Journal of Pharmaceutical Sciences xxx (2016) 1-9

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Research Article

Predicting the Fine Particle Fraction of Dry Powder Inhalers Using Artificial Neural Networks

Joanna Muddle [1, 2], Stewart B. Kirton [3, *], Irene Parisini [3], Andrew Muddle [4], Darragh Murnane [3], Jogoth Ali [3], Marc Brown [3, 4], Clive Page [2], Ben Forbes [1]

[1] Institute of Pharmaceutical Science, King's College London, 150 Stamford Street, London SE19NH, UK
[2] Sackler Institute of Pulmonary Pharmacology, Institute of Pharmaceutical Science, King's College London, 150 Stamford Street, London SE19NH, UK
[3] Department of Pharmacy, University of Hertfordshire, Hatfield, Herts AL109AB, UK
[4] MedPharm Ltd, R&D Centre, Units 1 and 3 Chancellor Court, 50 Occam Road, Surrey Research Park, Guildford GU27AB, UK

ARTICLE INFO

Article history:
Received 9 July 2016
Revised 26 September 2016
Accepted 6 October 2016

Keywords:
artificial neural networks
in silico modeling
dry powder inhaler
next-generation impactor
in vitro performance
fine particle fraction

ABSTRACT

Dry powder inhalers are increasingly popular for delivering drugs to the lungs for the treatment of respiratory diseases, but are complex products with multivariate performance determinants. Heuristic product development guided by *in vitro* aerosol performance testing is a costly and time-consuming process. This study investigated the feasibility of using artificial neural networks (ANNs) to predict fine particle fraction (FPF) based on formulation device variables. Thirty-one ANN architectures were evaluated for their ability to predict experimentally determined FPF for a self-consistent dataset containing salmeterol xinafoate and salbutamol sulfate dry powder inhalers (237 experimental observations). Principal component analysis was used to identify inputs that significantly affected FPF. Orthogonal arrays (OAs) were used to design ANN architectures, optimized using the Taguchi method. The primary OA ANN r^2 values ranged between 0.46 and 0.90 and the secondary OA increased the r^2 values (0.53-0.93). The optimum ANN (9-4-1 architecture, average r^2 0.92 $\pm$ 0.02) included active pharmaceutical ingredient, formulation, and device inputs identified by principal component analysis, which reflected the recognized importance and interdependency of these factors for orally inhaled product performance. The Taguchi method was effective at identifying successful architecture with the potential for development as a useful generic inhaler ANN model, although this would require much larger datasets and more variable inputs.

Introduction

Dry powder inhalers (DPIs) are increasingly popular for delivering drugs to the lungs for the treatment of respiratory diseases. Over the last 2 decades, device and formulation technologies have developed considerably, along with the scientific understanding of the determinants of inhaler performance.[1-7] *In vitro* impactor deposition is the principal pharmaceutical performance assay and is an essential assessment during the development and registration of new originator or generic DPI formulations.[8] The fine particle fraction (FPF) determined using *in vitro* deposition techniques has been used as a DPI performance characteristic for mechanistic modeling,[6] *in vitro-in vivo* correlation,[9] and to make estimations of clinical relevance.[5] FPF is defined as the proportion of the particles that are <5 μm in diameter, that is, the respirable dose. Modeling the influence of formulation and device variables on FPF could facilitate the development of the next generation of inhaled medicines[10] and the matching of test inhalers to reference products during the development of generic products.[8,11]

As the patents for inhaled drugs expire, many companies are expending much effort in developing generic versions of innovator inhaled medicines. The United States Food and Drug Administration and European Medicine Agency have clear processes by which a generic orally inhaled product can be developed for the market,[12] both of which include extensive *in vitro* studies.[13,14] The United States Food and Drug Administration recently issued specific guidance for salmeterol-fluticasone combinations aiming for registration as a generic Advair[15,16] including the requirement to match FPF. For generic product development, the matching of *in vitro* performance is the obvious first stage in product development. However, heuristic development of a product guided by

* *Correspondence to:* Stewart B. Kirton (Telephone: ■■■; Fax: ■■■).
E-mail address: s.b.kirton3@herts.ac.uk (S.B. Kirton).

http://dx.doi.org/10.1016/j.xphs.2016.10.002
0022-3549/

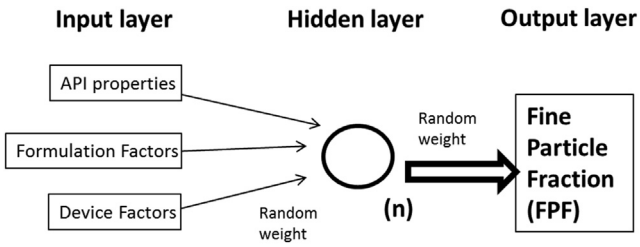

Figure 1. A simplified structure for a supervised ANN showing the feedforward links from the input layer to one neuron of the hidden layer which subsequently feeds onto the output layer (prediction of FPF). Input layer elements are the properties of API, formulation factors, and device factors. The output layer is product performance in terms of FPF.

in vitro FPF is a costly and time-consuming process. There has already been interest in using *in silico* models to streamline this process and there is a belief that modeling may become increasingly crucial in the design and development of inhaled drug products.[17] Linear regression models have been used to understand the complex formulation characteristics of DPIs,[6,18] for example, Kinnunen et al.[19] showed a best r^2 value of 0.861 between the physical properties of the carrier (amount of fines which are <4.5 μm in size) and the *in vitro* performance of the formulation, for example, FPF. Regression models make the assumption that relationships between physiochemical factors are linear. As such, prediction of pharmaceutical response based on these polynomial relationships is limited and can provide an incomplete estimate of the observed response in experimental determinations. Therefore, there is a need to investigate non-linear models as an alternative approach to determine whether or not this improves model quality.[20] The use of artificial neural networks (ANNs) is one such technique that could be exploited.

ANNs are computational models that replicate the way the human brain processes information, by linking input variables to a desired output via the selective activation of artificial "neurons" in a complex network (Fig. 1).[21] The ANN is typically separated into 3 sections: the input layer, the hidden layer, and the output layer. The input layer contains information that will be used as stimuli for the ANN, such as experimental variables. The elements of the input layer are connected to the neurons of the hidden layer via links. Each link has a weight value (w) associated with it, and each neuron receives numerical input from each of the elements of the input layer (element value × link weight). The hidden layer neuron can exist in one of the 2 states: inactive or active. Whether or not a neuron is active or inactive is determined by an activation threshold function. If the sum of the numerical inputs from the elements of the input layer exceeds the threshold value then the neuron is activated. Activated neurons then communicate with the output layer. Each neuron in the hidden layer is connected via a link (with an associated weight, w) to each element in the output layer. The elements in the output layer define one of a possible number of permissible outcomes from the input. The output value(s) selected by an ANN is a function of the neurons activated in the hidden layer.

Supervised ANNs, such as the ones used in this investigation, work by training the network to generate a desired output from input variables using an iterative process of calculating and minimizing the error between the generated and expected output value.[22] Error is minimized by manipulating the weights associated with the links between the 3 layers of the ANN via a process known as backpropogation.[23]

ANNs are powerful pattern recognition tools and have been utilized for analyzing inhaled drug delivery. Nazir et al.[24,25] produced ANNs for predicting the regional and total aerosol deposition in the human lungs and De Matas et al.[26-28] produced ANNs that predicted deposition and clinical effects for pulmonary drug delivery (improvements in the forced expiratory volume in 1 s and urinary excretion of the drug in 24 h). To create an optimal ANN architecture, many models need to be created and tested, which can be challenging and time-consuming. Hence, any techniques that can help to reduce the number of architectures that need to be investigated are useful with respect to streamlining the development process. As such, the Taguchi method has been widely used in formulation development[29,30] and in designing the architecture of ANNs.[31-33] Design of experiments has also been used to evaluate the effect of lactose size fractions on the performance of the dry powder formulation,[2] but has not been used to design an optimum ANN predicting the *in vitro* performance of a DPI. The objective of this study was to investigate the feasibility of using ANNs to predict FPF based on formulation device variables. The Taguchi method, which has not previously been applied to ANN to predict the FPF of DPI, was used to develop an optimized ANN.

Methods

Materials

Unscrambler™ was obtained from CAMO Software (Oslo, Norway), while Minitab™ was from On Line Computers (York, UK). Neurosolutions™ was from Neuro-dimension (Gainsville, USA).

Experimental Data for Modeling

The self-consistent dataset used in this study was assembled from studies into the effect DPI formulation factors on FPF reported by Hassoun et al.,[34] Muddle et al.,[35] and Parisini[36] (Table 1) where we define a self-consistent dataset as experimental data for a range of compounds and/or formulation types that has been obtained using identical experimental procedures. Two of these studies, investigating the effect of the carrier lactose particle, total fine lactose content and device resistance on FPF of salmeterol xinafoate[34] and salbutamol sulfate,[35] were performed in the same manner to allow direct comparison. In brief, the active pharmaceutical ingredient (API) was blended with 3 coarse lactose grades (Respitose® ML001, Respitose ML006, and Lactohale® LH200) and different amounts of fine lactose were added (0%-20%) (Lactohale LH300). These formulations were tested using a next-generation pharmaceutical impactor to measure their *in vitro* deposition with 3 different inhalers (Aeroliser®, Handihaler®, and Rotahaler®). The third study[36] investigated salmeterol xinafoate and salbutamol sulfate blended with a different coarse lactose (Respitose SV003) ± addition of 5% fine lactose. These 4 formulations were then tested using a Cyclohaler® at 2 different pressure drops (2 and 4 kPa). Thus, the combined dataset had 2 APIs (salmeterol xinafoate and salbutamol sulfate) tested with 4 different inhalers (Aeroliser, Handihaler, Rotahaler, and Cyclohaler) and 4 different coarse lactoses blended with a variety of different fine lactose ratios to allow a large range of total fines content to be examined (Table 1).

To allow the datasets to be combined as described above, the device factors, resistance, test flow rate (Q), and pressure drop (P) were converted to a value for the power generated by simulated "inhalation" through the inhaler at each flow rate using an established method[37] (Eq. 1). This enabled the 3 original datasets to be combined by overcoming the problem that the 3 inhalers were originally tested at different flow rates.

$$Power = P.Q. \quad (1)$$

The records in the combined dataset (237 records) were split into a training set (60% of the records), a cross validation (CV, 20%

Table 1
Datasets for Salmeterol Xinafoate and Salbutamol Sulfate Dry Powder Inhaler Fine Particle Fraction Versus Formulation Device Variables, Showing the Input Variables and Number of Records for the Individual Datasets that Were Emerged to Form the Combined Dataset Used to Develop ANN

Dataset	ANN Input			ANN Output
	API	Device	Formulation	No. of Records
Hassoun et al.[34]	Salmeterol (0.58% wt/wt)	Aeroliser, Handihaler, Rotahaler	Coarse lactose (Respitose ML001, ML006, Lactohale LH200) Fine lactose 0%-20%	84
Muddle et al.[35]	Salbutamol (0.58% wt/wt)	Aeroliser, Handihaler, Rotahaler	Coarse lactose (Respitose ML001, ML006, Lactohale LH200) Fine lactose 0%-15%	120
Parisini[36]	Salbutamol and salmeterol (1.48% wt/wt)	Cyclohaler	Coarse lactose (Respitose SV003) Fine lactose 0%-5%	33
Combined dataset	Salbutamol and salmeterol (0.58% wt/wt and 1.48% wt/wt)	Aeroliser, Handihaler, Rotahaler, Cyclohaler	Coarse lactose (Respitose ML001, ML006, SV003; Lactohale LH200) Fine lactose 0%-20%	237
	Melting point; critical primary pressure of dispersion	Power	Total lactose fines; d_{v10}; d_{v50}; d_{v90}; intrinsic fines	FPF

Factors below the dotted line in the combined dataset are the inputs and outputs used in the optimized model.

of the records), and test set (20% of the records). Neurosolutions for Excel was used to allocate records to the training, test, and validation sets. The records with the maximum and minimum experimental FPF values were allocated to the training set. The remainder of the records was randomly allocated to the training, validation, and test sets, using the random allocation tool in Neurosolutions for Excel.

Principal Component Analysis

Principal component analysis (PCA) was carried out in Unscrambler X. The input variables were total fines content, concentration of fine lactose added, intrinsic fines content, mean drug content, particle size distribution (d_{v10}, d_{v50}, and d_{v90}), power of the simulated inhalation, molecular weight, aromatic proportion, melting point, log *P*, and critical primary pressure (CPP) for dispersion of the API during pressure titration laser diffraction analysis, which was established by Jaffari et al.[38] All the input variables and the associated FPF (output variable) were analyzed in a PCA loading plot. The input variable, which had the lowest absolute value on PC-1 in the loading plot, was removed and the PCA repeated with the remaining inputs. If the explained variance of PC-1 in the new model remained the same or increased, then the removed input variable was excluded from the model as it was irrelevant for predicting FPF. This process was repeated until only the factors relevant to predicting the output remained. If there was a decrease in the explained variance of PC-1, then the factor was put back into the model, as this represented a breakdown in the modeling process, that is, the minimal number of variables required to explain the variance in the dataset had been reached.

Orthogonal Array Design

Primary and secondary orthogonal arrays (OAs) were designed to optimize the architecture characteristics for an ANN to be used to predict FPF. Ten architecture characteristics were evaluated during the primary OA (Table 2). Each of these characteristics was set at 2 levels and the data were normalized between the values of 0.1 and 0.9 (Eq. 2, where x_n is the normalized value, x is the descriptor value, x_{max} is the maximum value, and x_{min} is the minimum value for each of the inputs). The options for the dataset were either the full dataset or the dataset with the reduced number of input fields which was decided upon by PCA. In the primary OA, the learning functions examined were momentum and quickprop, and the type of termination was either none or CV incremental, where an iteration can only decrease the error of the CV set by a negligible amount (0.001).

$$x_n = 0.1 + (0.9 - 0.1) * \frac{(x - x_{max})}{(x_{min} - x_{max})} \quad (2)$$

The primary OA was designed using Minitab to give an L_{12} array. The 12 networks were created in Neurobuilder. Once created, the networks were tested with the external test set. The OA designs were evaluated using the mean r^2 value of the predicted versus observed FPF from the training, CV, and test sets, from each individual network to assess the effectiveness of each of the ANNs. The output of each individual prediction was also analyzed to determine whether it was correctly categorized, that is, within 15% of the observed FPF, enabling the amount of the records which were correctly categorized in each of the data subsets (training, CV, and training set) to be compared.

The results of the primary OA were fed into the design of a secondary OA. The number of architecture characteristics was reduced to 7 and the number of levels were increased to 3 (Table 3). A secondary OA card (L_{18}) was designed. Each of these networks were tested and validated with the same training, CV, and test sets described previously. The mean r^2 value for the training, CV, and test sets and the percentage of the observations that were correctly categorized were determined for each network. The difference between the maximum and minimum average r^2 value for each of the architecture characteristics was used to determine their significance. The most significant architecture characteristic was

Table 2
The ANN Architecture Characteristics Used in the Primary OA to Design an Optimum ANN

Factor Name	Architecture Characteristics	No. of Levels	Level 01	Level 02
DT	Dataset	2	Full	Without PCA fields
NU	Number of units in the hidden layer	2	10 units	100 units
NL	Number of layers in the hidden layer	2	1 layer	2 layers
LF	Learning function	2	Momentum	Quickprop
SS	Step size	2	0.1	1
MO	Momentum	2	0.1	0.7
EP	Epochs	2	50	1000
BT	Backpropagation type	2	Online	Batch
TE	Termination	2	None	CV incremental

Each factor was split into 2 levels to develop the best architecture for predicting FPF in the experimental dataset.

Table 3
he ANN Architecture Characteristics Incorporated in the Secondary Orthogonal Array and their Associated Levels

Factor Name	Architecture Characteristics	No. of Levels	Level 01	Level 02	Level 03
NU	Number of units in the hidden layer	3	50 units	100 units	200 units
NL	Number of layers in the hidden layer	3	2 layers	3 layers	4 layers
LF	Learning function	3	Quickprop	Delta Bar	Step
SS	Step size	3	0.3	0.5	0.7
MO	Momentum	3	0.1	0.2	0.3
EP	Epochs	3	500	1000	2000
TE	Termination	3	None	Maximum	Minimum

the one that had the greatest difference between the maximum and minimum mean r^2 value of the 3 levels. Any architecture characteristic which had a difference of 0.05 (5%) or greater between the maximum and minimum average r^2 value of the 3 levels was deemed to be a significant architecture factor in the creation of an ANN to predict FPF.

The Taguchi method uses a signal to noise ratio to calculate which of the levels of each of the architecture characteristics formed the optimum ANN. The signal to noise ratio was calculated from the mean r^2 value for each network in the secondary OA using Equation 3, where n was the number in each mean r^2 value. The ANN (a multilayer perceptron [MLP]) was created in Neurobuilder, the same method as the creation of all the other ANNs, using the optimum characteristics. The r^2 values and proportion of the records which were correctly categorized were used to evaluate the training, CV, and test sets as described above.

$$S/N = -10 \log \frac{\sum \frac{1}{r^2}}{n} \quad (3)$$

Results and Discussion

Selection of Input Factors Using Principal Component Analysis

The PCA loading plots were examined to identify any factors which did not contribute to explaining the overall variance of the dataset, that is, added noise to the analysis. Initially, the input variables specifically relating to the API were melting point, aromatic proportion, log *P*, and CPP. However, only 2 of these input variables, CPP and melting point, were retained, as analysis showed that the others did not contribute to explaining the variance in FPF in the dataset in the first 2 PCs (90% of the variance in the dataset). Of these, melting point was shown by PCA to be most important in explaining the variance in the dataset. However, this does not imply that melting point of an API is directly related to the aerodynamic properties of a substance. Melting point has been linked to other solid-state properties such as crystal packing/crystal structure, which can impact on the aerodynamic properties of the substance, and this may be reflected in these findings. Also, as the number of APIs increase and melting point range becomes more of a continuum, it is possible that this would influence its relative importance on the PCA loading plot. Hence, the baseline value was obtained from an input vector of total fines content, intrinsic fines content, concentration of the fines added, mean drug content, particle size distribution (d_{v10}, d_{v50}, and d_{v90}), power of the inhaler, melting point, amount of API in formulation, and CPP of the API (Table 4, No. 1).

The input factor which had a loading value close to zero on PC-1, that is, the factor contributing least to explaining the overall variance of the dataset, was mean drug content (PC-1 loading value −0.010). This factor was removed from the input vector and the PCA was regenerated. The second PCA resulted in an increase in the explained variance in PC-1 (Table 4, No. 1: 42%, No. 2: 45%) suggesting that the mean drug content was not contributing to explaining the variance in the dataset. However, there was limited variation in drug content, as the powders were formulated and tested for homogeneity and compliance with 95%-105% of target drug content in the original studies.

The goal of the PCA was to ascertain/identify which metric of fines content provided the best predictive strength to include in the ANN model. Concentration of fine lactose was the next input variable with the lowest explained variance of PC-1, and therefore was removed (PCA loading value: 0.011). Removal of concentration of fine lactose (PCA No. 3) caused an increase in the explained variance of PC-1 from 45% to 48%. There was also an increase in the explained variance of PC-2 from 21% to 23%, hence an increase in the overall variance which had been explained by the first 2 PCs (Table 4). The total fines content was calculated as the sum of the amount of fine lactose added and the intrinsic fines content of each formulation. Thus, there was an over-representation of the data by including all 3 input variables. The removal of concentration of fines reduced noise in the dataset and increased the predictive strength of the ANN model, which was represented by the increase in variance explained by PC-1. Removal of any other input factors caused breakdown in the orthogonality of the score plots (Table 4, No. 4-7). The remaining input factors which were important in describing FPF when analyzed using PCA were total fines content, intrinsic fines content, particle size distribution (d_{v10}, d_{v50}, and d_{v90}), power of the inhaler, melting point, amount of API, and CPP of the API.

Table 4
PCA loading Plots Were Used to Remove Input Factors which Caused Over-representation of Certain Variable in the Dataset

No.	What Has Been Removed?	PC1	PC2	Total Explained × Variance		
				C1	C2	C3
1	Baseline values	42.0	20.0	42.1	62.0	74.9
2	Without mean drug content	45.0	21.0	45.0	66.2	79.9
3	Without mean drug content and concentration of fines	48.0	23.0	48.2	70.9	84.2
4	Without mean drug content, concentration of fines and power	55.0	26.0	55.0	80.7	93.9
5	Without mean drug content, concentration of fines and d_{v10}	49.0	23.0	48.9	71.5	85.4
6	Without mean drug content, concentration of fines and d_{v50}	51.0	19.0	50.7	69.8	83.6
7	Without mean drug content, concentration of fines and d_{v90}	50.0	23.0	50.3	73.3	87.0

Input factors were evaluated for whether there was a reduction in the explained variance of PC-1 compared to the original PCA. PC-1 and PC-2 were taken from the PCA score loading plot, whereas the Explained × Variance was calculated from the explained variance plot when they were removed. The Explained Variance was calculated by the variance explained by PC-1 (C1) was then added the Explained Variance of PC-2 (C2). Then, the Explained Variance of PC-3 was added to C2 for C3. After C3, the Explained Variance was almost 100%.

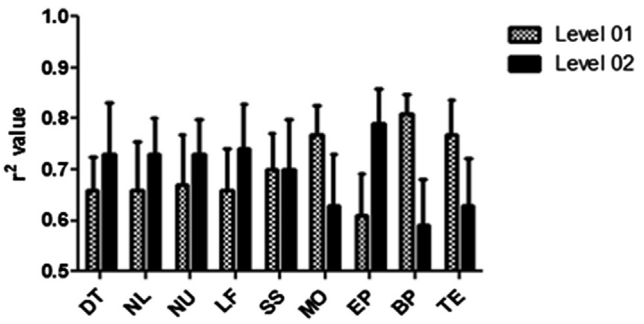

Figure 2. Mean r^2 value between the ANN predicted and observed FPF for the training, CV, and testing sets from the primary OA, separated into 9 architecture characteristics at 2 levels. These r^2 values were achieved from each of the ANN defined in the primary OA (Table 4). The architecture characteristics which were evaluated were dataset (DT), number of layers in the hidden layer (NL), number of units in the hidden layer (NU), learning function (LF), step size (SS), momentum (MO), number of epochs (EP), backpropagation type (BT), and termination (TE).

Optimization of ANNs Using the Orthogonal Arrays

The primary OA suggested that the creation of 12 networks was required to assess the 10 architecture characteristics. The architecture characteristics evaluated were dataset (DT), number of hidden layers (NL), number of units in the hidden layer (NU), learning function (LF), step size (SS), momentum (MO), number of epochs (EP), backpropagation type (BT), and termination (TE) (Table 1). Analysis was completed by comparing the mean and the individual r^2 values of the training, CV, and test sets for the observed versus predicted FPF for each of the 12 networks of the primary OA and at each of the 2 levels (Fig. 2). Also, the percent of the records that were correctly categorized was analyzed. Some of the networks created in the primary OA demonstrated a weak correlation between the observed FPF and the predicted FPF (Table 5, Network 9: r^2 value = 0.31, Network 2: r^2 value = 0.46, Network 5: r^2 value = 0.49). Many of the networks investigated were over-training, as indicated by the higher proportion of records which were categorized correctly in the CV compared to the training set where we define the percentage of records that were correctly categorized as the number of the records, that the ANN generated, which were within 15% of the experimentally observed values. Network 12 categorized the highest number of records successfully (average of 73.5%) and along with Network 8 did not show signs of over-training, where over-training was seen to have occurred when the network simply memorized the training set and was unable to generalize the problem (i.e., gave poor test set predictions). As such, reliance on the r^2 values as a measure of model predictivity was considered an over-simplification, and the percent of the records which were correctly categorized (any value produced by the ANN that was <15% variation from the desired value) was seen as a better measurement. Network 12 was the best performing network, seemingly as a result of the selection of online backpropagation and the reduced dataset without PCA input fields as input parameters.

The effect of changing the experimental factor level was the most significant for the backpropagation architecture (Fig. 2). Online learning was optimal when compared to batch learning (Level 01 online range of r^2 between 0.66 and 0.85, Level 02 batch range of r^2 between 0.31 and 0.79). Backpropagation affects the updating of the weights associated with the neurons and hence the error between the predicted and observed FPF. Given this, it is unsurprising that this is the most significant architecture characteristic. Online backpropagation updates the weights on the neurons after the presentation of each exemplar, whereas "batch" waits until the entire training set has been presented and then updates the weights.[39] With a larger dataset, online backpropagation has more records to update all the weights associated with the neurons, hence the error had a chance to reach the minima. The weight changes of batch learning were significantly larger with larger datasets, thus lead to larger step sizes being used which then in turn lead to unstable learning and over-shooting of the local minima.[39] Commonly for larger datasets, online learning can update the weights significantly quicker, as it does not have to reduce the learning rate and was used for future creation of ANNs.

One of the other most significant architecture characteristics in the primary OA was the dataset (DT) (Fig. 2; difference in $r^2 = 0.07$). The optimal dataset was the dataset which had the inputs removed that caused noise in the models (Level 02: $r^2 = 0.73$), as determined by PCA. Hence, PCA aided the understanding of which factors explain the largest variance of the dataset and this knowledge can then be used to remove any factors which contributed noise to the data.

The other architecture characteristics were analyzed to determine values for use in the secondary OA, which were then used to design the optimal ANN. When the number of epochs (EP) was set at 1000, this had the most profound effect on the r^2 value. Therefore future experiments were planned around this value (Level 02: $r^2 = 0.79$). Momentum (MO) had a larger effect at the rate of 0.1 (Level 01), hence further optimization occurred around this value (Level 01: $r^2 = 0.77$). CV incremental had a negligible effect on the r^2 value (Level 02: $r^2 = 0.63$) when understanding the optimal termination (TE) and therefore it was removed from analysis in the secondary OA.

Table 5
Analysis of Networks Designed by Primary OA Analysis: r^2 Value for Predicted Versus Observed FPF from the Training, CV, and Test Subsets and the Percent of Correctly Categorized Outputs (±15% of observed FPF)

No.	DT	NL	NU	LF	SS	MO	EP	BP	TE	r^2 value				% Correctly Categorized			
										Training	CV	Test	Mean ± SD	Training	CV	Test	Mean ± SD
1	Full	1	10	Momentum	0.1	0.1	50	Online	None	0.754	0.651	0.575	0.660 ± 0.09	63.0	44.7	51.1	56.9 ± 16.0
2	Full	1	10	Momentum	0.1	0.7	1000	Batch	CV	0.403	0.494	0.474	0.457 ± 0.05	41.2	29.8	48.9	42.6 ± 11.1
3	Full	2	10	Quickprop	1	0.1	50	Online	CV	0.816	0.790	0.738	0.781 ± 0.04	61.3	63.8	57.4	64.8 ± 7.8
4	Full	1	100	Quickprop	1	0.1	1000	Batch	None	0.887	0.824	0.713	0.808 ± 0.09	69.7	63.8	57.4	68.1 ± 13.3
5	Full	2	100	Momentum	1	0.7	50	Batch	None	0.460	0.532	0.478	0.490 ± 0.04	40.3	29.8	36.2	38.0 ± 9.2
6	Full	2	100	Quickprop	0.1	0.7	1000	Online	CV	0.847	0.814	0.709	0.790 ± 0.07	55.0	61.7	57.4	61.7 ± 4.3
7	Reduced	2	10	Quickprop	0.1	0.1	1000	Batch	None	0.895	0.916	0.879	0.897 ± 0.02	62.2	64.6	58.3	61.7 ± 3.2
8	Reduced	2	10	Momentum	1	0.7	1000	Online	None	0.903	0.950	0.849	0.901 ± 0.05	71.4	79.2	64.6	71.7 ± 7.3
9	Reduced	1	10	Quickprop	1	0.7	50	Batch	CV	0.431	0.249	0.246	0.309 ± 0.11	38.6	27.1	37.5	34.4 ± 6.4
10	Reduced	2	100	Momentum	0.1	0.1	50	Batch	CV	0.633	0.525	0.492	0.550 ± 0.07	37.0	25	37.5	33.2 ± 7.1
11	Reduced	1	100	Quickprop	0.1	0.7	50	Online	None	0.870	0.848	0.835	0.851 ± 0.02	63.9	60.4	56.3	60.2 ± 3.8
12	Reduced	1	100	Momentum	1	0.1	1000	Online	CV	0.923	0.887	0.873	0.894 ± 0.03	74.8	75.0	70.8	73.5 ± 2.3

Table 6
Analysis of Networks Designed by Secondary OA Analysis: r^2 Value for Predicted Versus Observed FPF from the Training, CV, and Test Subsets and the Percent of Correctly Categorized Outputs (±15% of observed FPF)

No.	NL	NU	LF	SS	MO	EP	TE	r^2 Value				% Correctly Categorized			
								Training	CV	Test	Mean ± SD	Training	CV	Test	Mean ± SD
1	4	200	Quickprop	0.3	0.2	1000	Minimum	0.904	0.946	0.911	0.920 ± 0.02	57.1	70.8	58.3	62.1 ± 7.6
2	2	50	Quickprop	0.3	0.1	500	None	0.877	0.948	0.902	0.909 ± 0.04	55.5	68.8	62.5	62.2 ± 6.6
3	4	100	Quickprop	0.7	0.1	500	Maximum	0.874	0.948	0.908	0.910 ± 0.04	58.8	77.1	64.6	66.8 ± 9.3
4	3	50	Quickprop	0.5	0.3	1000	Maximum	0.880	0.949	0.879	0.903 ± 0.04	61.3	79.2	60.4	70.0 ± 10.6
5	4	50	Delta Bar	0.5	0.3	500	Minimum	0.414	0.659	0.524	0.532 ± 0.12	31.1	39.6	33.3	34.6 ± 4.4
6	2	200	Delta Bar	0.5	0.1	1000	None	0.493	0.693	0.667	0.618 ± 0.11	30.3	52.1	47.9	43.4 ± 11.6
7	3	200	Quickprop	0.7	0.3	2000	None	0.928	0.951	0.936	0.938 ± 0.01	73.9	77.1	79.2	76.7 ± 2.6
8	4	100	Delta Bar	0.3	0.3	2000	None	0.352	0.671	0.572	0.531 ± 0.16	31.9	45.8	45.8	41.2 ± 8.0
9	2	50	Delta Bar	0.7	0.2	2000	Maximum	0.914	0.938	0.940	0.930 ± 0.01	71.4	68.1	68.1	69.2 ± 1.9
10	2	100	Step	0.3	0.3	1000	Maximum	0.904	0.938	0.932	0.925 ± 0.02	65.5	68.8	64.6	66.3 ± 2.2
11	4	200	Step	0.5	0.1	2000	Maximum	0.924	0.943	0.921	0.929 ± 0.01	73.9	72.9	66.7	71.2 ± 3.9
12	3	50	Step	0.3	0.1	2000	Minimum	0.913	0.934	0.899	0.915 ± 0.02	70.6	79.2	62.5	70.8 ± 8.3
13	2	200	Step	0.7	0.3	500	Minimum	0.893	0.927	0.894	0.905 ± 0.02	64.7	66.7	62.5	64.6 ± 2.1
14	4	50	Step	0.7	0.2	1000	None	0.919	0.942	0.910	0.924 ± 0.02	79.8	70.8	72.9	74.5 ± 4.7
15	3	100	Step	0.5	0.2	500	None	0.893	0.942	0.917	0.917 ± 0.02	63.0	70.8	64.6	66.1 ± 4.1
16	2	100	Quickprop	0.5	0.2	2000	Minimum	0.907	0.944	0.909	0.920 ± 0.02	72.3	79.2	70.8	74.1 ± 4.5
17	3	100	Delta Bar	0.7	0.1	1000	Minimum	0.900	0.934	0.913	0.916 ± 0.02	73.1	72.9	64.6	70.2 ± 4.9
18	3	200	Delta Bar	0.3	0.2	500	Maximum	0.886	0.926	0.915	0.909 ± 0.02	71.4	68.8	72.9	71.0 ± 2.11

The other 4 factors (number of hidden layers [NL], number of units [NU], learning function [LF], and step size [SS]) had little effect on the predictive ability of ANN with a difference of <5% between the r^2 values of the levels. Therefore, a similar range was chosen to further examine these in the secondary OA. It is likely that the number of units and number of hidden layers did not have a significant effect on the primary OA because the increase in number of units and the change from one to two hidden layers did not add enough complexity to the ANN to make a significant difference in the predictive ability of the ANN. Both step sizes and learning function performed similarly, so neither significantly improved the performance of the ANN. Overall, the primary OA helped to define the ranges of each of the architecture characteristics, whereas the secondary OA was used to refine these levels to produce the optimum ANN architecture characteristics.

The general validity of the networks was assessed using the values for the training, test, and validation sets during the ANN construction. The high r^2 values in the secondary OA showed a slightly biased view, as most networks accurately categorized around 60%-80% of the records correctly. Many of the networks demonstrated over-training with the CV or test set producing a higher percentage of records which have been correctly categorized when compared to the training set. The networks in the secondary OA which did not show signs of over-training were Networks 9, 11, 14, and 17. Analyzing the percent of the records which were correctly categorized and the r^2 values (Table 6) showed that these networks were not over-fitting the data, hence it demonstrated the plausible application of an ANN to predict FPF.

Initially, weights between links were set to random values between −1 and 1, as it is a common practice when establishing ANNs. Reproducibility of results is considered a limitation of ANN because of this random assignment. However, previous work carried out in the group (Ali J, unpublished data, 2014) has shown that similar overall results can be obtained for ANNs, which have different starting weights, as long as the parameters, such as the learning algorithm and the number of cycles, have been optimized; therefore the random assignment of weights was not considered to be a limiting factor in the general approach employed.

The secondary OA was important to identify the best learning function, the optimum number of hidden layers, step size, momentum rate, and the best type of termination. The most significant architecture characteristic in the secondary OA when improving the predictive ability of the ANN was the type of learning function (LF). It had the largest difference of 18% between the maximum and minimum r^2 value between levels (Fig. 3).

In the secondary OA, Level 02 of the learning function (Delta Bar) was detrimental to the predictability of the ANN (Level 02: $r^2 = 0.74$). However, step and quickprop (Level 01 and Level 03) had a similar effect (Level 01: $r^2 = 0.92$, Level 03: $r^2 = 0.92$, respectively). Learning function had the most significant effect because it was the way of transferring information between the layers. Delta Bar (Level 02) was an adaptive step size learning function, as it is updated according to the previous error values of the processing element. Therefore, if the current weight and past weight are both the same sign, this contributes to a quicker computational time, whereas if these weights were opposite signs then this indicates that the change to the value of the weight had gone too far. For a smaller sample size (<100 records), Delta Bar was the best learning function, as the Widow-Hoff error was involved to determine the amount the weights were changed by. However, Delta Bar (Level 02) was detrimental to the predictive ability of the ANN due to a combination of all the networks using online learning and the larger dataset allowing the other learning functions to thrive, as they had more training records to learn properly.[40] There was nothing to choose between the other 2 learning functions, step and quickprop. The learning function, Step, uses the step size to determine which direction to get to the local minima. Quickprop uses the momentum to provide the gradient descent toward the local minima. However, oscillations may occur if the momentum is set too high in the learning function of quickprop, which results in never reaching the local minima. This may have been one of the reasons that step was shown in the analysis to be slightly more beneficial.

The next most significant factor was the number of hidden layers (NL), and Level 02, 3 layers, produced the highest r^2 value out of the 3 levels (Level 03: $r^2 = 0.87$; Fig. 3). The number of hidden layers was one of the most significant factors because it affected the learning capabilities of the ANN (difference in $r^2 = 0.13$). The larger number of hidden layers allowed a greater complexity to be obtained in the ANN, hence the ANN can handle the complex dataset.

The other significant factors when predicting the FPF were step size, momentum, and the type of termination, as they had >5% variation between the r^2 values associated with the different levels.

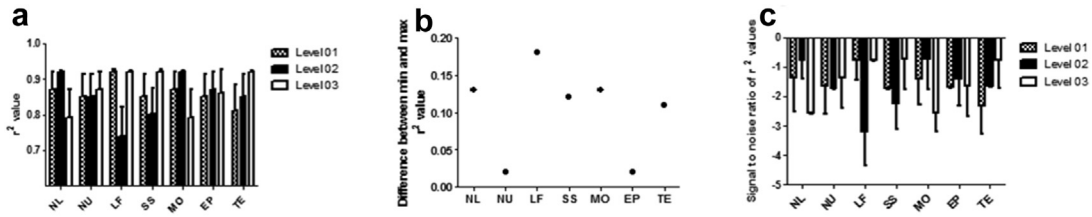

Figure 3. Evaluation of the number of hidden layers (NL), the number of units in the hidden layer (NU), learning function (LF), step size (SS), momentum (MO), number of epochs (EP), and termination (TE) (a) r^2 values between the desired and generated FPF from the average of the training, CV, and testing sets from the secondary OA (Table 5) (data represent mean ± SD), and (b) the difference in the average r^2 value between the maximum and minimum level for each architecture characteristics. (c) Signal to noise ratio from the training, CV, and testing sets (mean ± SD).

The optimal step size (SS) was Level 03, 0.7, as it produced the highest r^2 value for the 3 levels (Level 03: $r^2 = 0.92$). Also, the highest r^2 value from the 18 networks was produced when the momentum (MO) was set at 0.2 (Level 02: $r^2 = 0.92$) (Fig. 3). Step size and momentum were obviously significant architecture factors, as a non-optimal step size and momentum rate, especially with the learning function of step. The largest step size (Level 03) was large enough so the ANN reached the local minima without over-shooting it and causing the error to oscillate. As such, Level 03 was chosen for the optimum ANN as the smaller step sizes involved a greater amount of computational time and carried the risk of never reaching the local minima. Momentum was the rate at which the ANN reaches the local minima and if the momentum was too large it may over-shoot the local minima, hence the optimum condition was a small value of 0.2.

The best r^2 values were produced when the termination (TE) was set to "minimum" (Fig. 4; Level 03: $r^2 = 0.92$), which means that when the mean square error was less than the threshold value of 0.01 (default setting), the network training would terminate. The default setting was low enough that the network reached its local minima and did not over-train.

The other 2 factors assessed had a non-significant effect on the predictability of the ANN, although the results can still be examined to see which factor had the best response to help create the optimum ANN. The number of epochs showed a slight increase in the r^2 value when 1000 epochs were used (Level 02: r^2 value = 0.87) when compared to 500 epochs (Level 01: r^2 value = 0.86) and 2000 epochs (Level 03: r^2 value = 0.86). The number of the hidden units was optimal at the maximum level of 200 units in the secondary OA (Fig. 4; Level 03: $r^2 = 0.87$). This is unusual given that a commonly used heuristic in ANN design is that the number of hidden units should not be more than twice the number of inputs. Too few hidden units would result in high training and high generalization error, whereas too many hidden units would result in the model being over-fitted. The primary OA showed that models with 100 hidden units consistently outperformed those with 10 without being over-fitted. This formed the basis of the second array table where values of 50, 100, and 200 hidden units were examined. In these networks in the secondary OA, the greater number of units was not detrimental to the ANN performance, because the addition of units did not add any complexity to the ANN when compared to the addition of hidden layers.

Performance of the Optimized ANN

An "optimal" MLP architecture was designed from the results of the OA analysis (Fig. 3c) comprising 3 hidden layers, 200 units in each of these hidden layers, a learning function of step, a step size of 0.7, a momentum rate of 0.2, number of epochs of 1000, a termination procedure of minimum, and online backpropagation as the learning algorithm. The optimum MLP produced a high average r^2 value of 0.937 (Table 7). However, this ANN had a higher r^2 value and proportion of the records which were correctly categorized associated with the CV set when compared to the training set, which demonstrated that the ANN was over-fitting and could not perform at its best. This could have been because the levels selected for each architecture characteristic led to the network error missing

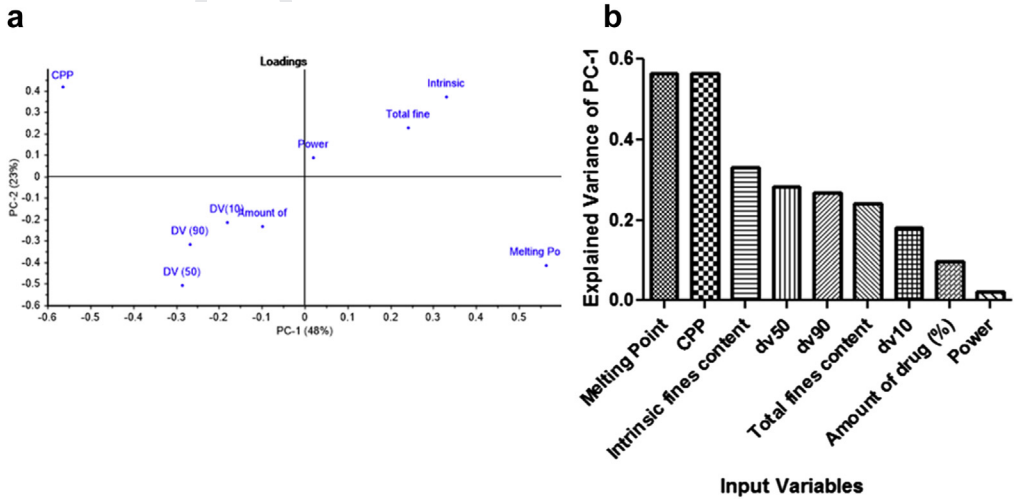

Figure 4. Explained variance of each of the final input variables used in the optimal ANN (a) loading plot of each of the input variables and (b) plot of the explained variance of PC-1 for each of the input variables.

Table 7
he r^2 Value and the Percent of Correctly Categorized FPF Records for Each Dataset Tested in the MLP Designed Using the Signal to Noise Ratio from the Taguchi Method, Compared to the Best Network from the Secondary Array (Network 14 from Table 6)

Set of Data	Network Designed Using Signal to Noise Ratio		Network 14 From Secondary OA
	r^2 Value Between the Expected and Generated FPF	Percent Which were Correctly Categorized	Percent Which were Correctly Categorized
Training	0.938	73.9	79.8
CV	0.952	79.2	70.8
Test	0.922	68.8	72.9
Mean ± SD	0.937 ± 0.015	73.9 ± 5.2	74.5 ± 4.7

the local minima, leading to the network over-training, but could also be a consequence of the small dataset used in the study.

In the primary and secondary OAs, some of the networks were over-fitting with higher r^2 values associated with the CV set and the optimum model due to the selection of certain ANN architecture characteristics (Fig. 3). However, this was not the case with Network 14 of the secondary OA which correctly categorized a greater amount of records when compared to the optimum ANN (74.5%) and was not over-fitted (Table 7). This indicated that some of the networks could be useful to predict FPF. Network 14's architecture include online backpropagation, 4 hidden layers, 50 units in each of these hidden layers, a learning function of step, a step size of 0.7, a momentum rate of 0.2, number of epochs of 1000, and no termination procedure. Network 14 from the secondary OA thus proved to be the best performing architecture for this dataset. Many of the ANNs created had a high mean r^2 value and the proportion of the records which were correctly categorized was above 70%. Thus, Network 14 in the secondary OA showed that ANNs could be a feasible modeling technique when analyzing the complex relationship between the formulation and device characteristics of a DPI.

This study indicated that it was possible to develop an ANN that could handle the complex relationship of different devices, formulations, and API to predict FPF, although given the small dataset used it cannot be claimed that this model is generalizable. In a similar study using ANNs, Kinnunen et al.[19] showed that the critical parameter in describing *in vitro* deposition patterns was the concentration of fine lactose which was <4.5 μm and >15 μm. This was determined by calculating the connection weights of the different input parameters in the ANN when the output was FPF and mass median aerodynamic diameter. However, the study covered less complexity with a single API, budesonide, and one device type, Cyclohaler. As r^2 values for these ANNs were not reported, a direct comparison between the 2 studies cannot be made beyond the success in using an ANN to predict the deposition of dry powder formulations with similar input factors. However, a comparison can be made between the input factors which were important in the ANN. The melting point and CPP were the most important input factors derived from the API in explaining the variance of FPF in the dataset, which was expected as these 2 input factors take one of two values. The next most significant input factors were those related to the formulation (Fig. 4), especially the amount of fine lactose similarly to the findings reported by Kinnunen et al.[19]

This work shows the potential to develop a useful generic model for predicting FPF, although this would require much larger datasets and more variable inputs, especially for the API and inhaler device. Preliminary investigations (data not shown) revealed that it was necessary to combine the experimental datasets to produce sufficient records such that a reliable network could be developed, as evident by higher r^2 values and the proportion of the records that were correctly categorized. Many of the projected networks showed signs of over-training, which are not unexpected given the small dataset, but the Taguchi method was able to identify which ANNs were successful. In general, ANNs are good at classifying items, so for future experiments the FPF may be grouped, which may lead to an improved proportion of records which are correctly categorized. Overall, relying on r^2 values alone as a measure of model predictability gave inflated results regarding model quality. However, the inclusion of the proportion of the records which were correctly categorized allowed a more realistic evaluation to mitigate against this. Even with these limitations, the potential for applying ANN to predicting FPF was shown and the dataset could be expanded to include a larger variety of inhalers and APIs, and the optimal ANN architecture characteristics for this specific dataset would evolve. Without such a sufficiently large and diverse dataset, ANN development will need to be completed for each prospective inhaled product on a case-by-case basis. With a considerably enlarged dataset, it may be possible to model formulation development to achieve a target performance specification, or match performance to that of a reference item. This technique could be used to aid the design and development of new products: the proposed material science tetrahedron proposes modeling approaches to probe the relationship between the structure, properties, processing of a medicine, and its performance. The optimum ANN used in this study allows the impact of structure and properties of inhaled product components on FPF to be predicted.

Conclusion

This study has demonstrated that ANNs can predict the effect of formulation and device design on the respirability of particles emitted from DPIs. The optimum ANN included inputs for the API, formulation, and device which reflected the recognized importance and interdependency of these factors for orally inhaled product performance. The Taguchi method was effective at identifying successful architecture with the potential for development as a useful generic ANN model for predicting FPF, although this would require much larger datasets and more variable inputs, especially for the API and inhaler device in order to be generalizable. To realize the full potential for reducing the impactor testing burden, the modeling could be extended to other impactor outputs, for example, emitted dose, aerodynamic particle size distribution, mass median aerodynamic particle size, and geometric standard deviation.

Acknowledgments

This study was sponsored by MedPharm Ltd in the form of research studentship for J.M.

References

1. Newman S, Busse W. Evolution of dry powder inhaler design, formulation, and performance. *Respir Med*. 2002;96(5):293-304.
2. Guenette E, Barrett A, Kraus D, Brody R, Harding L, Magee G. Understanding the effect of lactose particle size on the properties of DPI formulations using experimental design. *Int J Pharm*. 2009;380(1):80-88.
3. Stegemann S, Kopp S, Borchard G, et al. Developing and advancing dry powder inhalation towards enhanced therapeutics. *Eur J Pharm Sci*. 2013;48(1):181-194.

4. Hoppentocht M, Hagedoorn P, Frijlink H, de Boer A. Design and development of a disposable high dose dry powder inhaler for aminoglycosides. *J Aerosol Med Pulm Drug Deliv*. 2014;27(4):A22.
5. Demoly P, Hagedoorn P, de Boer AH, Frijlink HW. The clinical relevance of dry powder inhaler performance for drug delivery. *Respir Med*. 2014;108(8):1195-1203.
6. Tong Z, Kamiya H, Yu A, Chan HK, Yang R. Multi-scale modelling of powder dispersion in a carrier-based inhalation system. *Pharm Res*. 2015;32(6):2086-2096.
7. Haghpanah M, Marriott C, Martin G. Potential use of microencapsulation for sustained drug delivery to the respiratory tract. *J Aerosol Med*. 1994;7(2):185-188.
8. Forbes B, Bäckman P, Christopher D, Dolovich M, Li BV, Morgan B. In vitro testing for orally inhaled products: developments in science-based regulatory approaches. *AAPS J*. 2015;17(4):837-852.
9. Olsson B, Borgström L, Lundbäck H, Svensson M. Validation of a general in vitro approach for prediction of total lung deposition in healthy adults for pharmaceutical inhalation products. *J Aerosol Med Pulm Drug Deliv*. 2013;26(6):355-369.
10. Forbes B, Asgharian B, Dailey LA, et al. Challenges in inhaled product development and opportunities for open innovation. *Adv Drug Deliv Rev*. 2011;63(1-2):69-87.
11. Adams WP, Ahrens RC, Chen ML, et al. Demonstrating bioequivalence of locally acting orally inhaled drug products (OIPs): workshop summary report. *J Aerosol Med Pulm Drug Deliv*. 2010;23(1):1-29.
12. Fuglsang A. The US and EU regulatory landscapes for locally acting generic/hybrid inhalation products intended for treatment of asthma and COPD. *J Aerosol Med Pulm Drug Deliv*. 2012;25(4):243-247.
13. FDA. *Guidance for Industry. Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products*. 1998.
14. EMA. *Guideline on the Requirements for Clinical Documentation for Orally Inhaled Products (OIP) Including the Requirements for Demonstration of Therapeutic Equivalence between Two Inhaled Products for Use in the Treatment of Asthma and Chronic Obstructive Pulmonary Disease (COPD) in Adults and for Use in the Treatment of Asthma in Children and Adolescents*. 2009. CPMP/EWP/4151/00 Rev 1.
15. FDA. FDA Guidance Smoothes Advair Generic Approval Process. In: Wilson D, ed. Law360. 2013.
16. Fuglsang A. Futility rules in bioequivalence trials with sequential designs. *AAPS J*. 2014;16(1):79-82.
17. Sun CC. Materials science tetrahedron—a useful tool for pharmaceutical research and development. *J Pharm Sci*. 2009;98(5):1671-1687.
18. Zeng XM, Martin GP, Marriott C, Pritchard J. The influence of carrier morphology on drug delivery by dry powder inhalers. *Int J Pharm*. 2000;200(1):93-106.
19. Kinnunen H, Hebbink G, Peters H, Shur J, Price R. Defining the critical material attributes of lactose monohydrate in carrier based dry powder inhaler formulations using artificial neural networks. *AAPS PharmSciTech*. 2014;15(4):1009-1020.
20. Takayama K, Fujikawa M, Nagai T. Artificial neural network as a novel method to optimize pharmaceutical formulations. *Pharm Res*. 1999;16(1):1-6.
21. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Bradford Books/MIT Press; 1985.
22. Krogh A. What are artificial neural networks? *Nat Biotechnol*. 2008;26(1):195-197.
23. Winkler DA. Neural networks as robust tools in drug lead discovery and development. *Mol Biotechnol*. 2004;27(2):139-167.
24. Nazir J, Barlow DJ, Jayne Lawrence M, Shrubb I. Artificial neural network prediction of the patterns of deposition of polydisperse aerosols within human lungs. *J Pharm Sci*. 2005;94(9):1986-1997.
25. Nazir J, Barlow DJ, Lawrence MJ, Richardson CJ, Shrubb I. Artificial neural network prediction of aerosol deposition in human lungs. *Pharm Res*. 2002;19(8):1130-1136.
26. De Matas M, Shao Q, Biddiscombe MF, Meah S, Chrystyn H, Usmani OS. Predicting the clinical effect of a short acting bronchodilator in individual patients using artificial neural networks. *Eur J Pharm Sci*. 2010;41(5):707-715.
27. De Matas M, Shao Q, Richardson CH, Chrystyn H. Evaluation of in vitro in vivo correlations for dry powder inhaler delivery using artificial neural networks. *Eur J Pharm Sci*. 2008;33(1):80-90.
28. De Matas M, Shao Q, Silkstone VL, Chrystyn H. Evaluation of an in vitro in vivo correlation for nebulizer delivery using artificial neural networks. *J Pharm Sci*. 2007;96(12):3293-3303.
29. Varshosaz J, Tavakoli N, Minayian M, Rahdari N. Applying the Taguchi design for optimized formulation of sustained release gliclazide chitosan beads: an in vitro/in vivo study. *AAPS PharmSciTech*. 2009;10(1):158-165.
30. Jahanshahi M, Najafpour G, Rahimnejad M. Applying the Taguchi method for optimized fabrication of bovine serum albumin (BSA) nanoparticles as drug delivery vehicles. *Afr J Biotechnol*. 2008;7(4):362-367.
31. Khaw JF, Lim B, Lim LE. Optimal design of neural networks using the Taguchi method. *Neurocomputing*. 1995;7(3):225-245.
32. Saglam H. Tool wear monitoring in bandsawing using neural networks and Taguchi's design of experiments. *Int J Adv Manuf Technol*. 2011;55(9-12):969-982.
33. Wang Q, Stockton D, Baguley P. Process cost modelling using neural networks. *Int J Prod Res*. 2000;38(16):3811-3821.
34. Hassoun M, Ho S, Muddle J, et al. Formulating powder–device combinations for salmeterol xinafoate dry powder inhalers. *Int J Pharm*. 2015;490(1-2):360-367.
35. Muddle J, Murnane D, Parisini I, Brown M, Page C, Forbes B. Interaction of formulation and device factors determine the in vitro performance of salbutamol sulphate dry powders for inhalation. *J Pharm Sci*. 2015;104(11):3861-3869.
36. Parisini I. Improving the Aerosol Deposition Profiles from Dry Powder Inhalers. Department of Pharmacy, ed. University of Hertfordshire. 2015.
37. Louey MD, Van Oort M, Hickey AJ. Standardized entrainment tubes for the evaluation of pharmaceutical dry powder dispersion. *J Aerosol Sci*. 2006;37(11):1520-1531.
38. Jaffari S, Forbes B, Collins E, Barlow DJ, Martin GP, Murnane D. Rapid characterisation of the inherent dispersibility of respirable powders using dry dispersion laser diffraction. *Int J Pharm*. 2013;447(1):124-131.
39. Wilson DR, Martinez TR. The general inefficiency of batch training for gradient descent learning. *Neural Netw*. 2003;16(10):1429-1451.
40. Zhang DLH, Alippi MPC. *Advances in Neural Networks—ISSN*. Switzerland: Springer; 2010.