*Proceedings of the 2018 Winter Simulation Conference*
*M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.*

# A SIMULATION OPTIMIZATION APPROACH TO OPTIMAL CALIBRATOR LEVEL SELECTION FOR CLINICAL IMMUNOASSAYS

Utkarsh Vardhan
Varun Ramamohan

Department of Mechanical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi, 110016, INDIA

## ABSTRACT

Improving bioanalytical procedures (assays) to measure levels of immunologically relevant proteins such as haptoglobin requires minimizing the uncertainty associated with the measurement process. In this work, we consider the question of optimally selecting the concentration levels of calibrators (solutions with known concentrations of haptoglobin) used to calibrate the haptoglobin immunoassay. Specifically, we aim to select calibrator concentration levels such that the net measurement uncertainty at medical decision points is minimized. We accomplish this by constructing a Monte Carlo simulation of the haptoglobin immunoassay that estimates the measurement uncertainty associated with the assay, and then formulating the problem of optimally selecting calibrators to minimize the assay uncertainty at medical decision points as a discrete simulation optimization problem. We demonstrate the application of the NSGS and KN procedures to solve this problem. This work represents a first step towards the utilization of simulation optimization in the optimal design of clinical assays.

## 1 INTRODUCTION

Results of clinical laboratory measurements inform diagnosis and monitoring of disease, and also guide the determination of appropriate dosage levels of drugs. No clinical measurement process (also referred as an "assay") is free from uncertainty, and therefore, estimates of measurement uncertainty are required to quantify the quality of a clinical measurement. In particular, the calibration of clinical measurement processes is an important contributor to the net measurement uncertainty associated with an assay. In this study, we construct a Monte Carlo simulation model that estimates the uncertainty associated with the measurement of the immunoprotein haptoglobin in the clinical laboratory, and then use the simulation model to optimize the values of certain parameters associated with the calibration of the haptoglobin assay. Specifically, we consider the question of optimally selecting the concentration levels of the solutions with known concentration levels of haptoglobin (commonly known as standard solutions or "calibrators"). These calibrators are used to calibrate the measuring instrument used in the haptoglobin assay, and we aim to demonstrate the use of simulation optimization methods to optimally select these calibrators such that the measurement uncertainty of the haptoglobin assay at medical decision points is minimized. This represents a new approach for manufacturers of clinical assays, and potentially for manufacturers of measurement systems in general, to optimally select calibrators in terms of minimizing measurement uncertainty.

The concepts associated with measurement uncertainty, along with guidelines for its analytical modeling (including the "law of propagation of uncertainty", an expression to estimate the combined uncertainty associated with a measurement process) were formally described in 1993 in the Guide to the Expression of Uncertainty in Measurement (GUM) (Bureau International des Poids et Measures 1993), and in its later revised versions (Joint Committee for Guides in Metrology 2008b). The term uncertainty is formally defined in the GUM as "any parameter that characterizes the dispersion of the distribution of the values

that can be attributed to the result of a measurement". In this study, we utilize the standard deviation to quantitatively describe measurement uncertainty. We chose the standard deviation because the sources of uncertainty associated with the measurement process are all characterized to be Gaussian random variables, and the measurement result (hereafter the *measurand*) distribution was also determined to be Gaussian. We also note here that of the three stages of a clinical assay (Burtis et al. 2012), we only model the analytical stage because of the following reasons: (1) the identification and distributional characterization of the many sources of preanalytical uncertainty merits a separate study in itself, and (2) modeling postanalytical uncertainty is outside the scope of our study because it is usually introduced due to human error in recording or reporting of the result of the measurement.

Haptoglobin is a clinically important transport and acute phase protein. Decreased levels of haptoglobin indicate the presence of hemolytic anemia, which is indicative of underlying conditions such as chronic liver disease, bleeding in the liver or spleen, blood disorders, transfusion reactions, etc. Elevated levels of the protein indicate the presence of inflammation, typically due to peptic ulcers, infection, extreme stress, burns, etc. (Schwartz 2011). Two reagents are required for the haptoglobin assay - the first supplies the metal ion buffer and preservatives, and the second provides the antibodies (henceforth referred to collectively by the term "antibody") required for the reaction. The patient sample or a calibrator supplies the haptoglobin required for the assay. The bioanalytical principle underlying the assay is that of immunoturbidimetry, which involves a chemical reaction between the antigen (haptoglobin) and the antibody, yielding an insoluble antigen-antibody complex. The formation of this antigen-antibody complex increases the cloudiness or "turbidity" of the reaction mixture. The optical absorbances recorded as part of the haptoglobin assay is directly proportional to the concentration of the turbidity-causing antigen-antibody complex at that point in time, and hence measure the turbidity of the reaction mixture. The optical absorbance measurements are then converted into the quantity to be measured, the haptoglobin concentration, by a nonlinear calibration function. A detailed description of the immunoassay measurement procedure is provided in Section 3.

We now provide a brief overview of the literature in this area, and describe how our work is located in relation to the existing work on finding optimal calibrators for bioanalytical assays.

## 2 LITERATURE REVIEW

We found very few published studies that describe models of the net uncertainty of immunoassays. Our literature search yielded the following studies that dealt with the analytical estimation of immunoassay measurement uncertainty (Borg et al. 2002; Suchanek and Robouch 2009; Ramamohan et al. 2015; Pereira et al. 2015; Pereira et al. 2016). Borg et al. (2002) and Suchanek and Robouch (2009) develop models that involve utilizing the law of propagation of uncertainty to estimate the combined measurement uncertainty of the assay, whereas the studies by Pereira and colleagues apply models of diagnostic accuracy to experimentally collected data. Further, none of these studies involve the immunoassay under consideration here - haptoglobin.

In this work, we use a systems engineering approach to first develop a mathematical model that describes the biochemistry of the assay, and then utilize this model to integrate individual uncertainties of the measurement system components. We subsequently utilize Monte Carlo simulation to quantify the net assay uncertainty.The use of Monte Carlo simulation, as opposed to the law of propagation of uncertainty, to estimate the uncertainty associated with nonlinear models of measurement systems and with sources of uncertainty which require characterization using Type B ad-hoc non-statistical methods (see Section 3.1) is well supported, including in a supplement to the GUM (Joint Committee for Guides in Metrology 2008a) dedicated to the use of Monte Carlo simulation for such cases. A similar approach was first suggested by Aronsson et al. (1974), and subsequently by Krouwer (2002). Ramamohan et al. (2015) use such a systems perspective to build a Monte Carlo simulation model of the uncertainty of the rheumatoid factor (RF) immunoassay. In this study, we describe an approach similar to that utilized in Ramamohan et al. (2015) to model the haptoglobin assay uncertainty. However, in contrast to the authors' use of Monte Carlo simulation to investigate how uncertainty in the calibration process affects the correlation structure of the

calibration function, we use the Monte Carlo simulation model to identify optimal calibrator concentration levels such that the net measurement uncertainty at medical decision points is minimized. We use a simulation optimization approach - specifically the NSGS (Nelson et al. 2001) and KN (Kim and Nelson 2001) procedures - to identify the optimal calibrator concentration levels.

Our search of the literature yielded very limited work done regarding the determination of optimal calibration concentrations for bionanalytical assays. A key early study we identified was by Buonaccorsi (1986), who statistically examined the effect of the choice of "design points" or calibrators on a measurement system utilizing a linear calibration function. Another set of studies (Boer, Rasch, and Hendrix 2000; Melas 2005; Gauchi and Pázman 2006) employed statistical approaches focused on finding optimal designs for nonlinear calibration models, and were based on improving on a previously determined set of parameters for the calibration function. Bayesian methods for finding optimal designs (Müller 2005; Clyde and Chaloner 2002; Matthews and Allcock 2004) for nonlinear calibration models have also been developed. These methods require specification of a multivariate distribution for the parameters of the calibration function.

A key relevant study is the article by Forkman (2008). The author suggested a purely statistical approach for finding optimal design points for nonlinear univariate calibration functions that utilizes aspects of both Bayesian methods and methods that condition on a previously determined fixed set of calibration parameters. The method requires that the calibration function parameters are random variables with known expected values, variances and covariances, and finds the optimal design using first- and higher-order derivatives of the calibration function and its inverse. The objective function that is optimized is typically a function of the variances of possible patient sample concentrations. The author demonstrates the use of the method for the Michaelis-Menten and four-parameter logistic models.

Another key study is the work by Ramamohan et al. (2014). The authors develop a Monte Carlo simulation model of the serum triglycerides assay that utilizes a mathematical description of the assay biochemistry. They then utilize a brute force enumeration approach to explore the effect of the choice of calibrators for a linear calibration function on the net measurement uncertainty at medical decision points. This method will likely be prohibitively computationally expensive for assays with nonlinear calibration functions. Therefore we extend the work done in this study to assays with nonlinear calibration functions, and to find optimal calibrator concentration levels using more sophisticated methods that guarantee a minimum probability of selection of the optimal set of calibrators.

With the exception of the work in Ramamohan et al. (2014), our approach for identifying optimal calibrator sets differs from the above studies in the following ways: (a) no assumptions or prior knowledge regarding the distributions of the calibration curve parameters is required, (b) our model integrates a mathematical description of the assay biochemistry, the nonlinear calibration function, and distributional specifications regarding the sources of uncertainty operating within the assay, (c) no derivatives of the calibration function or its inverse are required, (d) our search did not yield a single study that involved the application of simulation optimization techniques to find optimal calibrator sets, and (e) heuristic methods are not used, and hence a minimum probability of selection of the optimal set of calibrators is guaranteed. In comparison with a purely statistical treatment of the problem, the development of a mathematical description of the assay biochemistry, its integration with the nonlinear calibration function and the distributional specifications regarding the sources of uncertainty yield a more comprehensive description of assay performance. This provides the model with substantial flexibility - for instance, the inclusion of a new calibration protocol on the optimal design can easily be explored using this model.

The paper is organized as follows: in Section 3, we describe the development of the Monte Carlo simulation model; and in Section 4, we describe the formulation of the optimization problem, the procedures applied to solve it, associated considerations and the results of the solution of the simulation optimization problem; in Section 5, we summarize the work, discuss limitations and outline avenues for future work.

## 3 IMMUNOASSAY MODEL DEVELOPMENT

The mathematical model of the haptoglobin immunoassay presented in this section was developed in collaboration with a large multinational diagnostics platform manufacturer.

In the case of the haptoglobin assay, the quantity to be measured or the measurand is the concentration of the antigen haptoglobin in the patient sample. The physical principle underlying the measurement of haptoglobin is immunoturbidimetry, which involves the determination of the turbidity of the reaction mixture. The turbidity of the reaction mixture increases due to the increase in concentration of the insoluble antigen-antibody complex that is formed as a result of the binding that occurs between the antigen and the antibody (the agglutination process). This binding process can be described in the form of a chemical reaction as follows:

$$\mathrm{Ag} + \mathrm{Ab} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{AgAb} \tag{1}$$

Here $Ag$ and $Ab$ represent the antigen (haptoglobin) and antibody (rabbit monoclonal antibodies) respectively, and $AgAb$ as is evident, denotes the antigen-antibody complex. The terms $k_1$ and $k_{-1}$ represent the rate constants of the forward and backward reactions, respectively. Two reagents, $R_1$ and $R_2$, and the patient sample $S$ are delivered into the reaction cell where the antigen-antibody binding occurs. $R_1$ contains the metal ion buffer and the preservatives necessary for the chemical reaction, and $R_2$ supplies the antibody that binds to the haptoglobin in the patient sample. Two optical absorbance measurements are made as part of the assay. The first measurement, denoted by $A_{x(0)}$, is made approximately at time $t = 0$ before $R_2$ is added to the reaction mixture, and the second measurement, denoted by $A_x$ is made 5.4 minutes after the first measurement is made. Since the first measurement is made at $t = 0$ before $R_2$ is added, the absorbance associated with the reaction mixture at $t = 0$ can be considered to be zero for all practical purposes.

The volumes of the reagents are denoted by $V_{r1}$ and $V_{r2}$. The volume of the calibrator or the patient sample is denoted by $V_s$.

The difference between optical absorbance measurements $A_x$ and $A_{x(0)}$ is converted into the concentration of the antigen haptoglobin in the patient sample, denoted by $C_x$, by a nonlinear calibration function given below. Since we assume $A_{x(0)}$ to be zero, only the term $A_x$ appears hereafter, and in the expression below.

$$C = \left[ \frac{(p_1 - (A_x - p_4))}{p_2\,(A_x - p_4)} \right]^{1/p_3} \tag{2}$$

This is the inverse log-logit function. Here, $p_1$, $p_2$, $p_3$ and $p_4$ are the parameters of the calibration function. $C$ is an intermediate concentration value that is converted into the desired patient haptoglobin concentration value using the following function:

$$C_x = (C + C_b)\, IF_A + IF_B \tag{3}$$

In the above expression, $C_b$ represents the concentration of haptoglobin in the standard 1 calibrator solution, which is typically a blank sample with a haptoglobin concentration of zero. $IF_A$ and $IF_B$ are instrument constants which in ideal error-free conditions represent a slope of 1 and intercept of zero. These instrument constants are typically not used in routine laboratory practice, and hence we do not consider them further in the analysis. Therefore, we only consider equation 2 in the following derivation. Equation 2 has four parameters of unknown value, and while a minimum of four calibrators are required to characterize the calibration function, six calibrators are used in practice for the haptoglobin assay. We refer to these calibrators as $Ag_1$, $Ag_2$, ..., $Ag_6$.

The analytical stage of the assay can be considered as consisting of the calibration phase and the sample analysis phase. In the calibration phase, the instrument is calibrated using standard solutions (solutions with known analyte concentrations) and the values of the calibration function parameters are estimated. In the sample analysis phase, the sample with unknown concentration (generally the patient sample) is evaluated

using the calibrated instrument. In terms of the uncertainty associated with the calibration function in equation 2, uncertainty in the calibration phase is represented by the uncertainty of the calibration parameters; uncertainty in the sample analysis phase is associated with that of the independent variable $A_x$; and the net assay uncertainty is represented by the uncertainty associated with the dependent variable $C_x$. The construction of the model for the calibration phase is presented below.

### 3.1 Calibration Phase

The parameters of the calibration function $p_1$, $p_2$, $p_3$, and $p_4$ are estimated in this phase. This is done using six calibrators and measuring their corresponding optical absorbance values. The parameters of the calibration function are determined from these six concentration values and their optical absorbance measurements using a nonlinear least squares curve-fitting method. Prior to describing how we model the uncertainty in the calibration process, we must mention the use of a reference function to generate absorbance values for the purposes of the simulation. This reference function is established using quality control data provided by the manufacturer, and its parameters are assumed to be error-free. The absorbance values (corresponding to given values of haptoglobin concentration) generated by this reference function are also therefore treated as 'true' or 'reference' values.

We now describe how uncertainty is introduced into the measurement of a single calibrator. Let $[Ag]$ represent the desired haptoglobin concentration of a calibrator sample. Based on input from the instrument manufacturer, three sources of calibrator uncertainty are considered: calibrator set-point uncertainty ($u_{c1}$), vial to vial variability ($u_{c2}$) and calibrator reconstituted stability ($u_{c3(t)}$). Uncertainty introduced in the calibrator concentration level during its manufacture and before it is used in the clinical laboratory is described by calibrator set-point uncertainty. The second source describes the uncertainty generated in the haptoglobin concentration level in each vial of the calibrator during its preparation from the manufacturer's batch. The deterioration (percent decrease per day of calibrator concentration) of the sample when the vial with the calibrator is reconstituted and stored after it is used in the laboratory each day (e.g., up to $N$ days) is represented by calibrator reconstituted stability.

Introducing the above sources of uncertainty associated with the calibrator into the model changes the value of $[Ag]$ according to the following equation:

$$[Ag]' = [Ag]\,(1+u_{c1})\,(1+u_{c2})\prod_{t=1}^{N}(1+u_{c3(t)}) \tag{4}$$

The sources of calibrator uncertainty, in addition to the other sources of uncertainty within the assay, are modeled by fitting appropriate probability density functions to the specifications given to us by the manufacturer for the sources of uncertainty. This approach towards modeling the sources of uncertainty belongs to a class of ad-hoc non-statistical methods known as the Type B method. Type B methods include relying on expert judgment, manufacturer-provided specifications, etc. to assign and parameterize distributions that describe the variation of sources of uncertainty. This in contrast to Type A methods, which involve determining the best fitting distribution to data (preferably experimental) for each source of uncertainty (Bureau International des Poids et Measures 1993). In the case of this study, a Type B method was used because we did not have access to experimental data for any of the sources of uncertainty included in the model. As an example of our approach towards modeling the sources of uncertainty, specifications for vial-to-vial variability were supplied by the assay manufacturer as a coefficient of variation (CV) of 1.5%. Based on information from the assay manufacturer indicating that deviations from the desired value for calibrator concentration after preparation of the calibrator vial were usually symmetrically distributed around a mean value, and that bias in the vial preparation process is negligibly small, a Gaussian distribution with an expected value of 0% and a standard deviation of 1.5% was assumed to model vial-to-vial variability. Therefore, if we assume that vial-to-vial variability was the only source of variation contributing to calibrator uncertainty, at a desired (error-free) calibrator concentration of 200 mg/dL, the haptoglobin concentration level in actuality would be represented by a Gaussian distribution with an expected value of 200 mg/dL and a

standard deviation of 3 mg/dL. Similar assumptions were made for the other sources of calibrator uncertainty as well, with the exception of calibrator reconstituted stability, wherein a non-zero bias was used to describe the deterioration of the calibrator with time and reconstitution. Further, simulation experiments with other distributions capable of symmetry and unimodality were conducted (e.g., the triangular distribution), but the estimates of measurement uncertainty proved to be robust to these changes, and hence the assumption of Gaussian sources of uncertainty was retained.

The measured calibrator concentration is also affected by the sources of uncertainty operating within the instrument. We identified three key sources of uncertainty associated with the instrument: sample pipetting uncertainty, reagent pipetting uncertainty and photometer uncertainty. The first two sources describe the uncertainties in the sample volume and reagent volume, respectively, delivered into the reaction test tube, and therefore changes the total volume of the reaction mixture and the number of molecules of the antigen and antibody in the reaction test tube at time $t$ = 0. In contrast, photometer uncertainty affects the optical absorbance measurement recorded at $t$ = 5.4 minutes. These sources of uncertainty are also statistically modeled as Gaussian distributions, similar to the sources of calibrator uncertainty. Estimates of the distributional parameters of the sources of uncertainty associated with the haptoglobin assay are listed in Table 1. It is evident from Table 1 that zero bias was assumed for all sources of instrument uncertainty.

Table 1: Probability distribution parameters of the sources of uncertainty.

| Source of uncertainty | Distribution | Mean (%) | SD (%) | Notes |
|---|---|---|---|---|
| Calibrator set-point uncertainty ($u_{c1}$) | Gaussian | 0.00 | 0.10 | |
| Vial-to-vial variability ($u_{c2}$) | Gaussian | 0.00 | 1.50 | |
| Reconstituted stability ($u_{c3(t)}$) | Gaussian | -1.25 | 0.42 | Decrease of haptoglobin concentration per day |
| Sample pipetting uncertainty ($x$) | Gaussian | 0.00 | 1.50 | |
| Reagent pipetting uncertainty ($y_1, y_2$) | Gaussian | 0.00 | 4.00 | |
| Photometer uncertainty ($u_p$) | Gaussian | 0.00 | 0.45 | |

The next step in the development of the model involves deriving the effect of the sources of sample and reagent pipetting uncertainty on the optical absorbance $A$ of the calibrator $[Ag']$. This portion of the model is essentially the same as that for the rheumatoid factor immunoassay, and due to space limitations we refer the reader to the work by Ramamohan et al. (2015) for the details of the derivation.

We provide below the expression relating the change in absorbance (denoted by $\delta A$) to the fractional changes in sample (calibrator) volume $V_s$ and total reaction mixture volume $V$ (denoted by $x$ and $z$, respectively) due to sample and reagent pipetting uncertainty:

$$\delta A = k\,[Ag_s]\frac{V_s}{V}\,(1-e^{-k_1 t})\left(\frac{x-z}{1+z}\right) \tag{5}$$

Note that the fractional change $z$ in total reaction mixture volume $V$ is related to the fractional changes in sample and reagent volumes $V_s$, $V_{r1}$ and $V_{r2}$ due to sample and reagent pipetting uncertainty in the following manner:

$$V+\delta V = V_s\,(1+x)+V_{r1}\,(1+y_1)+V_{r2}\,(1+y_2) \tag{6a}$$

That is,

$$z=\frac{\delta V}{V}=\frac{1}{V}\,(x\,V_s+y_1\,V_{r1}+y_2\,V_{r2}) \tag{6b}$$

We represent the fractional change in absorbance at time $t$ as a result of pipetting uncertainty, $\delta A/A$, as $u_{pc}$. At this stage, equation 5 describes the change at time $t$ in optical absorbance from the target

value that occurs prior to measuring the optical absorbance. When the absorbance is measured, uncertainty in the photometer further changes the optical absorbance by $u_p$. The equation for absorbance after the incorporation of all the sources of instrument uncertainty into the model is given below:

$$A^{'} = A\ (1+u_{pc})\ (1+u_p) \tag{7}$$

The above process is repeated for each of the six calibrators, which yields six calibrator concentrations $[Ag_1]'$, $[Ag_2]'$, ..., $[Ag_6]'$, and six corresponding optical absorbance measurements $[A_1]'$, $[A_2]'$, ...., $[A_6]'$. The $[Ag_i]'(i = 1-6)$ are functions of the sources of calibrator uncertainty, and can therefore be represented by equation 4. Similarly, the $[A_i]'(i = 1-6)$ denote absorbance values after the sources of uncertainty associated with calibration are incorporated into the model, and can therefore be represented by equation 7.

These six values of haptoglobin concentration and their corresponding optical absorbance values are used to determine the values of the calibration parameters $p'_1, p'_2, p'_3$ and $p'_3$. The dashed symbols for the calibration parameters represent the incorporation of calibration uncertainty incorporated into their values. The values of the calibration parameters are obtained by applying the Levenberg-Marquardt nonlinear least-squares algorithm. The calibration function with the uncertainty of the calibration phase now becomes:

$$C = \left[\frac{(p'_1 - (A_x - p'_4))}{p'_2\ (A_x - p'_4)}\right]^{1/p'_3} \tag{8}$$

### 3.2 Sample Analysis Phase

After the calibration process is complete, the process enters the sample analysis phase. The uncertainty around the independent variable $A_x$ in the calibration function represents the uncertainty of this phase; that is, the uncertainty associated with the optical absorbance measurement of the patient sample. Patient sample uncertainty (i.e., typically associated with the preanalytical stage, which is not considered in this study), reagent and instrument uncertainty operate within this phase. However, in scenarios wherein the target haptoglobin level in the sample is available before the analysis is performed (e.g., quality control experiments), sample uncertainty can be quantified. It is characterized by a Gaussian distribution with the mean representing the desired or true absorbance reading (assuming zero bias), and the standard deviation estimated based on the specifications obtained from the quality control sample manufacturer. We have established the effect of instrument uncertainty in section 3.1; so we represent the fractional change (at time $t$) in absorbance due to pipetting and clock uncertainty in the sample analysis phase as $u_{pc(m)}$. If we denote the true haptoglobin sample concentration as $[Ag_x]$, and represent the associated absorbance as $A_x$ (at time $t$), we can write the value of absorbance after including sample and instrument uncertainty as:

$$A'_x = A_x\ (1+u_x)\ (1+u_{pc(m)})\ (1+u_p) \tag{9}$$

$A^{'}_x$ in the above equation is the absorbance after incorporating the uncertainty of the sample analysis phase, and $u_x$ is the fractional change in sample haptoglobin concentration as a result of preanalytical uncertainty. The term $u_p$ represents photometer uncertainty - we emphasize here that this term represents the random variable characterizing photometer uncertainty, and therefore its value might be different each time it is sampled in the calibration and sample analysis phase. This value of the optical absorbance is then converted into the haptoglobin concentration in the patient sample by equation 8, as shown by the expression below:

$$C'_x = \left[\frac{(p'_1 - (A^{'}_x - p'_4))}{p'_2\ (A^{'}_x - p'_4)}\right]^{1/p'_3} \tag{10}$$

The net uncertainty of the above model is quantified by generating sample haptoglobin concentration values for a predetermined number of sets of randomly generated realizations of the sources of uncertainty (in this implementation, 100, based on the average number of haptoglobin assays conducted in one day

on such an analytics system in a typical laboratory). The standard deviation of these 100 haptoglobin concentration values is then estimated.

We depict the haptoglobin assay uncertainty model in a flow diagram below.

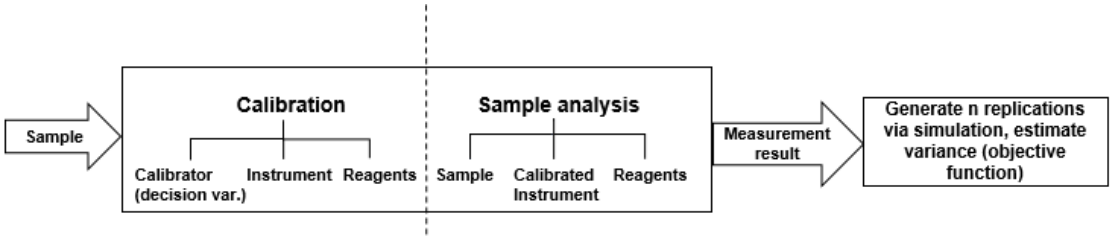


Figure 1: Haptoglobin assay uncertainty model.

Due to instrument access limitations, we were unable to directly validate the model by performing controlled experiments. Therefore, in addition to verifying the mathematical model of the assay biochemistry with two independent chemists, we compared net uncertainty estimates generated by the model to estimates of uncertainty generated under repeatability conditions, which were made available to us by the assay manufacturer. We adopted this approach towards model validation because of the following reasons. 1) The distributional parameters of the sources of uncertainty are based on specifications obtained by the assay manufacturer, and these specifications in turn are based on experimental data available for the instrument components and associated sources of uncertainty. Hence it followed that uncertainty estimates generated by the model are best compared to experimental estimates available to the manufacturer. 2) The estimates of uncertainty for the haptoglobin assay generated under repeatability conditions utilized samples with substantially lesser preanalytical uncertainty in comparison with patient samples usually evaluated in clinical laboratories, and we therefore reasoned that this is an appropriate comparison because we do not account for preanalytical uncertainty in the model.

Net measurement uncertainty estimates generated under repeatability conditions for the haptoglobin assay were supplied by the instrument manufacturer as coefficients of variation (CVs) of 1.4%, 1.5% and 1.7% at haptoglobin concentrations of 91.6 mg/dL, 120 mg/dL and 267 mg/dL, respectively. At these haptoglobin concentrations, model estimates of uncertainty were 1.3%, 1.38% and 1.71%. It can thus be inferred that the model underestimates the net measurement uncertainty of the haptoglobin assay at lower concentrations and overestimates the assay uncertainty at higher concentrations. However, since these comparisons are not equivalent to validation via the conduct of controlled experiments, they do not support definitive conclusions regarding the validity of the haptoglobin uncertainty model. Further, given that the overall trend of the model's estimates is the same as that of the manufacturer's estimates, and the differences between both sets do not exceed 10%, these comparisons indicate qualitatively that the model's estimates of uncertainty are reasonable in comparison to those seen in the laboratory.

## 4 OPTIMIZATION PROBLEM FORMULATION AND SOLUTION

Two formulations for the problem of finding optimal calibrator concentration levels suggest themselves: identify calibrator concentration levels that minimize the assay measurement uncertainty (a) at a medical decision point, or (b) across the possible range of patient sample concentrations. Given the preliminary nature of the work in this article, we focus our efforts on the first objective in this study.

The uncertainty associated with the assay can be written as follows:

$$u^2(C') = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \left( \left( \frac{p'_1 - A'_i + p'_4}{p'_2 \, (A'_i - p'_4)} \right)^{1/p'_3} - \bar{C}' \right)^2 \right] \tag{11}$$

The above expression represents the estimated variance of the result of the haptoglobin assay, assuming $N$ measurements are taken at a desired patient sample concentration of $C$, which may represent a medical decision point. The $i^{th}$ ($i$ = 1-$N$) estimated sample concentration $C_i'$ (we use the primed notation henceforth to indicate that the measurements are subject to uncertainty) is obtained by inputting the measured absorbance $A_i'$ into the calibration function, also subject to uncertainty from the calibration phase (equation 8). $\bar{C}'$ represents the mean of the $N$ measurements.

Note that the $p_i'$ ($i$ = 1-4) can be considered to be functions of the calibrators $Ag_i'$ ($i$ = 1-6). That is, $p_i'$ = $g_i(\mathbf{Ag}')$, $i$ = 1-4. Also note that the vector $\mathbf{Ag}'$ represents the vector of calibrator concentration values ($Ag_i'$) subject to measurement uncertainty ($i$=1-6).

The objective function for the optimal calibrator selection problem at medical decision points can be stated as follows:

$$\begin{aligned} &\underset{\mathbf{Ag}}{\text{minimize}} \ u^2(C') \\ &\text{subject to} \\ &Ag_i \in I_i^+ \ , \ i = 1-6 \end{aligned} \tag{12}$$

We assume that the $Ag_i$ ($i$ = 1-6) will assume a finite set of non-negative values (in the above formulation, $I_i^+$ represents the finite set of non-negative values that a calibrator $Ag_i$ can possibly take), and hence the problem in equation 12 becomes a discrete optimization problem. Ranking and selection techniques therefore are a natural choice to identify the optimal **Ag**, provided the number of possible solutions is not very large. In applying ranking and selection methods to solving the above problem, we have assumed that manufacturers of calibrators have the capability to manufacture only limited sets of calibrators. We apply two methods: the NSGS procedure (Nelson et al. 2001), which we apply when the number of feasible solutions is small (20-25), and the KN procedure (Kim and Nelson 2001), which we apply when the number of feasible solutions is both small (20 systems) and relatively large (100 systems, 400 systems).

We provide a short description of the NSGS and KN procedures for ranking and selection. More detailed descriptions are given in Kim and Nelson (Kim and Nelson 2006).

The NSGS procedure requires that the replications from each system (i.e., a feasible solution - in this case, a set of calibrators) are i.i.d. and follow the Gaussian distribution, and that the replications from each system be independent of those from other systems. This implies that common random numbers cannot be used for these systems. The KN procedure also requires i.i.d. Gaussian replications from each system, but allows common random numbers. Both procedures employ the indifference zone approach - the methods guarantee that, given $k$ systems, the best system will be selected with a probability $1-\alpha$, given that the difference between the best system and the next best system is at least $\delta$. The experimenter is "indifferent" to systems within $\delta$ of the best system yielding the "indifference zone" nomenclature for this type of algorithm. The values of $\alpha$ and $\delta$ can be set by the experimenter. Both approaches involve generating an initial set of replications $n_0$ for each system, and then based on the values of $\delta$ and $\alpha$, identifies a subset $B$ of the original $k$ systems that is guaranteed to contain the best system with $1-\alpha$ probability. This is the screening stage. Note that the NSGS procedure utilizes Rinott's constant (Wilcox 1984) and the estimated variances for each system during the screening stage. The KN system uses only the variances of the pairwise differences between the replications of each system.

Once the subset $B$ is identified, both procedures move on to the ranking and selection stage, and the NSGS and KN procedures diverge in a significant manner in this stage. For each system in $B$, the NSGS procedure calculates the number of additional replications to be generated from the simulation. The mean for each system is updated after these additional replications are generated, and the system with the lowest mean is selected as the best system.

The KN procedure is a fully sequential procedure in the ranking and selection stage. The procedure involves generating an additional replication for each system and updating the subset $B$ using the screening criteria until only system is left in $B$. This is the best system as selected using the KN procedure.

Note that even though the NSGS procedure can be applied to a large number of systems - Nelson (Nelson et al. 2001) demonstrates its application for up to 500 systems - we prefer the KN procedure when the number of systems is relatively larger (100, 400, etc.) because its sequential ranking and selection stage makes it computationally less expensive.

We now describe the solution of the above problem when a small number of feasible solutions are available. The current default set of six calibrator concentration levels used by the assay manufacturer are as follows: 0, 32.9, 65.8, 131.5, 263 and 601 mg/dL. The 0 mg/dL calibrator is a water blank, and we assume that it is a fixed value for any feasible set of calibrators - therefore, the number of variables in the problem reduces to five. We demonstrate the use of the NSGS procedure to solve the above optimization problem assuming that the number of feasible solutions is 20. Each of these 20 solutions is generated randomly in the following manner: each calibrator concentration level in an alternative calibrator set is picked randomly from a set of ten values in the neighbourhood of the default value, each of which differ by 1. For instance, these ten alternative values that may substitute for $Ag_2 = 32.9$ are the set $\{27.9, 28.9, 29.9, ..., 37.9\}$. The reference range in humans for haptoglobin concentration values is 30-250 mg/dL, so the medical decision point (patient sample concentration) is chosen to be 250 mg/dL.

29 replications, with each replication being the variance of the haptoglobin assay at a patient sample concentration level of 250 mg/dL, are generated for each of the 20 systems, and then their means and variances are calculated. A normality test (the Anderson-Darling test) was conducted for the replications from these systems, and the replications from each system were found to be normally distributed. A $\delta$ value of 3 was chosen and the $\alpha$ value was set at 0.1. Six systems were included in the screening subset $B$ after the screening stage, and the number of additional replications required for each system was calculated. Upon generating the additional replications and updating the average of the variances for the systems in $B$, the system with calibrator concentration levels $\{0, 29.9, 63.8, 134.5, 263, 604\}$ was identified as the best system.

In order to verify that using the best system does indeed yield a reduction in uncertainty over the other feasible systems, we compared the uncertainty estimated for the best system and the "worst" system (the system identified as having the largest uncertainty based on the $n_0$ replications generated during the screening stage) by generating 500 replications for each. Based on these 500 replications, the best system yielded a reduction of x% in uncertainty when compared to the worst system. Note that we compare the best system to the worst system (and not, for instance, to a "mid-range" system) because each of these systems are equally likely to be picked in the absence of such information.

We then apply the KN procedure to the above scenario (with the same 20 systems), and to scenarios with 100 systems and 400 systems. The set of feasible solutions are generated in manner similar to that for the 20-system scenario with the NSGS procedure. In each case, we use a $\delta$ value that is based on the average differences between the mean values of the replications from each system. For each scenario, Table 2 lists the values of $\delta$, the cardinality of $B$ after the screening stage, the calibrator set identified as the best system, and the difference between the best and worst systems estimated based on 500 replications from each system during the verification stage.

Table 2: Results of applying NSGS and KN procedures.

| Algorithm, size of solution space | $\delta$, $n_0$ | Best System | Uncertainty (95% CI), best system (%) | % reduction from worst system (95% CI) | $\|B\|$ |
|---|---|---|---|---|---|
| NSGS, 20 | 3, 29 | 29.9,63.8,134.5,263,604 | 1.69 (1.64-1.74) | 2.91 (1.60-4.22) | 6 |
| KN, 20 | 3, 29 | 28.9,61.8,128.5,262,602 | 1.65 (1.60-1.70) | 1.60 (0.53-2.67) | 4 |
| KN, 100 | 1, 100 | 29.9,62.8,132.5,262,605 | 1.65 (1.62-1.68) | 2.82 (1.54-4.10) | 64 |
| KN, 400 | 1.5, 100 | 31.9,63.8,128.5,258,605 | 1.66 (1.63-1.69) | 2.77 (1.75-3.79) | 107 |

The value of $\delta$ in each scenario is selected after examining mean values for each system generated in the screening stage (from $n_0$ replications) so that a reasonable number of systems are retained in $B$ after the screening stage. This is done to ensure that the manufacturer has multiple "good" systems to choose from in case the optimal system cannot be used due to constraints that cannot be incorporated into the model. However, $\delta$ is initialized to a minimum value of 1 since we assume it is the least difference between the best and other systems that is worth detecting.

## 5 DISCUSSION AND CONCLUSIONS

We present here a simulation optimization approach to selecting calibrator concentration levels such that the uncertainty of the haptoglobin assay is minimized at medical decision points. This study represents a first step towards exploring the use of simulation optimization methods to optimize bioanalytical assays. Further, given that our search of the literature did not identify any previous work that utilized a simulation optimization approach to optimally select calibrators for a measurement procedure, this study serves as proof of concept for the use of such methods to optimize measurement processes.

From Table 2, the difference between the best and worst systems may not appear to be significant in terms of the reduction in uncertainty at medical decision points. However, the significance of the difference depends on how close the uncertainty with the default set of calibrators already is to the clinically permissible level of uncertainty. Further, we have picked the set of candidate systems at random in the neighborhood of the default set of calibrators to demonstrate the application of our proposed approach, and hence we do not necessarily expect the "best" system to yield a substantial reduction in uncertainty when compared to the default or "worst" set of calibrators. A complete exploration of the solution space will require a continuous formulation of the problem, and application of appropriate (e.g., gradient-based) methods to identify the optimal solution. This is deferred for future work.

Another future avenue of research is the formulation of the problem in terms of minimizing the uncertainty across the range of possible patient sample concentration values instead of focusing only on medical decision points. Weights can be included in the objective function so that minimizing the uncertainty at medical decision points is given the appropriate consideration.

## REFERENCES

Aronsson, T., C.-H. de Verdier, and T. Groth. 1974. "Factors Influencing the Quality of Analytical Methods: A Systems Analysis, with Use of Computer Simulation". *Clinical Chemistry* 20(7):738–748.

Boer, E., D. Rasch, and E. Hendrix. 2000. "Locally Optimal Designs in Non-linear Regression: A Case Study of the Michaelis-Menten Function". In *Advances in Stochastic Simulation Methods*, edited by V. B. Melas and N. Balakrishnan, 177–188. Berlin, Germany: Springer.

Borg, L., J. Kristiansen, J. M. Christensen, K. F. Jepsen, and L. K. Poulsen. 2002. "Evaluation of Accuracy and Uncertainty of ELISA Assays for the Determination of Interleukin-4, Interleukin-5, Interferon-$\gamma$ and Tumor Necrosis Factor-$\alpha$". *Clinical Chemistry and Laboratory Medicine* 40(5):509–519.

Buonaccorsi, J. 1986. "Design Considerations for Calibration". *Technometrics* 28(2):149–155.

Bureau International des Poids et Measures 1993. *Guide to the Expression of Uncertainty in Measurement*. 1st ed. Geneva: Bureau International des Poids et Measures. https://www.iso.org/standard/45315.html. Last accessed: April 12, 2018.

Burtis, C. A., E. R. Ashwood, and D. E. Bruns. 2012. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. Amsterdam, the Netherlands: Elsevier Health Sciences.

Clyde, M., and K. Chaloner. 2002. "Constrained Design Strategies for Improving Normal Approximations in Nonlinear Regression Problems". *Journal of Statistical Planning and Inference* 104(1):175–196.

Forkman, J. 2008. "A Method for Designing Nonlinear Univariate Calibration". *Technometrics* 50(4):479–486.

Gauchi, J.-P., and A. Pázman. 2006. "Designs in Nonlinear Regression by Stochastic Minimization of Functionals of the Mean Square Error Matrix". *Journal of Statistical Planning and Inference* 136(3):1135–1152.

Joint Committee for Guides in Metrology 2008a. *Evaluation of measurement data – Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement" – Propagation of distributions using a Monte Carlo method*. 1st ed. Geneva: Bureau International des Poids et Measures. https://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf. Last accessed: April 9, 2018.

Joint Committee for Guides in Metrology 2008b. *Evaluation of measurement data Guide to the Expression of Uncertainty in Measurement*. 1st ed. Geneva: Bureau International des Poids et Measures. https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Last accessed: April 10, 2018.

Kim, S.-H., and B. L. Nelson. 2001. "A Fully Sequential Procedure for Indifference-Zone Selection in Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11(3):251–273.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". *Handbooks in operations research and management science* 13:501–534.

Krouwer, J. 2002. "Setting performance and evaluating total analytical error for diagnostic assays". *Clinical Chemistry and Laboratory Medicine* 48(6):919–927.

Matthews, J., and G. Allcock. 2004. "Optimal Designs for Michaelis–Menten Kinetic Studies". *Statistics in medicine* 23(3):477–491.

Melas, V. B. 2005. "On the Functional Approach to Optimal Designs for Nonlinear Models". *Journal of Statistical Planning and Inference* 132(1-2):93–116.

Müller, P. 2005. "Simulation Based Optimal Design". *Handbook of Statistics* 25:509–518.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2001. "Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large". *Operations Research* 49(6):950–963.

Pereira, P., B. Magnusson, E. Theodorsson, J. O. Westgard, and P. Encarnação. 2016. "Measurement Uncertainty as a Tool for Evaluating the Grey Zone to Reduce the False Negatives in Immunochemical Screening of Blood Donors for Infectious Diseases". *Accreditation and Quality Assurance* 21(1):25–32.

Pereira, P., J. O. Westgard, P. Encarnação, and J. Seghatchian. 2015. "Evaluation of the Measurement Uncertainty in Screening Immunoassays in Blood Establishments: Computation of Diagnostic Accuracy Models". *Transfusion and Apheresis Science* 52(1):35–41.

Ramamohan, V., J. T. Abbott, G. Klee, and Y. Yih. 2014. "Modeling, Analysis and Optimization of Calibration Uncertainty in Clinical Laboratories". *Measurement* 50:175–185.

Ramamohan, V., J. T. Abbott, and Y. Yih. 2015. "Effect of Uncertainty in Calibration on the Correlation Structure of the Rheumatoid Factor Immunoassay Calibration Function". In *Proceedings of the Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1537–1548. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schwartz, R. S. 2011. "Autoimmune and Intravascular Hemolytic Anemias". In *Goldman's Cecil Medicine, 24th ed.*, edited by L. Goldman and A. I. Schafer. Philadelphia, PA: Elsevier Saunders.

Suchanek, M., and P. Robouch. 2009. "Measurement Uncertainty of Test Kit Results – the ELISA Example". *Clinical Chemistry and Laboratory Medicine* 47(7):808–810.

Wilcox, R. R. 1984. "A Table for Rinott's Selection Procedure". *Journal of Quality Technology* 16(2):97–100.

## AUTHOR BIOGRAPHIES

**UTKARSH VARDHAN** is a junior undergraduate student in the Department of Mechanical Engineering at the Indian Institute of Technology Delhi. His e-mail address is Utkarsh.Vardhan.me215@mech.iitd.ac.in.

**VARUN RAMAMOHAN** is an Assistant Professor in the Department of Mechanical Engineering at the Indian Institute of Technology Delhi. He holds a Ph.D. in Industrial Engineering from Purdue University. His email address is varunr@mech.iitd.ac.in.