An Approach to Specification Setting for the Manufacture of a Diagnostic Immunoassay

Phil Turk

July 1, 1998

Contents

1 Introduction

With the push towards Total Customer Satisfaction in recent years, determining the quality of product produced has become a major issue in industry. Very often, this quality status is dichotomous in nature (pass or fail) and is determined by comparing some measured variable to what are known as specifications. The crux of this paper deals with how one might determine specifications based on an example from the pharmaceutical industry. This approach can be generalized to many other areas.

All too often in the past, specification setting relied more on educated guesswork than statistics. Furthermore, the process by which specifications were set occurred after the product had already been fully developed. If companies today wish to remain competitive and profitable, this can no longer be the case. One suggestion the author would like to

put forth immediately is that any company wishing to embark on improving the quality of their products through specification setting should appoint one or two people to solely concentrate on that purpose. One reason this paper was written was because of the author's experience that those who actually worked in industry had little time to explore and develop new procedures. Yet, by the mid-1980's, the amount of literature on specification setting was substantial and available for anyone to use. Thus, the rewards could be substantial relative to the cost of investment.

A brief description of the example product referenced throughout this paper will now be given. In medicine today, we are oftentimes interested in measuring the quantity of a certain analyte in a patient's blood. A product by which this objective can be accomplished is called a diagnostic immunoassay. Assume that we are interested in a certain analyte which we will call Antigen X. The product used to quantify Antigen X is a kit containing several bottles or reagents. As is representative of many diagnostic immunoassays suppose the kit for Antigen X contains the following components:

1. 6 bottles of calibrators with known increasing concentrations of Antigen X. These are used to generate what is known as a calibration curve which is further used to read the concentration of Antigen X in the patient's blood sample.

2. 3 bottles of standards with known increasing concentrations of Antigen X. These are used to ensure the calibration curve is delivering correct readings. These are called low, medium and high controls.

3. 1 bottle of microparticles. The microparticles are coated with antibody specific to Antigen X. When a patient sample is mixed with the microparticles, any Antigen X will be bound by the antibodies.

4. 1 bottle of conjugate. Conjugate is antibody specific to Antigen X joined with an enzyme called alkaline phosphatase (AP). The conjugate will actually attach to Antigen X that has already been bound by the microparticles. Once this happens, the resulting entity is often referred to in immunology as a "sandwich".

5. 1 bottle of MUP. MUP is a chemical that is turned into what is known as MU by AP. This being done, MU fluoresces. Since the fluorescence will be proportional to the amount of AP present, we have a mechanism by which to measure the amount of bound Antigen X.

6. 1 bottle of buffer solution in which this reaction takes place.

The kit is used in conjuction with both manual setup labor and an automated machine equipped with a fluorometer which can read the counts per second generated by the reaction. Finally, a curve fitting procedure, generating, for example, a 4 parameter logistic model, is used by the machine to enable the user to obtain a concentration reading of Antigen X.

The reader should not have the impression from the previous descriptions that the process of generating an Antigen X concentration reading is quite simple. Experience shows instead it is not! The point for giving the previous descriptions is to provide the reader a brief background to a trilogy of issues that are the crux of specification setting. First, each

component is actually produced across several steps or "stages". Second, each component may have several attributes measured at one or more of its manufacturing stages to make sure all is progressing well. Some of these attributes may be unique to the component. Third, and finally, all the components are brought together to form the kit itself.

We will now discuss specification setting for one response from one component.

2 A univariate response: the simplest case

Consider the example of the immunoassay kit designed to detect and quantify an antigen X. We are dealing with a kit of multiple components, each of which plays a role in generating the final responses seen by the customer. For illustrative purposes, consider a particular response. One of the components of the kit is a known low control standard which the customer uses to make sure the kit is working ("reading") correctly via a concentration reading. We have the following relationship

$$\sigma^2_{\text{low}} = \sigma^2_p + \sigma^2_m$$

where σ^2_{low} is the variance associated with the low control reading, σ^2_p is the variance associated with the low control lot-to-lot production process, and σ^2_m is the variance associated with our ability to actually measure the low control reading.

A major issue immediately arises. Note that for our example σ^2_{low} will include not just the lot-to-lot variation associated with the low control production process, but also includes the effect of **all** the other kit component production processes as well. This is because some components could be interrelated with one another such that a change in one might impart a change in another or even several others. Thus, setting specifications just for the low control concentration by itself fails to consider the kit as a holistic entity because setting specifications for this particular response whether by manipulating any one component, reducing the production process variability or devising a particular measurement scheme might not be an appropriate course of action for other responses. This necessitates using a multivariate approach. Initially, our discussion will center around the simplest scenario, the effect of one component on a single response. However, later in this paper we will briefly look at the issue of multiple components and multivariate responses to obtain a holistic view of the kit. A thorough investigation of these particular topics would require additional intensive research.

2.1 A Multistage Process

There are many examples in industry today where a particular product has to be manufactured across multiple, sequential stages before it reaches the customer. Once again, consider the low control example. The production pathway for the low control follows four critical manufacturing stages. First, at Stage 1, a research team designs the initial formulation at a very small scale. This is done each time low control is made, because new lots of material are used and the procedure involved in making it is extensive. Once this very small lot of low control is made, it is measured using components from a "gold standard" kit to ensure that it is on target before being passed on to Stage 2. At Stage 2, the low control "recipe" is handed

over to another factory where scale up to a huge volume occurs. Again, the low control must measure accurately against the gold standard kit before it can go onto the next stage. At Stage 3, the bulk low control is aliquoted and batched with the other components that go into the kit. Stage 3 is where the kit officially comes into existence, being the combination of several different components, each with their own respective manufacturing process. Lastly, at Stage 4, the antigen X kits are shipped to the customers where they will be used by the customers for patient testing.

2.2 Decision Making and Error

At each stage of a manufacturing process, a decision is made as to whether the product being produced should be sent to the next stage. The criteria upon which this decision is made are called specifications. Typically, if the product falls between a lower specification and a higher specification, the product is deemed "good" and sent on to the next stage. However, if the product falls outside the lower and higher specifications, the product is deemed "bad" and the manufacturing process stops. It is important to note that a correct decision can never be made with 100% certainty. One reason, of course, is due to the inherent variability in the production process. The focus of this paper, however, will deal with the notion that our ability to measure what is being made is limited by factors such as equipment, human error, environment, etc. Because our decision rule is guided solely by the specifications themselves, our ability to measure what we have made (also known as measurement error) can lead to incorrect decisions. For example, a company may consistently make "good" product but measurement error might cause the product to read outside the specifications, thus appearing "bad". Conversely, measurement error can cause "bad" product to read inside the specifications, thus appearing "good". Table 1 summarizes the four possible consequences of our decision: To limit the errors in our decision making process it is necessary

	True State of Product	
Decision Made	Good	Bad
Good	No error	Loss to customer
Bad	Loss to company	No error

Table 1: Possible losses for any decision rule

to determine the specifications that minimize the amount of customer complaints. In the past, manufacturers often used control limits as specifications and emphasized their costs and profits while customer complaints were a secondary concern. Recently, more attention has been focused on Total Customer Satisfaction and allowing the customer to set the final specifications. The manufacturer is then responsible for making product that meets these customer needs. These needs are balanced with the necessity of minimizing costs to the company.

Let's focus our attention on decision error caused by measurement variation. If the company has an estimate of the amount of measurement variability the customer typically sees (at stage 4), this variability estimate can be deducted from the customer specifications to

act as a buffer. This gives the manufacturer an idea of where the specifications need to be set at stage 3 in order to produce the type of product the customer requires. To illustrate, let's say the customer wants the upper specification for the low control from our kit example to read 1.20 units/ml. Previous studies have shown that the measurement standard deviation customers have seen if they run the assay as instructed is .04 units/ml. We know, assuming the normal distribution for measurement error, that were the low control to have a true lot mean of 1.20, it could read anywhere from 1.08 to 1.32 over 99.7% of the time due to measurement variability. The consequences of shipping a lot with a true mean of 1.20 to the customer then would be that half of the time, the lot will read outside the upper specification and possibly generate a customer complaint. To compensate for this, the company can back off the customer specification by three measurement standard deviations and set their stage 3 upper specification at 1.08. In this way, if the company makes sure that at stage 3 they never manufacture a lot of low control greater than 1.08, then the probability of the customer reading the lot as being greater than 1.20 is very low!

With our multistage process, the specification setting procedure can become more complex. Continuing our example, at stage 3, can the company be sure it is making product with a true mean no greater than 1.08? Doesn't the company have to concern itself with the measurement variability at stage 3? The answers are no and yes, respectively. A possible approach to address these issues would be to extend our "failsafe" philosophy. The thinking here is that each stage will have a certain amount of unique measurement variability which leads to an adjustment of its specifications at that stage. For our illustration, if the company knows that the stage 3 measurement standard deviation is .01, the stage 3 upper specification will actually have to be 1.08 - $(3 \times .01) = 1.05$.

Again, our philosophy behind adjusting specifications for measurement error is ultimately to meet the customer requirements. One must be careful sequentially cascading specifications inwards as one marches backwards through the manufacturing stages because setting specifications too stringently may increase company costs due to failed product.

At this point, we need to introduce the notation and an understanding of the problems associated with each stage before we can actually develop a statistical approach to setting the specifications.

2.3 Random Variables and Decision Problems

For a given process, we assume a state of statistical control i.e. the variability present in the production process is confined to chance variation. It is up to the individual company via their quality control standards to verify the validity of this nontrivial assumption. For any stage, process variable levels are adjusted to provide measurements about the process target μ. Let the random variable X be the true unobserved value for any given lot or batch and assume that $X \sim N(\mu, \sigma_p^2)$ where σ_p^2 is the lot-to-lot variance. Also, define a random variable D that represents the difference between the observed and unobserved responses. This difference represents our measurement error. Assume $D \sim N(0, \sigma_m^2)$ where σ_m^2 is the measurement variance for a given manufacturing stage and that D is independent of X. Therefore $Y = X + D$ will be a random variable that represents an observed response. $Y \mid X$ is also assumed to be $N(X, \sigma_m^2)$. By integrating the joint distribution $f_{X,Y}(x, y)$ over X, it is

easy to obtain the resulting distributions $Y \sim N(\mu, \sigma_m^2+\sigma_p^2)$ and $X \mid Y \sim N(\frac{y\sigma_p^2+\mu\sigma_m^2}{\sigma_m^2+\sigma_p^2}, \frac{\sigma_m^2\sigma_p^2}{\sigma_m^2+\sigma_p^2})$. It is possible, if not probable, that the distributions for the last three stages will vary due to the measurement variability being different from stage to stage. As an example, this could be due to different stagewise measurement schemes. Therefore, we will denote σ_{m2}^2, for example, as being the measurement variability associated with stage 2.

Returning to our example from the immunoassay kit, we define for each of the 4 stages of the manufacturing process, a lower specification l_i and an upper specification u_i for the ith manufacturing stage. The production pathway would proceed as follows. First, at stage 1 there exists an unobservable X_1 and an observable Y_1. To reiterate for this example, X_1 is the true, unknown mean for the low control lot, while Y_1 is the measured, observed response value we obtain. Whereas this stage is one of initial development, there are no stage 1 specifications and the product automatically goes on to stage 2. Next, at stage 2 there exists an unobservable X_2 (recall the low control is remade at this stage) and an observable Y_2. If $Y_2 \in (l_2, u_2)$, the product passes on to stage 3. If $Y_2 \notin (l_2, u_2)$, the product fails and is either disposed of or reworked. At stage 3, we still deal with the same unobservable X_2 but now have an observed Y_3. This is because the same lot is simply being remeasured, not remanufactured. It is highly unlikely the responses Y_2 and Y_3 will be the same because of measurement variation. The product will pass to stage 4 if $Y_3 \in (l_3, u_3)$, otherwise it will fail and be disposed of or reworked. At stages 2 and 3, there must be a very low probability that $Y_2 \notin (l_2, u_2)$ or $Y_3 \notin (l_3, u_3)$ respectively, since in these cases the company will incur a "loss". Finally, at stage 4, we are still dealing with the same unobservable X_2 but now have an observed Y_4. Again, we are dealing with the same lot being remeasured. Since this is the stage where actual customer testing takes place, there must be a very high probability that $Y_4 \in (l_4, u_4)$, otherwise the product will "fail" in the eyes of the customer and a potential complaint to the company could occur. It is important to note that given X_2, the observed responses Y_2, Y_3, and Y_4 are independent of each other and vary only due to measurement variability.

There will be specific problems in the manufacturing of the low control that are unique to each of the stages. Each of these problems must be addressed before the decision to ship the low control material along to the next stage can be made. At stage 1, the small-scale recipe for the low control must be accurate with good precision. In other words, the low control must be at or near the target with small measurement variability in terms of repeatability. Repeatability is defined in this sense as multiple measurements on the same lot. Next, at stage 2, we have to readjust the settings of our process for each new lot of low control. To be at or near the target, the process variability, σ_p^2, must be low. Furthermore, the measurement error must now be low both in terms of repeatability and reproducibility. For this example, reproducibility is the variability that occurs when taking multiple measurements across different measurement factors (machines, time, etc.). These measurement factors may even interact with the now present lot effect. Stage 3 brings an important special problem because of the introduction of other kit components and their respective variance components. If the low control lot had passed stage 2, but failed stage 3, some other kit component might be the cause. There could be other measurement factors arising in this stage along with potential interaction involving the process factors. Finally, at stage 4 even more measurement factors are being introduced which inflate the measurement

error. Hence, we expect the relationship $\sigma^2_{m4} > \sigma^2_{m3} > \sigma^2_{m2} > \sigma^2_{m1}$.

2.4 Defining Costs and Goals

Some general notions can be drawn in terms of defining costs associated with a particular manufacturing process First, the company will incur fixed costs from the testing of the product at each stage of the manufacturing process. These costs may even vary from stage to stage. Second, the company will incur a cost when a product lot is rejected. Included in this cost would be lost profits, down time, product rework, wasted materials, disposal, etc. This cost would depend on the stage at which the product was rejected. For example, a lot that failed at stage 3 would be more costly to the company than if the lot failed at stage 2. Additionally, if the product was to fail in the hands of the customer, the company might incur an additional cost in terms of a product recall and replacement. Collectively, all these aforementioned points are defined as *tangible company costs.*

One of our goals is to minimize the tangible company costs. This is done by minimizing:

$$\begin{aligned} E[\text{cost}] \;=\; & [\text{cost incurred at stage 1}] \times P[\text{rejecting at stage 1}] \\ & + \sum_{i=2}^{n} [\text{cost incurred at stage } i] \times P[\text{rejecting at stage } i \mid \text{passed stage } i - 1] \end{aligned}$$

where n is the number of manufacturing stages the product goes through while in the hands of the company. For our low control example, we are therefore trying to minimize the following:

$$\begin{aligned} E[\text{cost}] \;=\; & [\text{cost incurred at stage 1}] \times P[\text{fail stage 1}] \\ & + [\text{cost incurred at stage 2}] \times P[\text{fail stage 2} \mid \text{pass stage 1}] \\ & + [\text{cost incurred at stage 3}] \times P[\text{fail stage 3} \mid \text{pass stage 2}]. \end{aligned}$$

Finally, there exists a second category of costs defined as *intangible customer complaint costs.* These types of costs can be assumed by either the company and/or the customer. However, they are difficult to measure and assess in terms of a dollar amount. For example, there would be a cost to the company in terms of loss of customer satisfaction if the lot "failed" in his or her hands. Also, there exists a cost to the customer if he or she rejects the product as being of poor quality. Such a cost might be, for example, loss of potential revenues due to down time. So, the other goal we are striving for is to minimize customer complaints. In other words, assuming each complaint generates a constant cost, we seek to minimize:

$$E[\text{potential number of customer complaints} \mid \text{passed stage } n].$$

This expectation equals:

$$(\text{number of kits produced}) \times P[\text{fail stage } n+1 \mid \text{pass stage } n].$$

Because the number of kits produced is fixed, it turns out that the expected number of customer complaints can be minimized by minimizing the following probability:

$$P[\text{product fails customer specs} \mid \text{passed last manufacturing stage}].$$

In our low control example this would be:

$$P[\text{product fails customer specs} \mid \text{pass stage 3}].$$

It is important to note that because a value cannot be assigned to the intangible customer complaint costs, it will be on a different and separate scale from the tangible company costs. Therefore, comparisons between the two types of costs will be difficult. We already know that tangible company costs are directly minimized by minimizing $E[\text{cost}]$. This goal could be accomplished, at the expense of the customer, by widening the specifications such that the company never rejected a lot. On the other hand, intangible customer complaint costs are minimized by minimizing $E[\text{number of customer complaints} \mid \text{passed stage } n]$. This goal is accomplished, at the expense of the company, by tightening the specifications such that the customer never "rejected" a lot. Hence, from an optimization standpoint, we must balance the needs of the customer versus the costs to the company. By working towards balancing these two goals, we can develop a sound, cost effective approach to specification setting. Otherwise, by favoring either aspect of this balance, the <u>total</u> costs will rise.

2.5 Probability Calculations

To develop the probability theory behind our approach, we again refer to the immunoassay kit example. It is possible that some of the probability calculations could be inaccurate because of a process shift from the target, an increase in process variability, or even both. In this paper, however, we assume that all processes have been studied and understood, and are in a state of statistical control. Finally, notice the use of conditional arguments. It would be incorrect to assume that probabilities associated with the various stagewise events are automatically independent.

In the previous subsection, one of our goals was to calculate and minimize the following:

$$P[\text{product fails customer specs} \mid \text{pass stage 3}]$$

which can be rewritten as:

$$\frac{P[\text{fail stage 4 and pass stage 3}]}{P[\text{pass stage 3}]} = \frac{P[Y_4 \notin (l_4, u_4) \text{ and } Y_3 \in (l_3, u_3)]}{P[Y_3 \in (l_3, u_3)]}$$

This calculation will broken into two pieces. We first start with the denominator. Notice that:

$$\begin{aligned} P[\text{pass stage 3} \mid \text{pass stage 2}, X_2 = x_2] &= \frac{P[\text{pass stage 3 and pass stage 2} \mid X_2 = x_2]}{P[\text{pass stage 2} \mid X_2 = x_2]} \\ &= \frac{P[Y_3 \in (l_3, u_3) \text{ and } Y_2 \in (l_2, u_2) \mid X_2 = x_2]}{P[Y_2 \in (l_2, u_2) \mid X_2 = x_2]} \end{aligned}$$

Given $X_2 = x_2$, note that Y_2 and Y_3 are independent, thus:

$$P[\text{pass stage 3} \mid \text{pass stage 2}, X_2 = x_2] = P[\text{pass stage 3} \mid X_2 = x_2] = P[Y_3 \in (l_3, u_3) \mid X_2 = x_2]$$

Similarly:

$$\begin{aligned} P[\text{pass stage } i \mid \text{pass stage } i-1,\, X_{i-1} = x_{i-1}] &= P[\text{pass stage } i \mid X_{i-1} = x_{i-1}] \\ &= P[Y_i \in (l_i, u_i) \mid X_{i-1} = x_{i-1}] \end{aligned}$$

assuming that $X_{i-1} = x_{i-1}$, the unknown true mean, is the same for all subsequent stages.

Given the normality assumption:

$$P[\text{pass stage } 3 \mid X_2 = x_2] = \Phi(\frac{u_3 - x_2}{\sigma_{m3}}) - \Phi(\frac{l_3 - x_2}{\sigma_{m3}})$$

where σ_{m3} is the SD of $Y_3 \mid X_2 = x_2$ and Φ is the standard normal CDF. Next, we remove the dependence on $X_2 = x_2$ by integrating over the range on $X_2 = x_2$:

$$\begin{aligned} P[\text{pass stage } 3] &= \int_{-\infty}^{\infty} P[\text{pass stage } 3 \text{ and } X_2 = x_2] dx_2 \\ &= \int_{-\infty}^{\infty} P[Y_3 \in (l_3, u_3) \mid X_2 = x_2] f_{X_2}(x_2) dx_2 \end{aligned}$$

Because $F_{X_2} = \Phi(\frac{x_2 - \mu}{\sigma_p})$, the pdf $f_{X_2} = \frac{1}{\sigma_p}\phi(\frac{x_2 - \mu}{\sigma_p})$. We can rewrite:

$$P[\text{pass stage } 3] = \int_{-\infty}^{\infty} [\Phi(\frac{u_3 - x_2}{\sigma_{m3}}) - \Phi(\frac{l_3 - x_2}{\sigma_{m3}})] \frac{1}{\sigma_p}\phi(\frac{x_2 - \mu}{\sigma_p}) dx_2 \qquad (1)$$

We approximate this probability using numerical integration to finally evaluate the denominator.

To find the numerator, we use the same logic as before. That is, we first condition on $X_2 = x_2$ and use the fact that Y_3 is conditionally independent of Y_4. Then, we integrate over the range on $X_2 = x_2$. First, note that:

$$\begin{aligned} P[\text{fail stage } 4 \text{ and pass stage } 3 \mid X_2 = x_2] &= P[\text{fail stage } 4 \mid \text{pass stage } 3,\, X_2 = x_2] \\ &\quad \times P[\text{pass stage } 3 \mid X_2 = x_2]. \end{aligned}$$

Because Y_3 and Y_4 are independent of each other, given $X_2 = x_2$:

$$P[\text{fail stage } 4 \text{ and pass stage } 3 \mid X_2 = x_2] = P[\text{fail stage } 4 \mid X_2 = x_2] P[\text{pass stage } 3 \mid X_2 = x_2].$$

To remove the dependence on $X_2 = x_2$:

$$\begin{aligned} P[\text{fail stage } 4 \text{ and pass stage } 3] &= \int_{-\infty}^{\infty} P[\text{fail stage } 4 \text{ and pass stage } 3 \text{ and } X_2 = x_2] dx_2 \\ &= \int_{-\infty}^{\infty} P[\text{fail stage } 4 \text{ and pass stage } 3 \mid X_2 = x_2] \\ &\quad \times f_{X_2}(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} P[\text{fail stage } 4 \mid X_2 = x_2] P[\text{pass stage } 3 \mid X_2 = x_2] \\ &\quad \times f_{X_2}(x_2) dx_2 \end{aligned}$$

Or, rewritten:

$$= \int_{-\infty}^{\infty} P[Y_4 \notin (l_4, u_4) \mid X_2 = x_2] P[Y_3 \in (l_3, u_3) \mid X_2 = x_2] f_{X_2}(x_2) dx_2$$

which is equivalent to:

$$= \int_{-\infty}^{\infty} [\Phi(\frac{l_4 - x_2}{\sigma_{m4}}) + 1 - \Phi(\frac{u_4 - x_2}{\sigma_{m4}})][\Phi(\frac{u_3 - x_2}{\sigma_{m3}}) - \Phi(\frac{l_3 - x_2}{\sigma_{m3}})] \frac{1}{\sigma_p} \phi(\frac{x_2 - \mu}{\sigma_p}) dx_2 \quad (2)$$

where σ_{m4} is the SD of $Y_4 \mid X_2 = x_2$. We approximate this probability using numerical integration to evaluate the numerator.

Finally, we can calculate our desired objective from (1) and (2) as:

$$P[\text{product fails customer specs} \mid \text{pass stage 3}] = \frac{P[\text{fail stage 4 and pass stage 3}]}{P[\text{pass stage 3}]} \quad (3)$$

Before leaving this section, note that similar calculations could be extended to other stages further back in the manufacturing chain in order to calculate $E[\text{cost}]$ from section 2.4.

2.6 An Illustrative Example

In industry, one might be interested in automating the specification setting procedure through the use of a computer. Although this certainly can save time and allow for an examination of a variety of interesting simulations, it is strongly recommended that such an automation not be used as a "black box". Part of the American Statistical Association's ethical guidelines states that we must "recognize that automated statistical computation alone does not constitute adequate statistical analysis; it is also necessary to understand the theory, the data, and the methods used in each statistical study." (ASA, (1998)).

Discussion will now focus on Reference 1. This would be the basic framework for an automated spreadsheet for our example using the low control from Antigen X. In this case, we have chosen to use Excel. The spreadsheet could easily be used as a front end overlay to a SAS program. The portion of the SAS program using Proc IML in order to calculate Equation (3) is on Reference 2. Basically, the user inputs into the spreadsheet through Excel while SAS analyzes the data and outputs to the spreadsheet. To facilitate understanding, the spreadsheet will now be discussed in pieces.

First, a gauge capability study was conducted to determine significant measurement factors for reading low control concentration. This type of study is also known as a gauge (gage) R and R study. The actual design and implementation of such studies is another topic in of itself and is not discussed in this paper. For an introductory discussion of such studies, see Montgomery and Runger (1993). Regardless of the details of the gauge capability study, it is critical that such a study use a probability sampling design to ensure the factors are truly random effects. Studies not incorporating this planned randomness will not allow any conclusions reached to extend beyond the sample.

Using the modified method-of-moments estimation procedure described by Milliken and Johnson (1992), the three significant measurement factors are listed at the top of the spreadsheet along with their associated degrees of freedom. Also listed at the top are approximate

Reference 1.

Antigen X Kit Specifications

Response: Low Control Concentration

Results of Reduced Model

Measurement Factors	Estimated Variance Components	DF	Approximate 95% Confidence Limits Lower	Upper
Machine	0.000074	18	0.000000	0.000240
Machine*Lot	0.000228	36	0.000000	0.000470
Rep	0.001318	216	0.001100	0.001600

The Upper Customer Specification is = 1.20
The Lower Customer Specification is = 0.60

If the customer runs the following scenario...

Machine	Machine*Lot	Rep
1	1	1

The Estimated Measurement Error SD for this scenario is = 0.040249

Therefore, we must minimize the risk of ever passing material with the following criteria:

Lot has true mean greater than ... 1.08
Lot has true mean less than ... 0.72

	Stage 3
Upper Specification	1.05
Target	0.90
Lower Specification	0.75
# of Machine	6
# of Machine*Lot	6
# of Rep	60
Est. Measurement SD	0.008503
Est. Process SD	0.043635
Est. Total SD	0.044456

Potential Number Of Customer Complaints Per Million Given The Lot Has Passed Stage 3
----------> 0.12

Reference 2.

```
proc iml;
reset log;

  start fun1(x_2);
  u_3 = 1.05;
  l_3 = .75;
  sig_m3 = .008503;
  sig_p = .043635;
  mu = .90;
  t = ((x_2 - mu)/sig_p);
  normpdf = exp(-(t**2)/2);
  denom = ((probnorm((u_3 - x_2)/sig_m3)) - (probnorm((l_3 - x_2)/sig_m3)))*(1/sig_p)*normpdf;
  return(denom);
  finish;
a = {0 .P};
call quad(denom, "fun1", a);
denom = (1/(sqrt(arcos(-1)*2)))*denom;

  start fun2(x_2);
  u_3 = 1.05;
  l_3 = .75;
  sig_m3 = .008503;
  sig_p = .043635;
  mu = .90;
  t = ((x_2 - mu)/sig_p);
  normpdf = exp(-(t**2)/2);
  u_4 = 1.2;
  l_4 = .6;
  sig_m4 = .040249;
  num = ((probnorm((l_4 - x_2)/sig_m4)) + 1 - (probnorm((u_4 - x_2)/sig_m4)))*
        ((probnorm((u_3 - x_2)/sig_m3)) - (probnorm((l_3 - x_2)/sig_m3)))*(1/sig_p)*normpdf;
  return(num);
  finish;
a = {0 .P};
call quad(num, "fun2", a);
num = (1/(sqrt(arcos(-1)*2)))*num;

print denom[format = E21.14];
print num[format = E21.14];
result = (num/denom)*1000000;
print result;
quit;
```

100(1 - α)% confidence intervals for the estimated variance components. These intervals are based on asymptotic results, so large sample sizes are required. In some gauge capability studies this assumption will not be met. Therefore, it is recommended that exact confidence intervals be calculated if possible or confidence intervals be calculated using Satterthwaite's approximation. How to estimate variance components from the several procedures available and the calculation of their confidence intervals is also not the focus of this paper. Here again, Milliken and Johnson (1992) would be a good reference on this topic as well as Burdick and Graybill (1992).

Next, the section in the middle of the spreadsheet has been described previously in this text. The general form for the calculation of the estimated measurement error standard deviation is:

$$\hat{\sigma}_m = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \cdots + \frac{\hat{\sigma}_k^2}{n_k}}$$

where the $\hat{\sigma}_k^2$s are the estimated variance components associated with the k significant measurement factors and n_i for $i = 1, \ldots, k$ is the actual number of replicates (repeats) the customer is running for each of the k significant measurement factors.

Finally, our attention is turned towards the grid at the bottom of the spreadsheet. It is here we place the result for Equation (3) using the appropriate estimated standard deviations listed at the bottom of the grid. The estimated measurement standard deviation is calculated from the stage 3 measurement scheme using the general calculation from the previous paragraph. An estimate of the process standard deviation was obtained in this case from a control chart. Notice this example contains specifications adjusted for Stage 3 measurement error as described back in Section 2.2. However, these specifications can be edited by the user such that they reflect various measurement schemes as costs and time permit. Thus, specifications will change according to the measurement standard deviation. One may even elect to not adjust the specifications for measurement error at all. The corresponding Equation (3) result will help assess each measurement scheme's effectiveness or lack thereof. It is also strongly recommended that a table of expected costs be generated for the different candidate specification levels.

There are a couple of other issues in need of attention concerning our spreadsheet. One will be the resolution of the measurement unit itself. In other words, a common question might be "How many decimal places should I use to record the response?". A general statement would be that it is worse to record too few places rather than too many. Wheeler (1989) has discussed this topic and what he calls the effective resolution of measurement. The second issue is that there are other diagnostics used in industry to assess quality control that one might be interested in adding to the spreadsheet. Such examples would the C_p or C_{pk} ratios (Montgomery (1997)) or the Discrimination ratio (Wheeler (1989)). It would be relatively easy to add these and other diagnostics to the spreadsheet as the situation dictates.

In the low control example discussed throughout this paper, notice that on Reference 1 we have a potential number of customer complaints per million tests given the lot has passed stage 3 as being equal to .12. Obviously, this is a highly desirable result. It is critically important to conduct simulations whenever using an automated procedure such as this one. For example, the specs at stage 3 where deliberately set to (.72, 1.08) i.e., they

were not adjusted for stage 3 measurement error. Next, the upper 95% confidence limits for the variance components were used to estimate measurement error standard deviations. Hence, the estimated customer measurement error standard deviation was set to .048062 while the estimated stage 3 measurement error standard deviation went up to .012042. This scenario gave a result of approximately 3.30 potential customer complaints per million which is still a highly desirable result. Nonetheless, notice the relatively large increase over the original estimate of customer complaints. It is best to spend some time plugging in different specifications, measurement schemes, and measurement and process error estimates to derive the most benefit from such a spreadsheet.

3 Multiple Components

Up to now we have only considered one component at a time, namely our low control. To look at the kit in a holistic fashion, we will have to focus on the relationships between the low control and the other kit components. By studying the variation in the production of each reagent and combining this information with our model, one could pick out the critically important reagents which must be held within strict specifications. For example, our low control might be on target with small variance when measured with gold standards (stage 2), but when measured with the reagents produced for a lot of kits (stage 3), the low control might appear off-target or have excessive variability due to fluctuations in the mass-produced reagents. The final manufacturing specifications before shipment to the customer must take bias and inflated variability into consideration.

How do we account for the inflation of variability from the pooling together of all our kit components? We must modify our distributions using the Delta Method. Many mathematical statistics books, such as Casella and Berger (1990), carry a discussion of this method. For our case here, the Delta Method is based on the linearization of the Taylor Series expansion of some function g about $E[\mathbf{w}] = \boldsymbol{\mu}_W$. (Let W_i, for $i = 1, \ldots, k$ be a set of random variables. Then $\mathbf{w} = \{W_i\}$ is said to be a random k-vector. Furthermore, suppose that $E[W_i] = \mu_i < \infty$, for $i = 1, \ldots, k$. Let $\boldsymbol{\mu}_W = \{\mu_i\}$.) To the first order, for any random variable B that is a function of $\mathbf{w}$:

$$B = g(\mathbf{w}) \doteq g(\boldsymbol{\mu}_W) + \mathbf{f}'(\mathbf{w} - \boldsymbol{\mu}_W)$$

where $\mathbf{f}$ is the gradient vector:

$$\mathbf{f} = \left.\frac{\partial g(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\boldsymbol{\mu}_W}$$

Having expressed B as being approximately equal to a linear function of $\mathbf{w}$, we are now able to find the expectation and variance of the linear function:

$$\mu_B \doteq g(\boldsymbol{\mu}_W)$$
$$\sigma_B^2 \doteq \mathbf{f}' \text{ Var}[\mathbf{w}] \ \mathbf{f}$$

Using our immunoassay kit example, let's say we have k components. When the kth component is being measured, each of the other components will have recorded their own important

responses W_i where hopefully, $W_1 \sim N(\mu_1, \sigma_1^2), \ldots, W_{k-1} \sim N(\mu_{k-1}, \sigma_{k-1}^2)$. Usually, since they are measured before kit assembly, we expect the W's to be independent except for those components that may have more than one response associated with it. To make it clear that we are dealing with an observed response as measured with the other kit components we will define a new random variable Z. Based solely on an examination of the kit, it is reasonable to hypothesize that some or all of the W_i's will have an effect on Z defined as $B = g(\mathbf{w})$. Another way of thinking about this would be to think of the random variable B as representing the bias introduced by the other kit components being brought into the measurement process. Therefore, the observed response Z is now equal to $X + D + B$. Because B is normally distributed, the marginal distribution for Z with the $k-1$ other kit components would be $N(\mu_X + g(\boldsymbol{\mu}_W), \sigma_p^2 + \sigma_m^2 + \mathbf{f}' \text{ Var}[\mathbf{w}]\ \mathbf{f})$. Note that we assumed independence between X, D, and B. The conditional distribution of $Z \mid X, W_1, \ldots, W_{k-1}$ is $N(X + g(\mathbf{w}), \sigma_m^2)$. Notice that if the other kit components do not affect Z , then $g(\boldsymbol{\mu}_W) = 0$.

Future research should focus on developing an approach for determining the functional relationships between kit components i.e. the $g(\mathbf{w})$'s. A possible approach incorporating historical data would be to collect response readings for a kit, both pass or fail, at the final stage before it is shipped to the customer. Then, these data would be compared to the previous stage's response readings, both pass or fail, for the specific components that went into making that kit. Estimates of the $g(\mathbf{w})$'s could be obtained through multiple regression. It should be noted that, from a biological and/or chemical perspective, some of the relationships between component responses could readily be obtained.

Before leaving this section, it should also be noted that if the company personnel have an understanding of the product such that they have made it to this point, the probability calculations from section 2.5 and the spreadsheet would have to be redone in order to take advantage of the improvement in our knowledge of costs and complaints. In order to calculate P[product fails customer specs | passed last manufacturing stage], the general strategy would require conditioning on X and $\mathbf{w}$. A k-fold integration would also have to be done. Obviously, this is computationally intense and would be a good topic for a future paper.

4 Multiple Responses

We have now arrived at the most complex part of this paper. It is here that we bring together not only all the kit components, but all the responses measured on the components as well. This type of multivariate analysis again considers the kit as a holistic entity rather than setting specifcations using the traditional "one-at-a-time" approach. Obviously, this would require a serious commitment to product quality on the part of those who pursue this avenue. In this section, the framework for a basic approach will be outlined. As previously mentioned, more research needs to be conducted along this vein to look at feasability and implementation.

Recall that our assay kit is made up of k components and each component is measured for at least one, and possibly several, attributes. Therefore we will have a total of $p \geq k$ measurements. The measurement p-vector, with one or more Z_i's possibly being measured

for each kit component, will be defined as:

$$\mathbf{z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix}$$

Usually Z_i will be independent of Z_j. However, there will be exceptions when multiple measurements are taken on a single kit component. For example, suppose that binding affinity and concentration are measured on microparticles in the Antigen X example. Thus, we know our covariance structure for $\mathbf{z}$ will not be a diagonal matrix.

We will once again use the Delta Method to carry out a Taylor Series expansion of a vector-valued function $\mathbf{g}$ about $E[\mathbf{w}] = \boldsymbol{\mu}_W$. Let $\mathbf{b}$ be a random p-vector functionally related to $\mathbf{w}$ such that:

$$\mathbf{b} = \mathbf{g}(\mathbf{w}) = \begin{pmatrix} g_1(\mathbf{w}) \\ \vdots \\ g_p(\mathbf{w}) \end{pmatrix}.$$

Therefore, we have:

$$\mathbf{b} = \mathbf{g}(\mathbf{w}) \doteq \mathbf{g}(\boldsymbol{\mu}_W) + \boldsymbol{\Delta}'(\mathbf{w} - \boldsymbol{\mu}_W)$$

where $\boldsymbol{\Delta}$ is the gradient matrix:

$$\boldsymbol{\Delta} = \left.\frac{\partial \mathbf{g}'(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\boldsymbol{\mu}_W} : k \times p.$$

Continuing:

$$\boldsymbol{\mu}_\mathbf{b} \doteq \mathbf{g}(\boldsymbol{\mu}_W)$$
$$\boldsymbol{\sigma}^2_\mathbf{b} \doteq \boldsymbol{\Delta}' \text{ Var}[\mathbf{w}] \ \boldsymbol{\Delta} : p \times p.$$

Here, $\mathbf{z} = \mathbf{x} + \mathbf{d} + \mathbf{b}$ where $\mathbf{x}$ is the corresponding vector of unobserved means, $\mathbf{d}$ is the corresponding vector of measurement errors, and $\mathbf{b}$ is the corresponding vector of biases. Assuming independence between these normal random vectors, we find $\mathbf{z}$ is multivariate normal i.e. $\mathbf{z} \sim N_p(\boldsymbol{\mu}_X + \mathbf{g}(\boldsymbol{\mu}_W), \boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_m + \boldsymbol{\Delta}' \text{ Var}[\mathbf{w}] \ \boldsymbol{\Delta})$. $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_m$ are the covariance matrices for the process and measurement variances respectively. The conditional distribution of $\mathbf{z} \mid \mathbf{x}, \mathbf{w}$ is $N_p(\mathbf{x} + \mathbf{g}(\mathbf{w}), \boldsymbol{\Sigma}_m)$.

As was noted in the previous section, once the distributions for $\mathbf{z}$ and $\mathbf{z} \mid \mathbf{x}, \mathbf{w}$ have been estimated, this information would then be used to redo the calculations necessary to derive P[product fails customer specs | passed last manufacturing stage]. Since we are now dealing with multiple components and multiple responses as well, the calculations become very difficult and, as before, would make a good topic for a future paper. The author feels the return on this investment of time would be sizeable.

5 Conclusion

In closing, it is worth mentioning that a change from any familiar setting can often be discomforting to a company seeking to improve quality in its products. It is suggested that

any company seeking to implement this approach not do so lightly. This means hiring and/or consulting with personnel who have a well-rounded statistics background. To discourage the temptation to embrace "business as usual", let's briefly summarize the real benefits of this specification setting approach that have been discussed in this paper.

First, the approach outlined in this paper allows new reliable specification settings for each stage of production of each component in a product. Through a willingness to learn, the company will gain a clear understanding of the importance of each product component in meeting final customer specifications via improved models, both in terms of well-specified variance components and through the inclusion of functional relationships between product components. This holistic approach which considers all product components and their variance components to find problem areas is done up front. This minimizes the frustration of the traditional approach which is to ship the product to the customer, wait for "fires" to break out, and hope for the best outcome.

Second, we can now apply a more accurate statistical rigor based on integrating over conditional densities. A complete computer algorithm would receive input from the production team members and output estimates of costs and numbers of complaints which would then be used to set specifications, from an optimal standpoint. The algorithm could also be made applicable to any other products via modification.

Finally, because a long-term improvement is realized in the product, the company would save money by reducing both tangible company costs and intangible customer complaint costs.

Several areas have been mentioned as being topics for future work, such as incorporating multiple components and multiple responses into the probability calculations from section 2.5. There may be other interesting areas specific to the needs of the user. One such area is to account for the expected degradation of reagents over time and include the stability information in setting specifications. In closing, this paper is intended to represent a starting point from which future research topics are based.

6 Bibliography

1. American Statistical Association: "Ethical Guidelines for Statistical Practice", Issue #253, Amstat News, Alexandria, VA, 1998.

2. Burdick and Graybill: "*Confidence Intervals on Variance Components*", Marcel Dekker, NY, 1992.

3. Casella and Berger: "*Statistical Inference*", Duxbury Press, CA, 1990.

4. Milliken and Johnson: "*Analysis of Messy Data. Volume I: Designed Experiments*", Van Nostrand Reinhold, NY, 1992.

5. Montgomery: "*Introduction to Statistical Quality Control*", 3d ed., John Wiley & Sons, NY, 1997.

6. Montgomery and Runger: "Gauge Capability and Designed Experiments. Part I: Basic Methods", *Quality Engineering*, 6(1), 115-135 (1993-94).

7. Montgomery and Runger: "Gauge Capability and Designed Experiments. Part II: Experimental Design Models and Variance Component Estimation", *Quality Engineering*, 6(2), 289-305 (1993-94).

8. Wheeler and Lyday: "*Evaluating the Measurement Process*", 2d ed., SPC Press, Knoxville, TN, 1989.