Inspection Strategies for Multistage Production Systems with Time-Varying Quality

Michael H. Veatch
Department of Mathematics
Gordon College
Wenham, MA 01984
veatch@gordon.edu

April 1999

Abstract

An economic, cost of quality model is formulated for a production system with multiple inspection locations and sampling. The model is applied to a thermal printer for digital photographs to determine when inspection of incoming material is appropriate. A simple dynamic model of quality is used, allowing for common-cause defects and the evaluation of sampling plans. The opportunity to correct quality problems more quickly when they are detected earlier is also considered. The cost structure used allows the costs of quality improvement, not just defective units, to be considered.

1 Introduction

The design of sampling plans is a fundamental issue for quality programs. Classical (risk-based) and Bayesian (economic) methods are well developed for a single inspection location. However, the cost implications of sampling are less well understood in multistage inspection. Most multistage models have focused on screening inspection, where defects occur independently and costs are proportional to the number of defects detected at each stage. In this paper, a multistage sampling inspection model is formulated that emphasizes dependency between defects and a more realistic cost structure. The main objective of the paper is to attempt to model several phenomena that result from current practice in continuous quality improvement programs. We also demonstrate an application of the multistage model, mostly using data already being collected by the quality organizations, to determine which parts are appropriate for incoming sampling inspection.

The most important conceptual difference between this model and typical multistage inspection allocation models is the importance of common cause defects. For variable parameters, designing for quality leads to a situation where planned process variability essentially never causes defects. The remaining defects are largely due to other events, such as specification errors, shipping the wrong part, damaged material, or changes in the supplier's process. Such events generate common cause defects, not just independently occurring defects.

A second implication of quality improvement programs is that quality improvement becomes a major justification for inspection. In the short term, the costs of investigating and responding to supplier problems can outweigh the cost of replacing material. An analysis of these programs needs to consider the costs–mostly labor–of corrective action, as well as the potential reduction in defects. In particular, we evaluate the benefit of detecting problems sooner, shortening the feedback loop and reducing the time-average defect rate.

A third implication is that defect rates will tend to be small, particularly for piece-parts and subassemblies. Small defect rates make the interactions between different defects (different attributes in a part or different parts in a product) less significant. Hence, there is less need to model these interactions in a single large optimization problem.

The older literature on multistage inspection allocation is surveyed by Raz (1986).

These models find the set of inspection locations to minimize cost or maximize outgoing quality. Most assume independently occurring defects at a known constant defect rate. In this setting, Lindsay and Bishop (1964) and others show that if individual units, not lots, are inspected (and inspection costs are concave) then the optimal inspection level at each stage is either 100% or 0%. Because the role of the inspection is to screen out discrepant items, not detect and respond to quality problems, sampling inspection is not justified. One of the most general models of this kind is Raz and Kaspi (1991).

The design of acceptance sampling plans is surveyed in Menipaz (1978). An example of more recent work is Park, Peters, and Tang (1991). Both classical statistics (risk-based) and Bayesian prior distributions of the defect rate (economic, expected cost models) have been used (Baker et al. 1994). A major challenge for the Bayesian method is the selection of a prior distribution (Case and Yeats 1982).

Less attention has been paid to sampling in the context of multistage inspection. Beightler and Mitten (1964) introduce a very simple dynamic model of quality where the defect rate is randomly sampled from two values for each lot. In Woo and Metcalfe (1972) the defect rate starts at the smaller of two values, randomly changes to a larger value when the process goes out of control, then is reset in the next lot after the change is detected. Both find the optimal sample size. Dietrich and Saunders (1974) use a more conventional beta prior distribution of lot defect rates. None of these papers distinguish between different kinds of defects or attributes; a unit is either defective or conforming. Multiple attributes, each inspected at a different stage, are considered using beta priors in Tang, Plante, and Moskowitz (1986) and Moskowitz, Plante, and Tang (1986). Economic and risk-based methods are combined in Evans and Alexander (1987) and Baker et al. (1994).

In this paper, a simple dynamic model of quality is used. The defect rate changes periodically, randomly sampled from an empirical distribution. The variation in defect rate represents changes in the supplier process or other changes that persist over a period of time, such as specification errors. These common cause quality problems provide a basis for justifying sampling inspection. The average days of stock is used as the time between changes in defect rate. This choice of time interval allows the benefit of earlier detection to be measured by interrupting high-defect rate intervals when the problem is detected. When a lot is rejected at the first stage (incoming inspection), the defect rate is reset.

Combined with the defect rate model is a multistage inspection model. Many of its features were motivated by the case study described below; other features common in the literature were added for completeness. To capture the costs associated with quality improvement, a fixed cost is incurred when a quality problem is detected, in addition to the usual cost for each defective unit. Hence, costs are nonlinear in the defect rate and are sensitive to the variation in defect rate, i.e., sensitive to the "size" of common cause quality problems. Three types of inspection are considered: Sampling, screening, and screening with sorting of on-hand inventory when a threshold is crossed. Defects are divided into different types for the purpose of applying costs; they are also divided into different stages at which they are first detectable. The model is computationally tractable for a large number of stages, since each part number is analyzed separately.

The model was applied to a thermal printer for digital photographs. Management was seeking a more objective rationale for which parts should receive incoming inspection. Thus, inspection was treated as a given at all but the first stage. A multistage model was still important for assessing the cost of defects entering their manufacturing process. The analysis suggested a potential $50,000-100,000 annual cost savings on this product through reducing incoming inspection. Insights were also gained on what makes a part a good candidate for incoming inspection.

The next section describes the Cost of Quality (CoQ) model, developing equations for cost and escaping defect rates. Readers who are interested in the application may wish to skip to Section 3, which presents the thermal printer case study. The findings are summarized in Section 4

2 CoQ Model

This section presents a model for computing the quality-related cost of a given inspection policy. After defining some common elements in Section 2.1, three types of single-stage inspection are presented in Sections 2.2 - 2.4. Their use in a multistage model is explained in Section 2.5 and their use with a defect rate distribution in Section 2.6. Of the many options described, only lotted inspection, inspect and sort, replacement, and corrective action are used in numerical examples.

2.1 Single-Stage Model

As shown in Figure 1, parts enter an inspection stage and are either accepted–with or without having been inspected–or rejected. Rejected parts then either rejoin the accepted parts after some repair process or leave the system, i.e., they are scrapped or returned to the supplier. These scrapped/returned parts create a yield loss, requiring more production volume at upstream stages. This increased volume translates into more inspections and rejections and a larger CoQ, which is measured per outgoing part at the last stage.

Defects are divided into different *types* according to the action taken and associated cost. Each part is assumed to contain at most one defect, so that defects of different types are mutually exclusive events. An incoming part has a type i defect with probability σ_i, independent of other parts. The probability that a type i defect is detected when a part is inspected is

$$\lambda_i = (1 - \beta_i)\sigma_i, \tag{1}$$

where β_i is the type II error rate, i.e., the probability of accepting a part upon inspection given that it has a type i defect. We assume that there are no false defects, since they can be modeled as an additional defect type. Let z be the fraction of parts that are inspected. Then the probability that a type i defect is detected is

$$P(\text{reject for type } i \text{ defect}) = \lambda_i z. \tag{2}$$

Let ρ_i be the probability that an outgoing part, either accepted or repaired, contains a type i defect. Also let y be the *yield* of the inspection stage, i.e., the average ratio of outgoing parts to incoming parts. Then $1 - y$ is the replacement rate. Because defect types are mutually exclusive, $\rho = \sum_i \rho_i$ is the probability that an outgoing part contains a defect and $\lambda = \sum_i \lambda_i$ is the probability of detecting a defect when a part is inspected.

Three models of repair will be considered:

Perfect repair. After rejected parts are repaired, they are defect-free.

Imperfect repair. After rejected parts are repaired, they have the same total defect rate as parts that are inspected and accepted.

Replacement. Rejected parts are replaced. Replacement parts incur the same costs and have the same defect rates as original parts.

Under perfect or imperfect repair, $y = 1$.

2.2 Screening Inspection

Screening inspection treats parts individually. When a defect is found, only the defective part is rejected. To derive the outgoing defect rate, we further assume that repair of one defect type cannot induce a defect of another type. There are three sources outgoing defects: not inspecting a part, inspecting but missing a defect, or not repairing the defect. Regardless of the type of repair,

$$\begin{aligned}\rho_i &= \sigma_i P(\text{outgoing part was not inspected}) \\ &\quad + P(\text{outgoing part was inspected, accepted and has type } i \text{ defect}) \\ &\quad + P(\text{outgoing part had type } i \text{ repair}) P(\text{type } i \text{ defect after repairing type } i).\end{aligned} \tag{3}$$

Under perfect repair, (3) reduces to

$$\rho_i = \sigma_i(1 - z) + \beta_i \sigma_i z. \tag{4}$$

Under imperfect repair, repaired parts have the same defect rate as parts that pass inspection, but it is concentrated in one defect type. Specifically, we assume that

$$\begin{aligned}& P(\text{type } i \text{ defect after repairing type } i \text{ defect }) \\ = \; & P(\text{defect} \mid \text{part is inspected and accepted}) = \frac{\sum_i \beta_i \sigma_i}{1 - \lambda}.\end{aligned} \tag{5}$$

Using (5),

$$\rho_i = \sigma_i(1 - z) + \beta_i \sigma_i z + \lambda_i z \left(\frac{\sum_i \beta_i \sigma_i}{1 - \lambda} \right). \tag{6}$$

If we assume that β_i is constant for all defect types (6) reduces to

$$\rho_i = \sigma_i(1 - z) + \frac{\beta_i \sigma_i}{1 - \lambda} z. \tag{7}$$

Equation (7) also applies if we assume that repaired parts have the same defect rate *for each type of defect* as parts that pass inspection. For these reasons, we use it in our calculations. If $\lambda = 1$, the last term in (7) becomes zero.

When replacement is used instead of repair, only incoming parts that are accepted become outgoing parts. Yield is just the probability of acceptance,

$$y = 1 - \lambda z. \tag{8}$$

The number of outgoing defects is the same as for perfect repair, but the expected number of outgoing parts is smaller by a factor of y; therefore, ρ_i is larger by a factor of $1/y$:

$$\rho_i = \frac{\sigma_i(1-z) + \beta_i \sigma_i z}{y}. \tag{9}$$

If $y = 0$, arbitrarily set $\rho_i = 0$.

The inspection stage incurs a cost h for each part inspected and c_i for each type i defect detected. A variety of costs may be included in c_i, such as the cost of scrapping, repairing, or returning the part; the cost of repairing an assembled unit containing the part; and the cost of investigating and correcting the problem. The expected cost per incoming part is

$$V = hz + \sum_i c_i \lambda_i z. \tag{10}$$

2.3 Lotted Inspection

Now we consider lotted inspection, where parts arrive in lots of size L. A sample of s parts from a lot is inspected and the entire lot accepted or rejected. We assume that a lot is rejected if one defect is found, since sampling inspection is often tied to supplier quality improvement programs that react to even one defect.

The probability that an inspected lot is accepted is

$$p_A = (1-\lambda)^s \tag{11}$$

and the probability it is rejected is $p_R = 1 - p_A$. Interpreting z as the probability that a lot is inspected, the probability that a lot is (inspected and) rejected is $p_R(z) = p_R z$ and the probability that a lot is accepted is

$$p_A(z) = 1 - (1-p_A)z. \tag{12}$$

Equation (3) applies with minor modifications:

$$\begin{aligned} \rho_i \quad = \quad & \sigma_i P(\text{outgoing part was not in an inspected lot}) \\ & + P(\text{outgoing part was in an inspected, accepted lot and has type } i \text{ defect}) \\ & + P(\text{outgoing part had type } i \text{ repair}) P(\text{type } i \text{ defect after repairing type } i). \end{aligned} \tag{13}$$

First we consider the number of type i defects in an inspected lot, X_i . Let X_i^u and X_i^t be the number of defects in the tested and untested portion of the lot, respectively, and A

and R be the events that the lot is accepted or rejected. The expected number of type i defects in an inspected accepted lot is

$$\begin{aligned} E(X_i|A) &= E(X_i^u|A) + E(X_i^t|A) \\ &= (L-s)\sigma_i + sP(\text{part has type } i \text{ defect}|A) \\ &= (L-s)\sigma_i + sP(\text{part has type } i \text{ defect}|\text{part is tested and accepted}) \\ &= (L-s)\sigma_i + s\left(\frac{\beta_i\sigma_i}{1-\lambda}\right). \end{aligned} \tag{14}$$

If $\lambda = 1$, the quotient becomes zero in (14) and below. To consider rejected lots, we sum over defect types, letting $X = \sum_i X_i$, and use the conditional expectation formula

$$E(X) = E(X|A)p_A + E(X|R)p_R, \tag{15}$$

or

$$\begin{aligned} E(X|R) &= \frac{E(X) - E(X|A)p_A}{p_R} \\ &= \frac{L\sigma - E(X|A)p_A}{p_R}, \end{aligned} \tag{16}$$

where $E(X|A) = \sum_i E(X_i|A)$ from (14). If $p_R = 0$, arbitrarily set $E(X|R) = 0$.

Under perfect repair, (13) and (14) combine to give

$$\begin{aligned} \rho_i &= \sigma_i(1-z) + \frac{E(X_i|A)}{L}p_A z \qquad (17) \\ &= \sigma_i(1-z) + \frac{1}{L}\left[(L-s)\sigma_i + s\left(\frac{\beta_i\sigma_i}{1-\lambda}\right)\right]p_A z. \end{aligned}$$

Under imperfect repair, we assume that all parts in a rejected lot are inspected, those that fail inspection are repaired, and repaired parts have the same defect rate as parts that pass inspection. As argued in Section 2.2, keeping track of the defect type before repair makes little difference, so we assume

$$\begin{aligned} P(\text{type } i \text{ defect after repair}) &= P(\text{type } i \text{ defect}|\text{part is inspected and accepted}) \\ &= \frac{\beta_i\sigma_i}{1-\lambda}. \end{aligned} \tag{18}$$

From (13), (14), and (18),

$$\rho_i = \sigma_i(1-z) + \frac{1}{L}\left[(L-s)\sigma_i + s\left(\frac{\beta_i\sigma_i}{1-\lambda}\right)\right]p_A z + \left(\frac{\beta_i\sigma_i}{1-\lambda}\right)p_R z. \tag{19}$$

Under replacement, entire lots are replaced. The yield is

$$y = p_A(z). \tag{20}$$

As argued in Section 2.2, ρ_i is larger by a factor of $1/y$ than for perfect repair:

$$\rho_i = \frac{\sigma_i(1-z)}{y} + \frac{1}{L}\left[(L-s)\sigma_i + s\left(\frac{\beta_i\sigma_i}{1-\lambda}\right)\right]\frac{p_A z}{y}. \tag{21}$$

If $y = 0$, arbitrarily set $\rho_i = 0$.

For lotted inspection, h is the cost of inspecting a *lot*. Rejecting a lot for a type i defect costs l_i plus c_i for each type i defect it contains. A rejected lot is assigned defect type i (for cost calculation) with probability

$$P_i = \lambda_i/\lambda. \tag{22}$$

Since lots may contain more than one defect type, (22) is an approximation. Also, if $\lambda = 0$, set P_i arbitrarily. We apply the cost c_i to *all* defective parts in a rejected lot. This in effect assumes perfect inspection when a lot is sorted, which differs from the assumptions used to compute ρ_i for imperfect repair. If the lot is scrapped rather than sorted, then $c_i = 0$ and scrapping cost is included in l_i. Again using (22) to assign defect type, the probability that a part incurs the cost c_i is approximated as

$$\frac{p_R(z)P_iE(X|R)}{L}. \tag{23}$$

Another approximation is made in the cost equation to reconcile lotted inspection costs with the inspect and sort costs described in Section 2.4. If an inventory is bigger than a lot, more than one lot may be defective in one inventory. However, since much of the cost in l_i is for identifying and resolving the underlying problem, this cost should only occur once, whether one inventory or several lots are rejected. Thus, we allow at most one defective lot cost per inventory. The number of lots per inventory is K/L, where L is the lot size and K is the inventory level. The probability that an inventory incurs the defective lot cost is

$$P(\text{reject at least one lot}) = 1 - [p_A(z)]^{K/L}. \tag{24}$$

The expected cost per incoming part is

$$V = \frac{1}{L}\left[hz + p_R(z)E(X|R)\sum_i P_ic_i\right] + \frac{1-[p_A(z)]^{K/L}}{K}\sum_i P_il_i. \tag{25}$$

2.4 Inspect and Sort

A third type of inspection, called inspect and sort, addresses the practice of sorting the entire on-hand inventory of parts when q defects are found within some time window. The inventory level varies over time, but is generally much larger than the lot size. The sort might involve inspecting each part or further investigation and sampling. We model this practice by assigning a cost to a sort that is proportional to K, the average inventory level. (This parameter is also used in Section 2.5 to measure the corrective action delay.) We assume that no part will be sorted more than once, based on the premise that all parts which pass the sort pass the inspection, so they do not trigger another sort. The sort error rate is the same as the inspection error rate

The probability of triggering a sort when inspecting a part depends on the number of parts inspected since the last sort and the time window. We use the following approximation. There will be Kz parts, rounded up to an integer if necessary, inspected out of a set of K parts. The probability of detecting at least q defects is

$$p_R(z) = 1 - \sum_{x=0}^{q-1} \binom{Kz}{x} \lambda^x (1-\lambda)^{K-x}. \tag{26}$$

Since a part is sorted at most once, we use $p_R(z)/K$ as the probability that a part triggers a sort. Treating sets of K parts as independent batches approximates a time window of K parts. Outgoing defect rate and yield are the same as screening inspection (4) to (9), and P_i from (22) will be used.

For inspect and sort, h is the cost of inspecting a part, c_i is the cost of each detected type i defect, and l_i is the additional cost of a sort attributed to a type i defect. The expected cost per incoming part is

$$V = hz + \frac{p_R(z)}{K} \sum_i P_i l_i + \sum_i c_i \lambda_i z. \tag{27}$$

2.5 Multistage Model and Corrective Action

The multistage model combines any sequence of inspection stages. The subscript $n = 1, \ldots, N$ will be added to previous notation to denote the stage. At stage n, a type i defect rate of δ_{in} is introduced. As in Chevalier and Wein (1996), these defects are first observable at state n and are assumed to be observable at all later stages. The incoming defects observable at stage n consist of defects not detected at stage $n-1$ and defects first detectable at stage n. Again, we assume that a part contains at most one defect, so these categories are mutually exclusive and

$$\sigma_{in} = \delta_{in} + \rho_{in-1}, \tag{28}$$

with $\rho_{i0} = 0$.

In addition to the sorting process, we also allow lotted inspection at the first stage to influence the incoming defect rate. This feature, called *corrective action,* is intended to model the more rapid correction of a supplier quality problem when it is detected at incoming inspection. Assume that a defect rate distribution is being used, as described in Section 2.6, and the inventory level (introduced in Section 2.4) satisfies $K > L$. Without corrective action, the defect rate remains constant for K parts. With corrective action, if a lot is rejected in stage 1 a new defect rate is sampled. Since most rejected lots have high defect rates, this "resetting" will lower the expected defect rate. Let y_0 be the expected fraction of an inventory until a lot is rejected at stage 1 and N_R be the number of lots until a rejected lot (at a fixed δ). Now, N_R has a geometric distribution with mean $1/p_R(z)$. The number of lots before resetting the defect rate is $\min\{N_R, K/L\}$. Hence,

$$\begin{aligned} y_0 &= \frac{E[\min\{N_R, K/L\}]}{K/L} \\ &= E[\min\{N_R L/K, 1\}] \\ &\approx \min\{[1/p_R(z)]L/K, 1\}. \end{aligned} \tag{29}$$

If $p_R(z) = 0$, then $y_0 = 1$. We will refer to corrective action as stage 0 because y_0 can be interpreted as the incoming yield, or expected fraction of the inventory received. Without corrective action $y_0 = 1$.

Multistage cost must take into account the yield loss across stages. Define the cumulative yield of stages 0 through n as

$$y_{0,n} = \prod_{j=0}^{n} y_j, \tag{30}$$

where y_n is the yield at stage $n = 1, \ldots, N$. In addition to the cost at each stage, each outgoing type i defect at the last stage incurs an escaping, or field, defect cost f_i. The expected cost per part entering stage 0 is

$$V = \sum_{n=1}^{N} y_{0,n-1} V_n + y_{0,N} \sum_i f_i \rho_{iN}. \tag{31}$$

2.6 Use of Defect Rate Distribution

In the previous sections the defect rate has been fixed. Now assume that the total defect rate, defined as $\delta = \sum_{n=1}^{N} \sum_i \delta_{in}$, has a distribution. Each inventory of parts has a defect rate, independent of other inventories, drawn from this distribution (actually, how often the defect rate changes makes little difference except for the corrective action option). Let $f(\delta)$ be the discrete probability density function of δ. Our approach is, essentially, to take the expectation of cost over δ. However, we must construct a joint distribution of δ_{in} . To avoid additional data requirements, we proceed as follows. Let $\delta_n = \sum_i \delta_{in}$, the defect rate first detectable at stage n. Assume that

$$f(\delta, n) \triangleq P(\delta_n = \delta, \delta_j = 0, j \neq n) = \gamma_n f(\delta), \tag{32}$$

where γ_n is the fraction of defects first detectable at stage n. In other words, all the defects in a given inventory are first detectable at the same stage. However, defects are divided into types in fixed proportions:

$$\delta_{in} = \tau_{in} \delta_n \tag{33}$$

with probability one, where τ_{in} is the fraction of defects that are type i of those first detectable at stage n.

Quantities computed in the previous sections now depend on δ and the stage first detectable, which we refer to by the index m. Expected yield over stages 0 through n is

$$\overline{y}_{0,n} = \sum_{\delta} \sum_{m=1}^{N} f(\delta, m) y_{0,n}. \tag{34}$$

In particular, the expected incoming "yield" due to corrective action is $\overline{y}_0 = \overline{y}_{0,0}$. The expected cost per part leaving the last stage is

$$\overline{V} = \left(\sum_{\delta} \sum_{m=1}^{N} f(\delta, m) V \right) \Big/ \, \overline{y}_{0,N} \tag{35}$$

and the outgoing defect rate at the last stage is

$$\overline{\rho_N} = \left(\sum_{\delta} \sum_{m=1}^{N} f(\delta, m) y_{0,N} \rho_N \right) \Big/ \, \overline{y}_{0,N}. \tag{36}$$

The incoming defect rate at the first stage (after corrective action) is

$$\overline{\delta} = \left(\sum_{\delta} \sum_{m=1}^{N} f(\delta, m) y_0 \delta \right) \Big/ \, \overline{y}_0. \tag{37}$$

Note that without corrective action (37) is the expectation of the given defect rate distribution. Similar equations can be written for other measures of performance.

The model computations can be summarized as follows.

For each δ, m, and n, compute

$\delta_{i,n}$ for each i from (33),

$\rho_{i,n}$ for each i from (4), (7), (9), (17), (19), or (21),

y_n from (8) or (20), and

V_n from (10), (25), or (27).

For each δ and m, compute

y_0 from (29),

V from (31), and

$f(\delta, m)$ from (32).

Finally, compute (34)-(37).

2.7 Optimal Inspection Policies

The CoQ model can be used to decide which of the possible inspection stages should be used for inspection. The optimization problem is to minimize $\overline{V}$ over $0 \leq z_n \leq 1$, where z_n is the fraction of parts or lots inspected at stage n. For many inspection problems it is known that the extreme points (0% or 100% inspection) are optimal (Raz 1986). In our model nonlinearity in z_n occurs in:

1. Replacement: outgoing defect rate (9) and (21) and division by $\overline{y}$ in cost (35).

2. Lotted inspection with more than one lot per inventory ($K > L$): cost (25) and corrective action (29).

3. Inspect and sort on more than one defect ($q > 1$): probability of sorting an inventory (26).

As a result, $\overline{V}$ might be minimized by inspecting only some of the parts or lots. However, given the other limitations of the model, we will only use it to compare the extremes of 0% or 100% inspection. There are then 2^N combinations of inspection locations.

For a single screening inspection stage it is instructive to express the inspection decision analytically. Assuming perfect repair, $\overline{y} = 1$ and cost is a linear function of δ. Hence, expected cost is just cost evaluated at the mean defect rate $\overline{\delta}$. The cost with inspection is

$$h + \sum_i c_i \lambda_i + \sum_i f_i \beta_i \sigma_i = h + \left[\sum_i c_i (1 - \beta_i) \tau_i + \sum_i f_i \beta_i \tau_i \right] \overline{\delta} \tag{38}$$

and without inspection is $\overline{\delta} \sum_i f_i \tau_i$. It is optimal to inspect when

$$h < \overline{\delta} \sum_i (f_i - c_i)(1 - \beta_i) \tau_i. \tag{39}$$

For (39) to hold, escaping defects must be significantly more costly than detected defects, inspection costs must be small, and the detected defect rate must be large.

Other aspects of the inspection policy could also be selected using this model. For example, a cost-minimizing sample size could be found by a search, or lotted inspection could be compared with screening.

3 Case Study

The model was used to analyze CoQ of a thermal printer for digital photographs. The thermal printer is a medium to low volume operation consisting of manual assembly of piece-parts and subassemblies, followed by extensive calibration and testing. All of the 350 externally-supplied parts in the printer were analyzed. Most of the model inputs were extracted from the manufacturer's existing quality databases, much of it at the part number level. Some exceptions are noted below.

The inspection process is modeled in four stages (Figure 2). Stage 1 is the sampling inspection of incoming material. Stages 2 and 3 represent the assembly process. The many stations where defects are reported were aggregated into two stages on the basis of cost: Stage 3 requires repair of the assembled unit, usually including calibration and functional testing, while stage 2 does not. Typically, stage 2 defects are visual defects found by the operator before installing a part. Stage 4 is a 25% quality assurance audit that essentially repeats the functional test of stage 3. Significant inventory is held between stages 1 and 2, but very little inventory is held on the shop floor, where they use demand flow technology. When several defective units are found on the line an investigation is conducted that usually leads to sorting inventory. Accordingly, stages 2-4 are modeled as inspect and sort. There are a variety of dispositions for rejected material; the most common is to return the entire lot or the defectives found in sorting the inventory to the supplier. As a baseline, the replacement and corrective action options were used.

Cost and inspection data are given in Table 1 for the top problem part in this product, an LED display panel. First, the table lists the fraction of defects first detectable (γ), type II error rate (β), inspection rate (z), and inspection cost (h) for each stage. Error rates, γ_1, and γ_3 could not be assessed from existing data; values were selected after discussions with their quality engineers. Inspection prior to assembly (stage 2) generally checks the same parameters as incoming inspection (stage 1), so no defects are first detectable at stage 2. Similarly, no defects are first detectable at stage 4 because it duplicates stage 3. Inspection at stages 2-4 is taken as a given, with 25% inspection at stage 4, and no effort is made to assign inspection costs to these stages. Only stage 1 is optimized. Inspection cost at stage 1 is based on a typical inspection time for all parts and the sample size for the specific part number.

The bottom of the table lists the fraction of defects of each type (τ), the cost per defective part (c), the cost per rejected lot or sorted inventory (l), and the cost per escaping defect (f). The cost of a service call is used for f. All of the possible material dispositions and actions taken are collected into three defect types: scrap, return, and use as is. Labor cost estimates were available for all of these activities except for repair of assembled units, which was assumed to be $500 (see footnote to Table 1). The cost of scrapping a part is assumed to be the full purchase price, even if the supplier is determined to be responsible for the defect. Costs incurred by the supplier could be discounted by separating type 1 into supplier responsible and manufacturer responsible defects. Defect type 3 illustrates the magnitude of investigation costs since it does not include any material costs. It should be noted that when replacement is used, the model does not handle defect type 3 in a precise manner because it includes this "use as is" material in the yield loss.

The defect rate distribution is shown in Table 2. This distribution was constructed from two years of data using the empirical Bayes technique described in Veatch (1998).

Table 1: Display panel inputs

Stage	γ	β	z	h (\$)
1	.50	.10	opt.	65
2	0	.10	1	0
3	.50	.01	1	0
4	0	.01	.25	0

	Stage 1			Stage 2-4			Escaping
Defect type	τ	c (\$)	l (\$)	τ	c (\$)	l (\$)	f (\$)
1. Scrap	.56	34.78	1008	.69	34.78	3828[1]	3800
2. Return	.19	17.39	1008	.23	17.39	3828[1]	3800
3. Use as is	.25	0	322	.08	0	481	0

Lot size $L = 119$, sample size $s = 13$, inventory level $K = 845$, rejection quantity $q = 3$.

[1] Add \$500 for stages 3 and 4.

Table 2: Defect rate distribution (mean = .076)

Defect rate	Probability
0	.664
.017	.042
.030	.046
.045	.042
.061	.042
.083	.042
.118	.042
.428	.034
1	.046

The shape of the distribution is based on all electrical parts in the thermal printer, with the mean adjusted to reflect the data for the display panel. Note that this is the distribution of the defect rate, not the number of defects in a lot. Using this distribution introduces much more variability in the number of defects in a lot than if the defect rate were fixed at its mean. The source of this excess variability is quality problems that affect a large number of units.

The CoQ results for the display panel are broken down by stage in Table 3. Costs are per outgoing part at the last stage and are computed from (35) and similar equations for $\overline{V}_n$. The rejection rate is the probability that a part or lot entering a stage is rejected, computed as $1 - \overline{y}_{0,n}/\overline{y}_{0,n-1}$ for stage n. The undetected defect rate only considers those

Table 3: CoQ results by stage

	Inspect			Don't inspect		
Stage	Rejection rate	Undetected defect rate	Cost ($)	Rejection rate	Undetected defect rate	Cost ($)
1	.0267	.0011	0.80	0	.0378	0
2	.0010	.0001	0.16	.0340	.0039	1.48
3	.0435	.0005	1.91	.0426	.0004	2.16
4	.0001	.00034	0.09	.0001	.00035	0.09
Out			0.99			0.96
Total[1]			3.93			4.69

[1]Costs may not add because of roundoff.

defects that could have been detected at that stage, computed from (36) and similar equations for other stages. Incoming inspection reduces CoQ from $4.69 to $3.93 (the unit cost of the display panel is $34.78) and reduces the escaping defect rate from .00035 to .00034. Thus, the model recommends inspection, as would be expected for the top problem part. The effectiveness of the sampling inspection at stage 1 can be seen in the reduction of the defect rate from the incoming rate of .0378 to an undetected rate out of stage 1 of .0011. Much of this reduction is the impact of corrective action, which reduces the incoming defect rate detectable at stage 1 from .0378 to .0072. Because 50% of defects are not visible at stage 1, corrective action eliminates less than half of all defects, reducing the total defect rate from .076 to .050.

The desirability of incoming inspection depends on a number of factors. First, the cost of defects must increase when they are detected further downstream. Table 1 shows a fourfold increase in l; however, at stage 1 this is the cost per rejected lot while at stages 2-4 it is the cost per sorted inventory. An inventory consists of more than seven lots, all of which might be rejected instead of performing one inventory sort. In this situation the assumption that only one rejected lot cost can be incurred per inventory plays a key role in making downstream defects more costly. When the inventory size is set equal to the lot size of 119, incoming inspection is preferred ($10.61 vs. $11.74) without relying on this assumption. In this case the cost of detecting a defect downstream is much greater, but there is no opportunity for corrective action at stage 1 because there is only one lot per inventory.

Figure 3 shows the sensitivity of CoQ to mean defect rate. The data labelled "distribution" were generated using the defect rate distribution of Table 2 and multiplying the probability of each (nonzero) defect rate by a scale factor, thus preserving the relative frequency of "small" and "large" quality problems. Incoming inspection is desirable when the mean defect rate is above .025. Comparison with the data labelled "fixed defect rate" shows that the results are sensitive to variability in the defect rate. With a constant defect rate, incoming inspection is never desirable. This is consistent with the well-known result that either 0% or 100% inspection, not sampling, is optimal for independent defects at a known, constant rate. Figure 3 also shows the nonlinearity of CoQ to defect rate, which is a result of the sampling and inspect and sort logic.

To analyze all 350 part numbers, source data that varies by part (unit cost, lot size, inventory level, and mean defect rate) was extracted from several data systems and used to prepare a model input file. Ten of the 350 parts are recommended for incoming inspection by the model. Inspecting just these 10 parts (optimal inspection) gives a CoQ of $80.23 per unit produced, compared with $86.85 if no parts are inspected and $177.32 if all part numbers are inspected. Most of the parts recommended for inspection have high unit cost; the three exceptions are mechanical parts around $1-$2 but with very large lot sizes and relatively large defect rates. Mechanical parts, as a whole, have about three times the defect rate of electrical parts, making them easier to qualify for inspection.

The model has primarily been used to assess inspection at a given sample size. However, it is interesting to observe the sensitivity to sample size (Figure 4). A sample size of 8 gives a slightly lower overall cost than the standard sampling plan, with a sample size of 13, that would normally be used for this part. Standard sampling plans are based on the probability of detecting specified defect rates while this model is based on CoQ, so it is not surprising that the sample sizes differ. One reason for a small sample size in the model is variation of the defect rate. The high variability in Table 2 suggests that most of the lots which should be rejected have a fairly large percent defective, which can be detected with a smaller sample size.

Sensitivity to some of the model assumptions was also checked. The corrective action feature was found to be critical. Without corrective action, none of the 350 parts are recommended for inspection.

4 Conclusion

A model of quality-related costs has been presented that allows various repair options, including sampling, to be meaningfully compared in a multistage production process. Although the data requirements of the model are fairly extensive, a case study demonstrated that the essential data could be obtained from a quality management data system. Such a model could interface with quality and product configuration data systems to generate summary reports of CoQ and inspection recommendations. Material quality managers could then compare these reports to current inspection practice to prepare more cost-effective inspection plans. The model appears to be applicable to a wide range of assembly processes. In fact, collecting quality data on more parts would increase the accuracy of the estimation techniques used.

For the question addressed in the case study of whether to perform incoming sampling inspection, it was found that inspection should be rare–only 10 of 350 part numbers were recommended for inspection. Inspection is cost effective only for parts that have a poor quality record or a very high unit cost. Sampling inspection was also shown to only be cost effective when there is significant variation in the defect rate between lots and when inspection enables faster corrective action. Both of these phenomena presuppose that a rejected lot can be replaced by a lot drawn from a smaller defect rate. Thus, the dynamic model of quality was crucial to the analysis.

Acknowledgments

The author thanks Peter Castro for his helpful suggestions. Simon Miner and Ken Hartlaub wrote the software used in this study and Jonathan Paige assisted with data collection and analysis. This work was supported by a Leaders for Manufacturing grant.

References

Baker, J. R., Clayton, E. R., Matheson, L. A., and Rakes, T. R., 1994, A multiple objective approach for constructing attribute-sampling plans for non-serial assembly items. *International Journal of Production Research,* **32**, 1599-1613.

Beightler, C. C., and Mitten, L. G., 1964, Design of an optimal sequence of interrelated sampling plans. *Journal of the American Statistical Association,* **59**, 96-104.

Case, K. E., and Keats, B. J., 1982, On the selection of a prior distribution in Bayesian acceptance sampling. *Journal of Quality Technology,* **14**, 10-18.

Chevalier, P. B., and Wein, L. M., 1996, Inspection for circuit board assembly. Working paper, Operations Research Center, MIT, Cambridge, MA.

Dietrich, D. L., and Saunders, J. L., 1974, A Bayesian quality assurance model for a multistage production process. *ASQC Technical Conference Transactions,* Boston, MA, 338-348.

Evans, G. W., and Alexander, S. M., 1987, Multiobjective decision analysis for acceptance sampling plans. *IIE Transactions,* **19**, 308-316.

Moskowitz, H., Plante, R., and Tang, K., 1986, Multistage multiattribute acceptance sampling in serial production systems. *IIE Transactions*, **18**, 130-137.

Raz, T., 1986, A survey of models for allocating inspection effort in multistage production systems. *Journal of Quality Technology*, **18**, 239-247.

Raz, T., and Kaspi, M., (1991) Location and sequencing of imperfect inspection operations in serial multi-stage production systems. *International Journal of Production Research*, **29**, 1645-1659.

Tang, K., Plante, R., and Moskowitz, H., 1986, Multiattribute Bayesian acceptance sampling plans under nondestructive inspection. *Management Science*, **32**, 739-750.

Veatch, M. H., 1996, Inspection strategies to reduce the cost of quality. Leaders for Manufacturing technical report, Massachusetts Institute of Technology, Cambridge.

Woo, W. K., and Metcalfe, T. E., 1981, Optimal allocation of inspection effort in multistage manufacturing processes. *Western Electric Engineer*, **16**, 8-16.