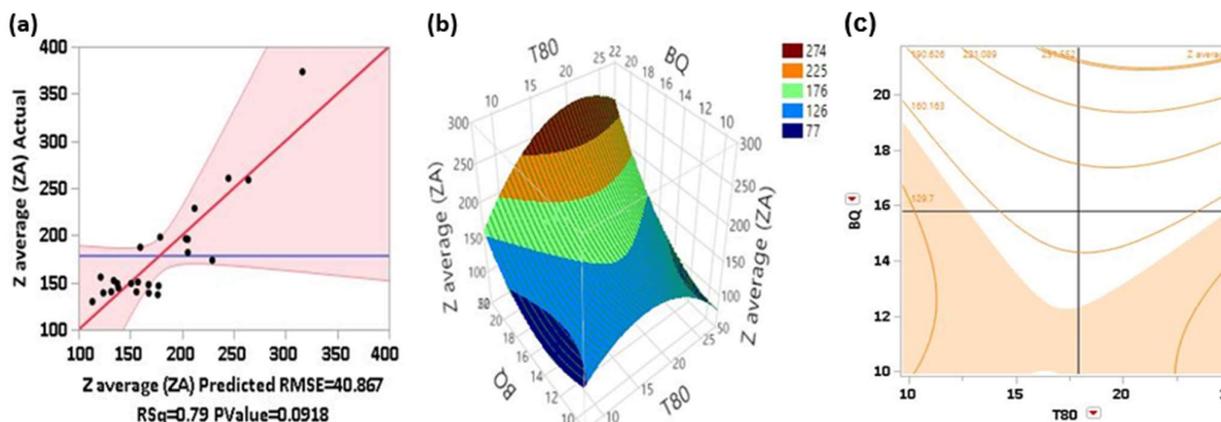


Figure 2. (a) Observed versus predicted plot for values of PSD (Y2) variable in the sonicated formulation. (b) A three-dimensional response surface plot between BQ and T80 concentrations and their combined effect on PSD of the sonicated formulation, at fixed levels of lecithin and PEG. (c) A two-dimensional contour plot depicting the effects of BQ and T80 concentrations on PSD of the sonicated formulation at constant levels of lecithin and PEG.

For the response surface plot Fig 2(b), as BQ increases, the PSD decreases sharply initially, indicating a strong influence on particle size reduction at lower concentrations. However, beyond a certain point, the decrease in PSD starts to plateau, showing diminishing returns with higher BQ concentrations. The influence of T80 on the PSD is also significant. At lower T80 concentrations, the PSD decreases, but this trend reverses at higher concentrations, likely due to micelle formation and increased surfactant layer thickness, which could increase particle size. The optimal region for the PSD is at the lower front corner of the plot, where both BQ and T80 concentrations are lower. This area corresponds to the smallest PSD values, indicating the most favorable conditions for achieving minimal particle size.

In the contour plot Fig 2(c), the highest PSD values (greater than 281.089) occur at the top of the plot, particularly where T80 is high and BQ is low. The lowest PSD values (less than 129.7) are at the bottom left, where BQ and T80 are low. The optimal region, where PSD is maximized, is indicated by the outermost contour lines at the top of the plot. This suggests that the highest PSD values occur when BQ is approximately 18-20 and T80 is approximately 20-25.

3.5.4. Effect of BQ and T80 on ZP

The R^2 value of approximately 0.5765 indicates that the second-order model predicts only about 58% of the variability in the actual ZP of the formulation. This suggests a moderate level of predictive power. The RMSE of approximately 9.1 indicates that on average, the model's predictions deviate from the actual ZP values by about 9.1 mV. Figure 3 shows the 3D response surface and 2D contour plots that explain the effect of BQ and T80 on the ZP at fixed levels of lecithin and PEG concentration.

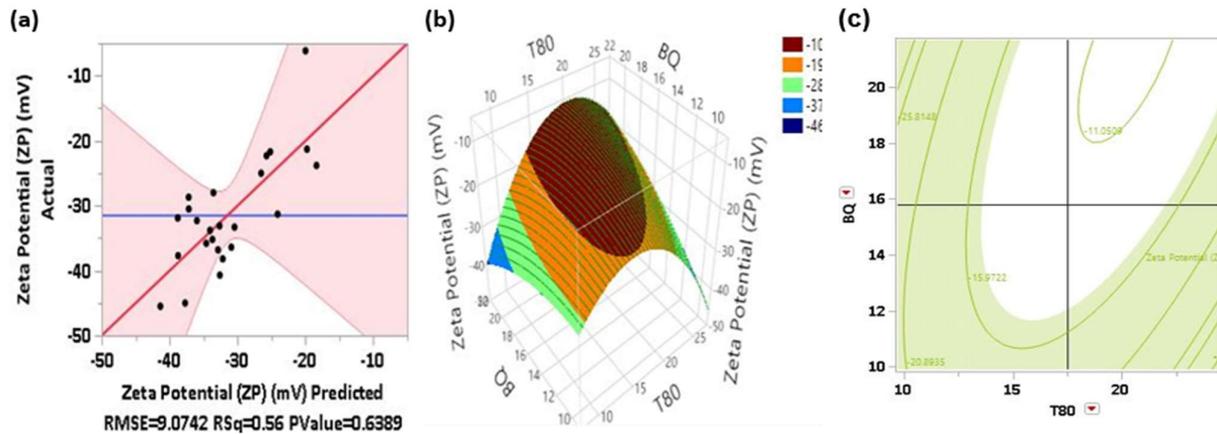


Figure 3. (a) Observed versus predicted plot for values of ZP (Y_3) variable in the sonicated formulation. (b) A three-dimensional response surface plot between BQ and T80 concentrations and their combined effect on ZP of the sonicated formulation, at fixed levels of lecithin and PEG. (c) A two-dimensional contour plot depicting the effects of BQ and T80 concentrations on ZP of the sonicated formulation at constant levels of lecithin and PEG.

The response surface plot Fig 3(b) shows a circular pattern around the peak region, indicating a local maximum for ZP. As BQ concentration increases from 10 to around 20 units, ZP values initially increase, reaching a peak. However, beyond this point, further increases in BQ concentration lead to a slight decrease in ZP. Similarly, an increase in T80 concentration from 10 to around 20 units increases ZP; reaching a peak value. The optimal region for ZP is around 20 units for both BQ and T80 concentrations, where ZP reaches its peak value. This region is marked by the highest contour levels (dark brown).

The contour plot Fig 3(c) shows that higher values of ZP (less negative) are observed in the regions where T80 is between 20 and 25, and BQ is between 14 and 18. Lower ZP values (more negative) are seen where T80 is lower and BQ is between 10 and 12. The optimal region, where ZP is less negative (around -11.059), is located where T80 is high (above 20) and BQ is around 18. Conversely, the most negative ZP values (around -39.995) are seen where T80 is low (around 10) and BQ is moderate (between 12 and 16).

3.6. Graphical Optimization

Optimization of the developed sonicated BQ SLN formulation was performed to find the levels of concentration of BQ (X_1), T80 (X_2), PEG (X_3), and Lecithin (Lec) (X_4). After plotting the desired values of dependent and independent variables on JMP software, it suggested that for the factors X_1 , X_2 , X_3 , and X_4 , the values were 19.43, 25.21, 39.20, and 200.00, and PdI (Y_1), PSD (Y_2) and ZP (Y_3) were 0.41, 250.99 μ m, and -25.95mV, respectively, see Figure 4.

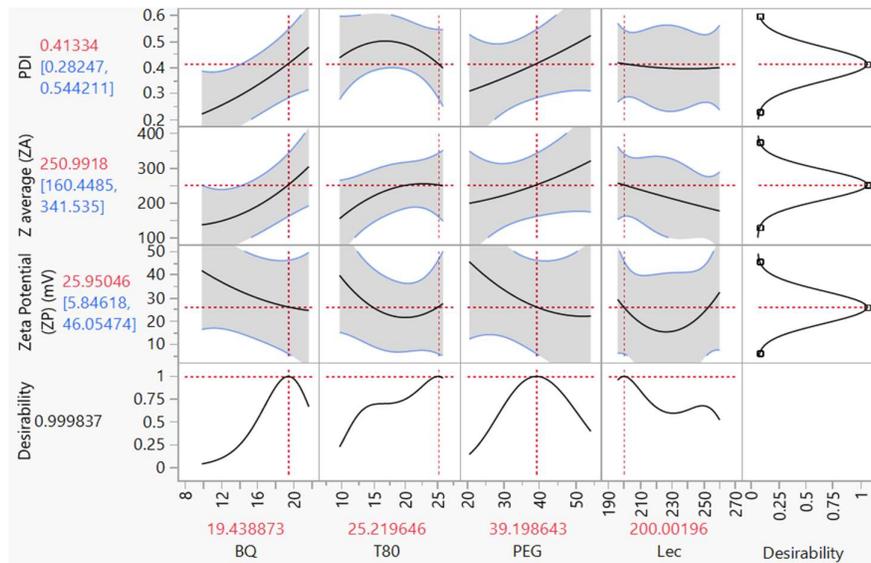


Figure 4. JMP software model profiler for sonicated BQ SLN formulations suggesting levels of experimental variables that can be used in defining product design space.

A statistical profiler tool predicts the optimal levels of input factors required to achieve output variables that provide a design space that maintains the CQAs of a target formulation. The JMP software model profiler for the second-order approach generated a high desirability level for the sonicated BQ formulations. The desirability value of 0.9998 indicates that the conditions under which this value is obtained are almost perfect in meeting the desired criteria. This suggests that the BQ SLN formulation or process conditions are nearly optimal, with the levels of the input factors set to produce response variables that almost perfectly align with the desired targets. With such a desirability value, further optimization may be minimally needed, as the current conditions yield optimal performance. The high level of desirability reflects strong confidence in the process or formulation, indicating excellent model fit and predictive capability. It also implies that the trade-offs between different response variables have been effectively balanced, resulting in a robust design space where small variations in input factors will not significantly affect the outcomes.

In designing a preliminary design space for the BQ SLN formulations, the overlay plot shown in Figure 5 demonstrated a white region representing the area where optimal conditions for all three response variables overlap. The design space region is bounded by PSD, PDI, and ZP. The respective values are $110.6\mu\text{m}$ (indicating smaller and more desirable particle sizes); 0.39 (showing a more uniform particle size distribution); and -19.87mV (indicating moderate stability); which are generally acceptable for the formulation. Moreover, it also illustrates that the optimal formulation is achieved at a T80 concentration of 15-20% and a BQ concentration of 14-18%. This range exhibits high ZP and optimal PSD, indicating stable and appropriately sized BQ SLN formulations. As T80 increases beyond 18%, ZP decreases, signifying reduced particle stability. The design space is robust within these concentrations but exhibits sensitivity to changes, with small deviations potentially leading to significant impacts on PSD and stability.

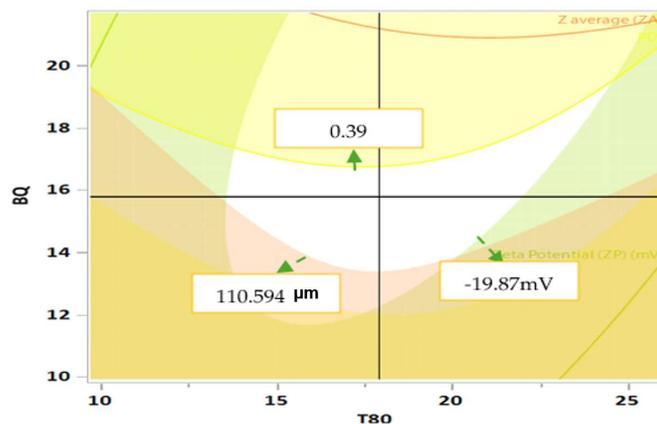


Figure 5. Overlaid Contour Plots: The optimum area (white region) suggests recommended design space parameters for BQ SLN formulations.

Therefore, maintaining concentrations within the optimal range defined in the design space is crucial to ensure the desired stability and performance of the BQ SLN formulations. Future steps will validate and standardize the optimized conditions to ensure consistent high-quality outcomes.

4. Discussion

BQ SLNs were generated from sonicated and unsonicated formulations. Based on the results obtained, the final regression model analyses focused on the sonicated formulations. The first-order model (linear regression) captures only linear relationships with fewer parameters; making it easier to interpret. In contrast, the second-order model (quadratic regression) includes linear, quadratic, and interaction terms; offering more flexibility to capture complex relationships and curvature. However, this requires more data and is therefore more complex. This study compared the performance matrices of these two regression equations to evaluate the model performance of the individual response variables (CQAs) for the BQ SLN formulations. This comparison aimed to identify which regression analysis—first or second—yielded models with superior characteristics in terms of fitness, sensitivity to variability, complexity, and prediction consistency. It also helps determine which model better explains the variability in the data and provides a more accurate representation of the underlying relationship between the variables. CCD was used to fit the quadratic models in this study. It is useful in RSM where the goal is to optimize a response influenced by several variables. Furthermore, CCD helps explore the design space and fit a quadratic model [9, 15].

4.1. Model fitting using first-order polynomials for sonicated and unsonicated samples

Several metrics were used in evaluating the adequacy of the statistical models used in this study; including R^2 , adjusted R^2 , predicted R^2 , Precision Ratios, standard deviation, CV, AIC, and BIC. These metrics assessed the model fit, complexity, and predictive performance. Each metric plays a crucial role in determining regression model adequacy. CV provides a measure of relative variability and helps compare model consistency. AIC and BIC are used for model selection, with AIC balancing fit and complexity and BIC imposing a stronger penalty for additional parameters [9, 17]. For the unsonicated formulations, the R^2 values were relatively low, indicating limited explanatory power. High RMSE and CV values also suggest substantial prediction errors and high variability [17, 19].

4.2. Model summary for first- and second-order regression for the sonicated samples.

Although the R^2 value for PdI was lower in the first compared to the second (0.6537 vs 0.8950), both regression models displayed a fair prediction; but showed varied results for the other two output variables. The highly negative Predicted R^2 (-15.8) from the first-order model for PSD indicates poor predictive performance. The Precision Ratio is extremely low (8×10^{-5}), and the standard

deviation is high (11.5), reflecting significant dispersion. The high CV shows considerable variability; similarly, high AIC and BIC values indicate a poor fit [20]. The linear model for ZP performs poorly with a very low R^2 (0.08) and a negative Adjusted R^2 (-0.12); indicating first-order modeling does not explain the variability well. Likewise, this negative Predicted R^2 also reflects poor predictive ability. The low Precision Ratio (0.003) and high standard deviation (2.1) suggest low precision and high variability. The AIC and BIC values are relatively high, indicating a poor model fit.

In contrast, the second-order models show improved performance. For the PSD response, the quadratic model has $R^2 = 0.7907$, and a predicted R^2 of -0.5806; which, while still negative, represents an improvement over the linear model. The Precision Ratio is higher (3.7789), and the lower CV indicated improved precision. The ZP model shows moderate improvement in $R^2 = 0.5623$ over the value from the first order. Although the predicted R^2 is still negative (-2.0822), the Precision Ratio is higher (1.2850), and the standard deviation is reduced (9.1). Although the quadratic model for PdI demonstrates high explanatory power with R^2 (0.8950) suggesting a clear and reliable signal [20], the presence of high Precision Ratios indicates that while the model performs well, further validation is required to ensure its predictive reliability [21].

4.3. Effects of input variables in first and second-order model sonicated BQ SLN formulation

The analysis of variance for both first-order and second-order models provides essential insights into sonicated BQ formulations. In both models, BQ concentration consistently shows significant effects on the PdI and PSD. This indicates that BQ plays a crucial role in determining the quality and characteristics of the nanoparticles. However, T80, PEG, and lecithin do not show significance suggesting their roles are less critical in these specific formulations or that their optimal concentrations might have already been achieved. Furthermore, the second-order model provided a better fit for the PdI, highlighting the importance of considering nonlinear effects and interactions in the formulation process. Overall, optimizing BQ concentration will yield significant improvements in PdI and PSD and a higher likelihood of achieving the desired quality attributes for the BQ SLN formulations.

4.4. Significance of the second-order polynomial model equations

The regression analysis of sonicated BQ formulations reveals crucial insights into how input variables affect key responses. T80 and BQ concentrations significantly influence the PdI and PSD in the sonicated formulations. Specifically, T80 decreases PdI due to its surfactant properties, and bedaqualine reduces particle size. Significant positive interaction terms between BQ and T80, and between T80 and PEG, indicate synergistic effects in reducing particle size. In the sonicated formulation, increasing concentrations of both BQ and T80 decrease ZP, potentially affecting nanoparticle stability.

4.5. Justifying the use of the Quadratic model in RSM optimization

The study compared linear and quadratic models in analyzing sonicated BQ formulations. The linear model showed weaknesses in explaining and predicting key characteristics like PSD. An important aspect of model adequacy is the statistical significance of its predictors. In the first-order models evaluated, only BQ showed significant effects on response variables in either of the formulations, while other predictors did not show statistically significant effects. This is evident from the high p -values associated with other predictors, which are all greater than 0.05. This lack of significance further underscores the inadequacy of the first-order models in explaining the variability of the response factors. Lastly, the evaluation of first-order models for the BQ formulation optimization reveals significant deficiencies. These models fail to adequately explain the variability in response variables, exhibit poor predictive performance, and lack statistically significant predictors. Overall, the quadratic (second-order) model is superior to the linear (first-order) model for the sonicated BQ formulation. It provides higher explanatory power, better predictive capability, improved precision, and lower variability across all response variables (PdI, PSD, and ZP), making

it the better model for optimization purposes. The result agrees with findings made in a study by Kura Beyene, where the quadratic model equation exhibited an RMSE of 0.1596, while the linear model equation demonstrated a significantly lower RMSE of 0.0747. The results reported the quadratic model provides superior accuracy and reliability in its predictions compared to the linear model [22]. To overcome these limitations, the study opted for a second-order model which was believed to ensure better efficacy and stability of the BQ formulations by providing a more accurate and reliable framework for optimization [16]. The model performed significantly better. It captured non-linear relationships between variables, allowing for more precise optimization of the formulation. This ultimately leads to better drug delivery through achieving the desired particle size and stability.

4.6. Optimization using 2D and 3D RSM

The 2D and 3D plots create visual representations to illustrate the relationship between input and response variables. The 2D RSM or contour plot two-dimensional graphs, typically show the effect of two independent variables on a response variable while keeping other variables constant. On the other hand, the 3D or response surface plot plots show the interaction between two or more independent variables and their combined effect on the response variable [10].

4.6.1. Effect of BQ and T80 on PdI

The sonication of the BQ SLN formulation process enhances particle size uniformity, thereby improving the overall quality of the formulation. The response surface analysis identifies a central region with lower PdI values, indicating uniform particle sizes crucial for consistent drug delivery. Sonication breaks down aggregates, resulting in a homogeneous mixture. The process shows a nonlinear, quadratic relationship between BQ, T80, and PdI, highlighting the need to optimize these concentrations. The quadratic model, with a high R-squared value of 0.65 and a significant p-value (<0.0073), indicates precise control is essential. The process's sensitivity to BQ and T80 changes ensures stability and efficacy. The strong model fit and predictive accuracy supports the reproducibility, reliability, and scalability of the sonicated BQ SLN formulations.

4.6.2. Effect of BQ and T80 on Z Average (PSD)

The 2D contour plot and 3D response surface explain the maximum PSD as a function of BQ and T80 concentration at a constant level of lecithin and PEG for BQ SLN formulations. Considering that at a lower concentration of T80, it was seen that the PSD initially increases significantly, attains maximum effect at values 16, and then levels off at a lower concentration. As the BQ concentration rises from 10 to 20, the quadratic effect for BQ is positive, indicating a convex curvature, meaning the decrease in PSD with increasing levels of BQ starts to level off at higher concentrations. Furthermore, the increase in PSD with higher T80 concentration post-sonication can be attributed to micelle formation, increased surfactant layer thickness, and potential reduction in sonication efficiency.

4.6.3. Effect of BQ and T80 on ZP

The reduction in ZP at a T80 medial concentration value (e.g. 18) can be attributed to the effective adsorption of surfactant molecules onto the lipid particle surface, enhancing stability through steric and possibly electrostatic mechanisms [23]. However, beyond a certain concentration, the formation of micelles and the saturation of the particle surface with surfactant molecules result in a decrease in ZP due to reduced surface charge, destabilizing interactions, and changes in the electrical double layer around the particles. Higher stability is generally believed to be associated with absolute ZP values >30 mV [24]. The optimal region for ZP is around 20 units for both BQ and T80 concentrations, where ZP reaches its peak value. This region is marked by the highest contour levels (dark brown in plot b of Figure 3), indicating the maximum ZP. This region indicates the levels of the input variables (T80 and BQ) that would yield maximum stability for the BQ SLN formulations [23].

4.7. Graphical optimization to define a design space for the BQ SLN formulation.

An overlay plot in the design space is a powerful visualization tool in DOE and RSM for optimizing processes involving multiple responses. It helps to identify the feasible region where all the specified criteria are met, enabling more informed and efficient decision-making in experimental design and process optimization [25-27]. Statistical tool profilers can be used to relate the feasible region in an overlay plot to optimal settings. Such prediction profilers use a systematic approach to identify feasible regions, interactively adjust variables, and fine-tune settings to generate acceptable ranges to create a design space. This integration ensures a comprehensive understanding and optimization of the design space, leading to robust and efficient process optimization [27]. In this study, a preliminary design space was created for the BQ SLN formulation from estimates from the prediction profiler obtained from the JMP software. By linking the feasible region in an overlay plot with the optimal settings identified in the prediction profiler, one can generate acceptable ranges within the design space where the BQ SLN formulation can consistently maintain its desirable CQAs. This integration ensures a comprehensive understanding and optimization of the design space, leading to robust and efficient process optimization.

4.8. Future studies

It is envisaged that further validation using the Monte Carlo simulation will be performed when larger scales of the BQ SLN formulations are made. The Stochastic optimization (Monte Carlo simulation) involves defining the design space, developing a DOE model, and running simulations to assess the variability and reliability of the experimental outcomes. This approach helps identify risk areas and informs decisions for process optimization by comparing simulation results with experimental data to ensure consistency and robustness [28].

4.9. Study limitations

This study represents our ongoing effort to discover the most effective approach for optimizing the process of formulating the BQ SLNs. We recognize that the initial models applied to optimize these formulation batches may not fully capture the complexities of the data. For example, it was observed that a reduced model was generated for the y variables when the study dataset was simulated using the k-fold cross-validation technique. This infers stronger predictive performance and optimal fit for the various fixed statistics or optimality criteria. In future studies, particularly during the scale-up stage, we plan to implement regularization techniques such as k-fold cross-validation or Lasso regression [29, 30]. These methods will help refine our models and achieve more effective outcomes with optimal fit statistics.

5. Conclusions

The study highlights the importance of model adequacy in assessing the predictive performance of regression models. The findings indicate that while first-order models may have limited explanatory power, the quadratic model significantly captures non-linear relationships between variables, allowing for more precise optimization of formulations. To overcome these limitations, the study opted for a second-order model which was believed to ensure better efficacy and stability of the BQ formulations by providing a more accurate and reliable framework for optimization. The study demonstrates that for optimal solid lipid nanoparticle formulation, maintaining BQ concentration between 10% to 20% and T80 concentration between 15% to 18% is crucial. This balance maximizes ZP and minimizes PdI; the ZP, ensuring particle stability and appropriate size distribution. The high precision ratio (8.5101) and significant model fit ($p < 0.0073$ for PdI) validate the reliability of these findings. The results highlight the importance of careful adjustment of BQ and T80 concentrations to achieve desired nanoparticle characteristics and optimize the formulation process for effective drug delivery systems.

Author Contributions: C.C.U.: conceptualization, data curation, formal analysis, investigation, methodology, writing - original draft. M.A.O.: conceptualization, formal analysis, investigation, methodology, project administration, supervision, writing - original draft. I.M.A.: data curation, formal analysis, methodology, visualization, writing - original draft. S.R.B.: conceptualization, funding acquisition, project administration, resources, supervision, writing - review.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: We acknowledge support from the Purdue University Industrial and Physical Pharmacy Department and the Biotechnology Innovation and Regulatory Science Program (BIRS); Kari Clase, Zita Ekeocha, Uche Sonny-Afoekelu, and Dan Smith from the same institution. Ashlee Brunaugh from the Department of Pharmaceutical Sciences, University of Michigan, Ann Arbor, USA. Michigan State University, School of Packaging.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Okezue, M.A., et al., *A Quality by Design Approach for Optimizing Solid Lipid Nanoparticles of Bedaquiline for Improved Product Performance*. AAPS PharmSciTech, 2024. **25**(6): p. 152.
2. WHO. *Global Tuberculosis Report 2023*, Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. 2023 [cited 2024 April 21]; Available from: https://reliefweb.int/report/world/global-tuberculosis-report-2023?gad_source=1&gclid=Cj0KCQjw8pKxBhD_ARIsAPrG45nT5QmUp6oan9ISNjIhoCUeTFNJLY6TPHYvbf8hRdGHCT9byb64Bp8aAmTnEALw_wcB.
3. Elbrink, K., et al., *Application of solid lipid nanoparticles as a long-term drug delivery platform for intramuscular and subcutaneous administration: In vitro and in vivo evaluation*. European Journal of Pharmaceutics and Biopharmaceutics, 2021. **163**: p. 158-170.
4. Hassan, H., et al., *Central composite design for formulation and optimization of solid lipid nanoparticles to enhance oral bioavailability of acyclovir*. Molecules, 2021. **26**(18): p. 5432.
5. Subramaniam, B., Z.H. Siddik, and N.H. Nagoor, *Optimization of nanostructured lipid carriers: Understanding the types, designs, and parameters in the process of formulations*. Journal of nanoparticle research, 2020. **22**: p. 1-29.
6. Sahu, J.N., J. Acharya, and B.C. Meikap, *Response surface modeling and optimization of chromium (VI) removal from aqueous solution using Tamarind wood activated carbon in batch process*. Journal of hazardous materials, 2009. **172**(2-3): p. 818-825.
7. Luiz, M.T., et al., *Design of experiments (DoE) to develop and to optimize nanoparticles as drug delivery systems*. European Journal of Pharmaceutics and Biopharmaceutics, 2021. **165**: p. 127-148.
8. Bhattacharya, S., *Central composite design for response surface methodology and its application in pharmacy, in Response surface methodology in engineering science*. 2021, IntechOpen.
9. Myers, R.H., D.C. Montgomery, and C.M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. 2016: John Wiley & Sons.
10. Sarrai, A.E., et al., *Using central composite experimental design to optimize the degradation of tylosin from aqueous solution by photo-fenton reaction*. Materials, 2016. **9**(6): p. 428.
11. Lamidi, S., et al., *Applications of response surface methodology (RSM) in product design, development, and process optimization*. 2022: IntechOpen.
12. Montgomery, D.C., *Design and analysis of experiments*. 2017: John Wiley & Sons.
13. Bushra, R., et al., *Formulation design and optimization of aceclofenac tablets (100 mg) using central composite design with response surface methodology*. Lat. Am. J. Pharm, 2014. **33**(6): p. 1009-1018.
14. SAS Institute Inc. JMP Pro 15.1.0 (426298). JMP Statistical Discovery From SAS 2019 [cited 2022 20 January]; Available from: <https://www.jmp.com>.
15. Behera, S.K., et al., *Application of response surface methodology (RSM) for optimization of leaching parameters for ash reduction from low-grade coal*. International Journal of Mining Science and Technology, 2018. **28**(4): p. 621-629.
16. Çaydaş, U. and A. Hasçalik, *Modeling and analysis of electrode wear and white layer thickness in die-sinking EDM process through response surface methodology*. The International Journal of Advanced Manufacturing Technology, 2008. **38**: p. 1148-1156.

17. Eid, A.M., N.A. Elmarzugi, and N.A. Jaradat, *Influence of sonication and in vitro evaluation of nifedipine self-nanoemulsifying drug delivery system*. Brazilian Journal of Pharmaceutical Sciences, 2019. **55**: p. e17497.
18. Mohammed, E.A., C. Naugler, and H. Behrouz, *Far. Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics*. Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology: Algorithms and Software Tool, 2015: p. 577-602.
19. Adamo, L., et al., *A Response Surface Methodology Approach to Develop a Multiphysics Simulation Model of a Tensile Friction Test*. Engineering Proceedings, 2022. **26**(1): p. 22.
20. Freckleton, R.P., *Special feature: 5th anniversary of Methods in Ecology and Evolution*. 2016, Wiley Online Library. p. 634-635.
21. Sadhukhan, B., N.K. Mondal, and S. Chatteraj, *Optimisation using central composite design (CCD) and the desirability function for sorption of methylene blue from aqueous solution onto Lemna major*. Karbala International Journal of Modern Science, 2016. **2**(3): p. 145-155.
22. Beyene, K.A., *Comparative study of linear and quadratic model equations for prediction and evaluation of surface roughness of a plain-woven fabric*. Research Journal of Textile and Apparel, 2023. **27**(2): p. 281-298.
23. Brunaugh, A.D., H.D.C. Smyth, and R.O. Williams Iii, *Essential pharmaceuticals*. 2019: Springer.
24. Malvern Instruments pdf. *Zeta potential - An introduction in 30 minutes*. 2015 [cited 2024 January 25]; Available from: <https://www.research.colostate.edu/wp-content/uploads/2018/11/ZetaPotential-Introduction-in-30min-Malvern.pdf>.
25. Corrie, L., R. Gundaram, and L. Kukatil, *Formulation and evaluation of Cassia tora phytosomal gel using central composite design*. Recent Innovations in Chemical Engineering (Formerly Recent Patents on Chemical Engineering), 2021. **14**(4): p. 347-357.
26. Kasemiire, A., et al., *Design of experiments and design space approaches in the pharmaceutical bioprocess optimization*. European Journal of Pharmaceutics and Biopharmaceutics, 2021. **166**: p. 144-154.
27. Manzon, D., et al., *Quality by design: comparison of design space construction methods in the case of design of experiments*. Chemometrics and Intelligent Laboratory Systems, 2020. **200**: p. 104002.
28. Lievense, R., *Pharmaceutical Quality by Design Using JMP: Solving Product Development and Manufacturing Problems*. 2018: Sas Institute.
29. Jung, Y., *Multiple predicting K-fold cross-validation for model selection*. Journal of nonparametric statistics, 2018. **30**(1): p. 197-215.
30. Roberts, S. and G. Nowak, *Stabilizing the lasso against cross-validation variability*. Computational Statistics & Data Analysis, 2014. **70**: p. 198-211.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.