Center for Quality and Productivity Improvement
University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53705
(608) 263-2520
(608) 263-1425 FAX
quality@engr.wisc.edu

Report No. 159

**Five Ways Statistical Tolerancing Can Fail,
And What To Do About Them**

Spencer Graves and Søren Bisgaard

September 1997

The Center for Quality and Productivity Improvement cares about your reactions to our reports. Please direct comments (general or specific) to: Reports Editor, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, WI 53705; (608) 263-2520. All comments will be forwarded to the author(s).

Five Ways Statistical Tolerancing Can Fail, And What To Do About Them

Spencer Graves

Center for Quality and
Productivity Improvement

University of Wisconsin-Madison, USA

Søren Bisgaard

Director, Center for Quality and
Productivity Improvement and
Associate Professor,
Department of Industrial Engineering

University of Wisconsin-Madison, USA

ABSTRACT

The traditional formula for statistical tolerance stack-up has repeatedly been criticized in the literature because, in some cases, the actual defect rate has exceeded that predicted by theory. Since the logic of the mathematics is correct, the problems must come from violations of assumptions. An awareness of those causes can help a user avoid difficulties that often accompany applications of statistical tolerancing. In this article we explore the general non-robustness of traditional root sum of squares statistical tolerancing and describe, in particular, five ways it can fail. These are [1] deficiencies in the functional model, [2] lower process capability in inputs than what is desired of outputs, [3] biases, [4] correlations, and [5] non-normality. We also show that statistical tolerancing is extremely non-robust to the first four types of causes. Moreover we provide examples of each type and discuss what to do about each.

Five Ways Statistical Tolerancing Can Fail, And What To Do About Them

Spencer Graves and Søren Bisgaard

The traditional formula for statistical tolerance stack-up has repeatedly been criticized in the literature because, in some cases, the actual defect rate has exceeded that predicted by theory. Since the logic of the mathematics is correct, the problems must come from violations of assumptions. An awareness of those causes can help a user avoid difficulties that often accompany applications of statistical tolerancing. In this article we explore the general non-robustness of traditional root sum of squares statistical tolerancing and describe, in particular, five ways it can fail. These are [1] deficiencies in the functional model, [2] lower process capability in inputs than what is desired of outputs, [3] biases, [4] correlations, and [5] non-normality. We also show that statistical tolerancing is extremely non-robust to the first four types of causes. Moreover we provide examples of each type and discuss what to do about each.

Introduction

The introduction of go/no-go gauges that operationalized lower and upper specification limits, L and U , was a major conceptual breakthrough in modern industrial mass production (Shewhart (1939, pp. 2-3)). This practice eventually permitted the development of economical manufacturing with interchangeable parts, as has lucidly been documented by Hounshell (1984). Early applications of the concept of specification limits appear to have been based more or less on intuition. By the early 1930s, however, tolerancing practice was moving towards more formal analysis, as exemplified by Rüdenberg (1929) and Shewhart (1931). For root sum of squares statistical tolerancing, it is assumed that an output quality characteristic, X_a , can be adequately approximated by a linear function of input quality characteristics, X_i 's, $i = 1, \dots, n$, as

$$X_a \equiv \theta_a + \sum a_i (X_i - \theta_i) \quad (1)$$

at least when $L_i \leq X_i \leq U_i$, where θ_a is the nominal for the assembly with X_a as its actual value, θ_i is the nominal for X_i , Σ denotes the sum over $i = 1, \dots, n$, and (L_i, U_i) denote the lower and upper specification limits, respectively. If X_a is a known and differentiable function of the X_i 's, $X_a = F(X_1, \dots, X_n)$, then (1) may be a Taylor approximation with

$$b_i = \left[\frac{\partial F}{\partial X_i} \right]_{X_1=\theta_1, \dots, X_n=\theta_n}$$

Suppose the mean of X_i is θ_i , the nominal. Subtract θ_a from both sides of (1), square both sides and take expectation. If X_i and X_j are uncorrelated, $i \neq j$, this produces the error transmission formula

$$\sigma_a = \sqrt{\sum b_i^2 \sigma_i^2} \quad (2)$$

where $\sigma_a^2 = \text{Var}\{X_a\}$ and $\sigma_i^2 = \text{Var}\{X_i\}$. With the further assumption that tolerance intervals T all are the same fixed multiple of the standard deviations for assembly as for components, we get the statistical tolerancing formula where the assembly's tolerance T_a is computed from

$$T_a = \sqrt{\sum b_i^2 T_i^2} \quad (3)$$

where $T_i = U_i - L_i$ is the width of the tolerance intervals for component i , $i = 1, \dots, n$, θ_i is its midpoint, and T_a is the tolerance for the assembly. (In this article, (3) is described as "traditional statistical tolerancing.") Historically, this formula goes back at least to Rüdenberg (1929), but it is also featured by Brandenberger (1946, pp. 72-77), Grant (1946, p. 326), Bruyevich (1946), Juran (1951, pp. 75-77; 1962, pp. 3-42/50), and others more recently.

As indicated above, the primary focus of our discussion in the remainder of this article, is based on the following five assumptions:

1. The linear functional relationship, (1), is exact.
2. All tolerances represent the same number of standard deviations.
3. The component distributions are all unbiased.
4. All component characteristics are uncorrelated.
5. All component characteristics are normally distributed.

Expression (3) can be proven using the first four of these assumptions. The fifth, normal distributions, is commonly used to estimate defect rates and costs.

Of course, in practice none of these assumptions will be exactly true. The important question therefore is how robust (or sensitive) expression (3) is to violations of these assumptions. This issue has previously been discussed in a cursory way by Juran and Gryna (1980, pp. 305-306) who mentioned the last three of these five assumptions, but not the first two, and did not explore in detail the consequences of violations of the assumptions.

In this article, we will explore in more detail the problems with (3) created by violations of each of these five assumptions as well as how possibly to avoid them. The problems sometimes include difficulties in producing parts from inferior materials, but may also involve extra work required to assemble discrepant components. In other cases, tolerancing problems can reduce the functionality or lifetime of a product.

We know of no careful study of tolerancing problems in industry. However, manufacturing and warranty problems are common, and some of these problems may be caused by inappropriate tolerances. When statistical tolerancing has been used, it is often easy for engineers to conclude that statistical tolerancing has contributed to the problems. The discussion in this article shows that there is some justification for this conclusion. In particular, the present article describes several seemingly modest violations of assumptions that can lead to substantial problems when statistical tolerancing has been used. This suggests that statistical tolerancing may be extremely non-robust to relatively common violations of the first four of the five listed assumptions. We hope this discussion will help engineers and others use statistical tolerancing confidently where it can reliably improve quality while reducing costs.

This article is organized as follows. The subsections of the next section provide examples of each of the five ways statistical tolerancing can fail. This is followed by a brief discussion of what can be done to reconcile statistical tolerancing and manufacturing reality.

Five Ways Statistical Tolerancing Can Fail

As previously indicated, statistical tolerancing does not fail because of some logical flaw in the formula. Rather, it fails because the idealized assumptions made in the derivation are often not met in practice and the usefulness of (3) critically depends on four of these five assumptions. Thus, in the following five subsections, we will describe problems arising from violations of each of the five assumptions.

Failure Mode 1. Inadequate Functional Model.

If the functional model, (1), is not accurate, tolerances based on it can lead to difficulties. Perhaps the most likely and serious cause of problems of this kind is the omission of important terms from (1). For example, Taguchi (1986, p. 75) suggested the inclusion in (1) of "external noise factors" such as environmental changes. The omission of such terms may often contribute to serious underestimates of the real variability. For example, in certain cases, ambient temperature, if ignored, can lead to designs that do not function at certain temperatures because of differential expansion of mating parts. For further examples of this, see also Taguchi, Elsayed and Hsiang (1989, p. 48).

In our own consulting work, we recently encountered a similar but slightly more subtle example involving the thickness of a transformer core laminated from 30 thin sheets of metal. In an initial analysis it was assumed the functional model for the thickness of the core was the straight algebraic sum of the thicknesses of the 30 laminates. With this assumption and assuming that all laminates were independent and identically distributed, it follows from (3) that the tolerance on the thickness of the core should be $\sqrt{30}$ times the tolerance on the thickness of the individual sheets. In practice, however, it was found that a high percentage of the cores were outside these specifications. A further analysis then revealed that part of the problem was due to laminates that were not geometrically flat or that had burrs. These problems substantially increased the thickness of the stack. Thus, in retrospect, it appears that an adequate tolerance analysis for this product should have included terms in (1) for flatness and burrs. If that was too complicated, some other allowance should have been made for likely violations of assumptions. A very simple fix of this nature is a tolerance inflation factor, to be discussed later.

Another source of model mis-specification problems is non-linearity. In that case an accurate analysis might require the inclusion in (1) of second and higher order terms from the Taylor series expansion. This issue was extensively investigated in an unpublished technical report by Tukey (1957) and later summarized by Ku (1966). They concluded that in most cases, the tolerance ranges are so narrow that all higher order terms can be safely ignored.

Although our experience largely supports Tukey's general findings, there are important cases where higher order terms are important. For example, Taguchi (1986) has advocated "parameter design," where the nominal values of certain design parameters are chosen so that first order terms in a Taylor expansion are at or near zero to minimize the transmitted variation. Incidentally, we are embarrassed to admit that we and many others previously

have overlooked that the basic idea of "parameter design" and "tolerance design," although not using that terminology, had earlier been set out by Shewhart (1931, p. 259). Nevertheless, parameter design is generally a very good practice when possible. However, tolerancing after parameter optimization must be done with caution.

Suppose, as an extreme example, that all the first order derivatives, b_i 's, were set to zero by parameter design. Then (3) would say that $\sigma_a = 0$. This is an unrealistic result that displays the ultimate deficiencies of the linear approximation, (1), because σ_a could never be reduced completely to zero. While this extreme case will rarely occur, it hints at a real problem that might occur when only one b_i is set at or near zero. One reason for trying to set a b_i to zero is to permit a wider tolerance interval, T_i . However, if T_i is widened too much, problems can arise that reduce substantially the validity of (3). For example, a second derivative can become important relative to the widened tolerance interval, even if it was negligible relative to a narrower tolerance interval. Similarly, a previously negligible interaction, b_{ij} with another X_j , might become important relative to the widened T_i . Thus, setting b_i to zero by tolerance design could tempt a user to expand T_i so much that the linear approximation, (1), is no longer sufficiently accurate. Since traditional statistical tolerancing, (3), assumes the validity of (1), excessive expansion of T_i could result in production problems not predicted by (3).

As a simple check of the importance of non-linearity for a given design, we recommend that tolerance engineers plot the functional relation around the optimal value. If such plots show marked curvature in a range covering $\pm 3\sigma_i$ for any of the input variables, it might be safer to include second order terms in (1)-(3) and/or do Monte Carlo simulations of a variety of tolerancing alternatives. However, more generally, as discussed by Bisgaard and Ankenman (1995) (see also Li and Wu (1996)), parameter design and tolerance design must be performed simultaneously if the system is actually to be optimized.

A more subtle problem related to mis-specification of the functional model occurs when the method of assembly has an impact on functionality. For example, a certain manufacturer had problems with excessive variability in the air-flow through a precision gas valve. To investigate this problem several valves were disassembled and the components reassembled in different combinations and subsequently tested. It was then found that the variability was large even when the same parts were reassembled. Thus it became apparent that there were sources of variability related to the assembly process itself not accounted for by the usual part-to-part variability. To deal with such problems, precise aspects of the assembly can, in principle at least, be introduced as additional terms in the functional

model, (1). Their presence in the error transmission formula, (2), would warn the user that accurate tolerancing depends on an appropriate assessment of assembly variability.

A related problem occurs if there is substantial measurement error, either in components or the assembly. Such measurement error can lead to good units being classified as "defective" and units outside specification being classified as "good". In some cases, measurement error should be included in the functional model, (1), and the resulting variation transmission model, (2).

Finally we wish to mention a type of problem with statistical tolerancing that has only recently surfaced in the literature where the use of (2) or (3) might be inappropriate. In some cases, the functional model is not continuously differentiable. Thus (1) might be a poor approximation to the real functional relation and everything based on it. In particular, (2) and (3) might be misleading. An interesting example of this involves a hinge consisting of two interlocking parts discussed by Parratt (1994) and Altschul and Scholz (1994). The functional model of the assembly involves the max or min functions and thus are not differentiable. A similar example involving the alignment of several holes was analyzed by Scholz (1996). For such problems, the safest approach seems to be Monte Carlo simulations.

To sum up this sub-section, we suspect that deficiencies in the functional model often lead to real, observed assembly variability that is higher than estimated by (3). This might lead to the assignment of component tolerances that are wider than required to properly assure assembly performance, which in turn might result in unpredicted assembly problems. However, the opposite problem may also occur. Suppose, for example, the value used for one of the b_i 's in the functional model (1) is too large. This would encourage an engineer to specify tolerances tighter than necessary on that component in order to achieve a desired tolerance for the assembly characteristic. In turn this might unnecessarily increase costs and in some cases could jeopardize the competitive position of the manufacturer.

Failure Mode 2. Desire Higher Process Capability in Assembly Than Delivered for Components.

As previously mentioned, one assumption required to obtain expression (3) from (2) is that all component tolerances represent the same number of standard deviations, $T_a = 2Z_a\sigma_a$ and $T_i = 2Z_i\sigma_i$ with $Z_a = Z_i$, $i = 1, \dots, n$. (In practice, many users assume that $Z_a = Z_i = 3$.) However, this subtle assumption can generate problems for a pro-

ducer who expects from (3) that $Z_a \geq 3$, without checking the actual process variances for the components. If for any of the components $Z_i < 2$, this might lead to serious underestimates of the true assembly variability and hence the compliance with T_a . A catchy way to say this is that "variances transmit; tolerances do not transmit." Thus (2) is more generally applicable than (3) since it relies on fewer assumptions.

Partly to assure that the component variances are reasonably in compliance with the assumption that $T_i = 2Z_i\sigma_i$ and specifically that $Z_i \geq 3$, it has increasingly become industrial practice to use "process capability indices" such as

$$C_{pk} = \frac{\min\{U - \mu, \mu - L\}}{3\sigma} \quad (4)$$

where μ is the process mean, and σ is the process standard deviation. For a comprehensive discussion of process capability indices, see Kotz and Johnson (1993).

This discussion of Failure Mode 2 will consider problems encountered when (a) C_{pk} is assumed to be good just because no defects are found, (b) C_{pk} is estimated from short term variability, treating longer term variability as bias, and (c) C_{pk} is estimated from prototypes and samples. Each of these situations has arisen in our industrial experience, and each can produce problems with traditional statistical tolerancing.

Failure Mode 2a. "Zero Defects" Does Not Imply Acceptable Process Capability. To see the relationship between (3) and (4) suppose all component processes are properly centered so that for the assembly characteristic $\mu = \theta$. Expression (4) then becomes $C_{pk} = T/6\sigma$, or $T = 6\sigma C_{pk}$. Since $T = 2Z\sigma$, this implies that $C_{pk} = Z/3$. Thus the assumption that $Z = 3$ is equivalent to assuming $C_{pk} = 1$, and assuming $Z_a = Z_i$ is equivalent to assuming $C_{pk,a} = C_{pk,i}$. Therefore, if a user assumes that (3) will give $C_{pk,a} \geq 1$ without checking the individual $C_{pk,i}$'s, the result can be quite misleading. This is especially true if an important $C_{pk,i}$ is substantially less than 1.

The subtle assumption that $C_{pk,a} = C_{pk,i}$ can often be a problem because some users naively may expect that if they haven't heard about quality problems from the production floor, the C_{pk} 's must be satisfactory. However, as we will now show, it is possible to have poor process capabilities even if none of the products fall outside the specifications. For example, a customer may not know that the vendor scrapped or reworked many units leading to an almost uniform distribution of the outgoing products. Thus the customer receiving these products will only see the (virtually) defect-free, roughly uniform distribution between the upper and lower specification limits delivered with a $C_{pk} = 0.58$ as a result of this sorting and reworking.

For another example, it used to be common practice to obtain precision resistors by sorting so that 2% resistors were what were left after the 1% resistors had been removed. If we assume perfect inspection, we might model such a production as a 50-50 percent mixture of one distribution that is uniform between $(-2, -1)$, and another uniform between $(+1, +2)$. In this case the customer may not see the rejects, but only the (virtually) defect-free outgoing distribution with a $C_{pk} = 0.44$. While the technology for producing precision resistors has improved substantially over the past few decades, it might still be possible to find components produced with (nearly) zero defects in spite of a low C_{pk} . Thus for both of these examples using (3) for the tolerances will be quite misleading because of the implicit assumption that $C_{pk} = 1$. A safer approach would be to compute the assembly variance using (2) and appropriately estimated component variances.

It is important to notice that the problem illustrated with the two examples above is not related to the non-normal distributional form, an issue we will discuss below under Failure Mode 5, but rather to low process capability. First we recall the Central Limit Theorem assures us the distribution of a product characteristic will be nearer to the normal than the distributions of components. Of course, there are situations where the Central Limit Theorem does not apply, as we will discuss under Failure Mode 5. However, our experience suggests that these cases may be rare. Moreover, when component distributions have zero defects (relative to two-sided specification limits), the assumptions of the Central Limit Theorem are always satisfied. Indeed, we will now show that with identical uniform distributions, the Central Limit effect works well even with only two components. To see that, we have in Figure 1 summarized the case of two components with identical distributions that are uniform between L and U , so $C_{pk} = 0.58$. It is relatively straightforward to show that the sum of two independent random variables with identical uniform distributions has a triangular distribution (Johnson and Kotz (1970b, p. 64)). When statistical tolerancing, (3), is used with the parameters of this example, the resulting triangle distribution has a defect rate of 8.6 percent. For comparison, a (centered) normal distribution with a C_{pk} of 0.58 has a defect rate of 8.3 percent, which is only slightly different from the 8.6 percent of the triangle distribution it approximates. While the normal approximation to this triangle distribution is not *this* good over its entire range, we see the high defect rate in this case is largely due to the low C_{pk} and not to the lack of normality.

This issue is not just academic because many manufacturers have some very capable processes, for example with $C_{pk} \geq 1.5$ for easily made parts, while at the same time some very poor process capabilities, for example with C_{pk}

Figure 1. Statistical tolerancing applied to the sum of two uniform distributions.

$\cong 0.58$, for parts that are difficult to make. One possible reason for this might be that the more difficult parts may require repeated adjustments to satisfy specifications, thereby likely producing approximately uniformly distributed parts between L and U .

It might seem obvious to some readers that a manufacturer should check an assumption as critical as the level of process capability available in components. We agree. Unfortunately, it may not be enough just to check this assumption; it must be checked *carefully* to avoid the problems we now describe with Failure Modes 2b and 2c.

Failure Mode 2b. Estimating Process Capability from Short Term Variability. If component characteristics are not in statistical control, then effective tolerances must be based on the *total variability*, not just within the subgroup or short term variability. For example, suppose we want the thickness of a flat washer to be between $L = 2.455$ mm and $U = 2.545$ mm, with a nominal $q = 2.50$ mm. Further suppose the process average on one day is 2.52 mm with a standard deviation of 0.007 mm. For that day, the process has a bias of $2.52 - 2.50 = 0.02$ and a C_{pk} of $(2.545 - 2.52)/0.021 = 1.19$. Thus the process superficially looks quite capable. However, suppose another check of the process the next day reveals that the process now averages 2.48 mm, again with standard deviation 0.007 mm and $C_{pk} = 1.19$. In that case the bias is $2.48 - 2.50 = -0.02$. A closer look now suggests the process has a serious problem be-

cause the daily averages also follow a distribution. This phenomenon is known as variance components; within days the process follows a distribution with a mean that itself varies randomly from day to day. Now suppose the standard deviation of the daily averages is 0.02 mm. Then the overall standard deviation is $\sqrt{(0.02)^2 + (0.007)^2} = 0.021$ mm. (This computation uses the fact that the overall variance is the expectation of the conditional variance plus the variance of the conditional expectation: $\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]$ where Y identifies the day from which the data were taken; see Rao (1973, sec. 2b.3).) For our process this makes the overall $C_{pk} = 0.045/[3(0.021)] = 0.71$. Now if this is a critical component characteristic and tolerances for the assembly are set using (3) thinking the $C_{pk,i}$'s are 1 or 1.2, this could result in unanticipated production problems. A somewhat similar situation was discussed by Shore (1997). He showed that autocorrelated observations can lead to overestimating C_{pk} . This can generate the same kinds of problems discussed here.

In general a process that seems acceptable based on data from short term samples may be lacking in overall process capability if there are substantial variance components or autocorrelation effects. Frequently therefore, short term standard deviation estimates are a gross underestimate of the total longer term variability. Hence if the short term variability is used as an estimate in calculating process capabilities, this could contribute to tolerancing problems.

Product designers reading this advice may reply that it is not relevant to them: They need to set specifications for products that have never been produced before and for which reasonably accurate process capability data will not exist until mature production is reached. In some cases, relevant process capability data can be obtained from related processes used to produce existing products. In other cases, they must rely on prototypes and pre-production samples from vendors, which we now consider.

Failure Mode 2c. Estimating Process Capability from Prototypes and Samples. Leading edge manufacturers must often set specifications based on a handful of prototypes. These prototypes may have been "tweaked" to get them to work properly and may not be representative of future production. The standard deviation and C_{pk} based on the prototypes could lead people to expect smooth production when substantial problems await the unwary. The opposite could also happen: If the new product introduction plan includes resources for rapid improvements in process capability, initial estimates of C_{pk} could be overly pessimistic, pushing engineers to tolerance certain components much tighter than will ultimately be necessary. A similar problem can arise from tests of samples that may have been carefully selected by a vendor who wanted to make the best possible impression as a basis for future business.

In some circumstances, the only data available come from engineering prototypes and / or vendor test samples. Setting tolerances based on these is generally risky because the prototypes and samples will not, in many cases, be sufficiently representative of future production. This problem can be handled by using an inflation factor that multiplies the standard deviation of the sample by some number larger than one. When this is done, a written record should be kept of the prototype data and later compared with mature estimates of the production process capability. In this way, an engineering organization can learn over time how much their own engineering prototypes differ from production, and how much a particular vendor's test samples differ from volume deliveries. This history can be used for future products to decide how much to inflate standard deviations estimated from early prototypes or vendor samples. These data can also be used to estimate the value of better data.

There is a second issue here: In many cases only a very small number of prototypes or samples are tested. This issue can be handled with what statisticians call "tolerance intervals" (Odeh and Owen (1980)). In many cases, however, the small number of units may be less of a problem than the question of whether the units are representative of future production.

To sum up this sub-section we hope the above discussion and examples help illustrate the general principle that if a critical component characteristic has substantially less process capability (i.e., larger standard deviation) than assumed, traditional statistical tolerancing based on (3) can lead to an unanticipated high defect rate for the assemblies.

Failure Mode 3. Bias.

If the process mean differs from the nominal, the distribution is said to be biased. The next three subsections describe three ways in which bias can result in poor process capability in an assembly, even though the process capability indices for components might look good.

Failure Mode 3a. Biases Accumulate Randomly. In some cases, biases may accumulate randomly. If both a shaft and a hole tend to be too large, the clearance will not be affected. However, if the biases accumulate randomly, some (but not all) of the biases may cancel. A simple example involves an extension of Failure Mode 2b above. Suppose, for example, that both shaft and hole have substantial bias within a day but are unbiased in the long term with overall $C_{pk} = 1$. Suppose also that tolerances are set using (3). From the normal approximation, we estimate that the long term defect rate will be approximately 0.27 percent. Thus, if shafts and holes are thoroughly mixed in inventory waiting to be assembled, then the defect rate will be in statistical control, averaging 0.27 percent each day. On the other hand, if Just-In-Time manufacturing is used, the lack of process control evident in the production of shafts and holes will be shared by the assembly. In that case, the 0.27 percent defect rate may appear as 2.7 days out of 1,000, roughly one day per year, during which nothing fits or the clearance is greater than specified. Thus, in some cases, biases accumulate randomly and some cancellation is achieved. However, it is not advisable to count on that without checking, as we explain in the next two subsections.

Failure Mode 3b. Biases are Systematic. While random biases do occur, systematic biases may be more common. To appreciate this let us consider a machine part with a critical outer dimension that must be 10.0 ± 0.03 mm and suppose the process standard deviation is 0.01 mm. Suppose further that it costs \$3 to rework a part if this dimension is too large, but \$60 to scrap it if it is too small. The total lifetime cost in assembly and use of the part may be approximately parabolic, as suggested by Taguchi (1986). A cost model of this nature may summarize "loss to society," as Taguchi says, or only the direct and indirect costs as calculated by a manufacturer's accountant. Nevertheless, in many cases, the machinists' pay and promotion are tied only to the parts produced within specifications,

and not to "loss to society." Thus in such situations, management may implicitly have established a loss function for the machinist that does not include considerations of possible downstream costs. Consequently a rational machinist may likely offset the process mean to reduce scrap while increasing rework. This offset may come close to minimizing the expected cost of scrap and rework even if this expected cost is never formally calculated. For our example, the combined expected cost is \$3 times the proportion of parts that are too large plus \$60 times the proportion that are too small. Assuming this dimension is normally distributed, this quantity is plotted in Figure 2 as a function of the mean. From the graph, we see the minimum of the costs visible to the machinist is achieved when the process mean, μ , is approximately 10.005 mm, where it is biased by 0.5σ . Therefore, in this or similar situations, it is not uncommon that machinists will offset the process mean to be on the safe side, reducing expensive and non-reworkable scrap.

The bias, in this example, may or may not be important depending on the biases of the other parts that must mate to produce an assembly and how they geometrically stack up. In some assemblies, biases may cancel and in others, they may not. For example, if the machine part discussed above is a shaft to be inserted into a hole, a typical ma-

chinist will try to make the shafts larger than the nominal and similarly make the holes, on average, smaller. However, the net result of this is that assemblers further downstream receive shafts that are large to put into holes that are small! Thus because of the geometry of the design and the machinist's loss function, the biases may add up most disadvantageously. In certain other situations, an experienced engineer may be able to determine from the design whether biases are likely to cancel, at least in part, or to accumulate disadvantageously. If they accumulate most disadvantageously, the results in some cases can be virtually 100 percent defective assemblies from 100 percent good parts, as we now explain.

Failure Mode 3c. Biases Accumulate Most Disadvantageously. Extreme examples of biases accumulating disadvantageously were described by Bisgaard and Graves (1997). One of their examples was another aspect of the transformer core problem described with Failure Mode 1. The 30 thin sheets of metal that formed the core were cut from large rolls that were thicker at the ends than in the middle. In fact, they were so much thicker (i.e., biased) that when cores of 30 laminates from the beginning and end of a roll were assembled, many would not fit in the required space.

Figure 2. Combined Expected Cost of Scrap and Rework.

To understand this problem, let us use the process centering coefficient discussed by Liggett (1993, pp. 60-62), defined as

$$C_c = 2|\bar{X} - \theta|/T \quad (5)$$

where μ is the process mean. Thus for later comparison, the Motorola "six sigma" program (Harry and Stewart (1988); Harry (1992)), which limits bias to 1.5σ , requires the process centering coefficient to be $C_c \leq 2(1.5)/12 = 0.25$ and $C_{pk} \geq 1.5$. In general, for tolerancing the height of a stack of n (distributionally) identical pieces, e.g., flat washers or laminates for transformer cores, we get from (3) that $T_a = T_1\sqrt{n}$, $\theta_a = n\theta_1$, and $\mu_a = n\mu_1$. Combining these with (5) gives

$$C_{c,a} = C_{c,1}\sqrt{n}. \quad (6)$$

Thus, if biases accumulate most disadvantageously and the process for an individual component has even the slightest bias, then for a sufficiently long tolerance chain, the process mean will be far outside the specification limits, and almost 100 percent of products will be defective! Moreover, the tolerance chain does not have to be long for problems to appear. To see this, we need to understand the impact of n and $C_{c,1}$ on $C_{pk,a}$.

First, to determine the general impact of C_c on C_{pk} , we combine (4) and (5) to get

$$C_{pk} = \frac{T(1 - C_c)}{6\sigma}.$$

Then, for the height of a stack of n (distributionally) identical components with bias accumulating most disadvantageously, we get

$$C_{pk,a} = \frac{T_1(1 - C_{c,1}\sqrt{n})}{6\sigma_1} = C_{pk,1} \left(\frac{1 - C_{c,1}\sqrt{n}}{1 - C_{c,1}} \right). \quad (7)$$

Thus we see that if the bias is non-zero, $C_{c,1}$ will be positive, and this will make $C_{pk,a}$ negative for a sufficiently long tolerance chain – no matter what $C_{pk,1}$ is! For the transformer with 30 laminates, suppose $C_{c,1} = 0.25$ and $C_{pk,1} = 1.5$, satisfying Motorola's 6 sigma requirements. Then from (6) and (7), we get $C_{c,a} = 1.37$, and $C_{pk,a} = -0.74$. A negative value for C_{pk} implies that the mean is outside the specification limits, so the good units must be at least $3(0.74) = 2.22$ standard deviations from the mean towards to the nominal, which makes the defect rate 98.7 percent. The same situation with $n = 2$ gives us $C_{c,a} = 0.35$, $C_{pk,a} = 1.29$, and 53 parts per million (ppm) defects instead of 3.4 ppm, the "6 sigma" standard. Thus, $n = 2$ will rarely create a problem with "6 sigma." However, if $T_a = 6\sigma_a$ instead of $12\sigma_a$, then with $C_c = 0.25$ the defect

rate is 2.8 percent when $n = 2$, and gets worse fairly rapidly as n increases. Thus, as mentioned above, the tolerance chain does not have to be long for problems to appear.

One might wonder if using a different process capability index might better signal this problem with bias. A logical candidate is the so-called Taguchi capability index (Kotz and Johnson (1993, ch. 3); Spiring (1997)):

$$C_{pm} = \frac{T}{6\sqrt{\sigma^2 + (\mu - \theta)^2}} \quad (8)$$

We now substitute (5) into (8) to get

$$C_{pm} = \frac{T}{6\sqrt{\sigma^2 + (C_c T/2)^2}}.$$

For the height of a stack of n (distributionally) identical components (e.g., flat washers or sheet steel laminates), we get

$$C_{pm,a} = \frac{T_1\sqrt{n}}{6\sqrt{n\sigma_1^2 + (nC_{c,1}T_1/2)^2}}.$$

We see from this that for large n ,

$$C_{pm,a} \approx 1/(3C_{c,1}\sqrt{n}),$$

regardless of the size of $C_{pm,1}$ and σ_1 , assuming non-zero bias so $C_{c,1} > 0$ as before. Thus this means that for sufficiently large n , $C_{pm,a}$ can be made arbitrarily small. Consequently, neither C_{pm} nor C_{pk} for the components will provide adequate warning of potential assembly problems if we have biased component distributions.

Before leaving this topic, we note that Taguchi and his followers have lately used a quadratic loss function to argue that processes should be centered. This has in fact been heralded as one of the key differences between his new approach to quality and the more conventional Western tradition. While we agree that quadratic loss approximates reality more closely than the 0-1 loss implied by specification limits, we don't need quadratic loss to argue that distributions should be centered, as the discussions in this section illustrated. It may have sometimes been forgotten, but the absence of bias has always been a key assumption behind traditional statistical tolerancing. Thus, we agree with Taguchi that ignoring bias will frequently generate serious difficulties, but for a more fundamental reason than quadratic loss.

Failure Mode 4. Correlation Between Component Characteristics.

In the derivation of (3), it is assumed that the component

characteristics are uncorrelated. This is often an assumption that is easily met in practice when the characteristics in a given tolerance chain are all on different parts made by different processes. However, it may not always be so. For example, consider a rubber gasket where both the diameter and thickness affect the seal. In that case thicker gaskets may tend to have smaller diameters and vice versa, so the two dimensions might be negatively correlated. For an accurate assessment of how this gasket and the rest of the assembly should be toleranced, a correlation like this might be important.

Another example where it might be unwise to assume that the dimensions are uncorrelated without checking is integrated circuits. Because of the way such circuits are manufactured, values of different parameters on a given circuit might very likely be correlated, and statistical tolerancing using (3) might be problematic.

A succinct analysis of problems with correlation requires vector notation, which allows us to write (1) as

$$X_a \equiv \theta_a + \mathbf{b}'(\mathbf{X} - \boldsymbol{\theta}) \quad (9)$$

where $\mathbf{b} = (b_1, \dots, b_n)'$ and $\mathbf{b}'(\mathbf{X} - \boldsymbol{\theta})$ is the product of a row and a column vector that produces the sum in (1). If we then transfer θ_a to the other side of the equality, square the result and take expected value we get

$$\sigma_a^2 \equiv \mathbf{b}' \boldsymbol{\Sigma} \mathbf{b} \quad (10)$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of $\mathbf{X}$. This expression is a generalization of (2) to the case of correlated component distributions.

To see how this works, let us again consider the example of tolerancing the height of a stack of n components (washers or sheet steel laminates), where now the correlation between the heights of any two components in the stack is ρ . Then the (i, j) th element of $\boldsymbol{\Sigma}$ is $\sigma_{ij} = \sigma^2$ when $i = j$ and $\sigma^2\rho$. For example, the flat washers discussed under Failure Mode 3a, might be correlated in this fashion. Specifically, if X_i and X_j are the thickness of two washers produced on the same day and μ_d is the mean for that day, then X_i can be written as $\mu_d + e_i$ with $Ee_i = 0$, $\text{Var}(e_i) = (0.01)^2$, $\text{Var}(\mu_d) = (0.03)^2$, and

$$\begin{aligned} \sigma_{ij} &= \text{Cov}(X_i, X_j) = E(X_i - 2.5)(X_j - 2.5) \\ &= E[(\mu_d - 2.5)^2 + (\mu_d - 2.5)(e_i + e_j) + e_i e_j] \\ &= \begin{cases} (0.03)^2 + (0.01)^2 = (0.032)^2 & \text{if } i = j \\ (0.03)^2 & \text{if } i \neq j \end{cases} \end{aligned}$$

Whence, $\sigma^2 = (0.032)^2$, as noted above, and if $\rho = (0.03/0.032)^2 = 0.90$.

For the height of an n -washer assembly, $\mathbf{b} = (1, \dots, 1)'$ in (9), and (10) becomes $\sigma_a^2 = n\sigma^2[1 + (n-1)\rho]$. Now if $\rho > 0$, then for large n , $\sigma_a \equiv n\sigma\sqrt{\rho}$. Thus we see that in this case, σ_a increases with n , not $\sqrt{n}$ as in the independent case. It can therefore be seen that in this and similar cases, the variability of the assembly can be substantially greater than the variability assumed by traditional statistical tolerancing using (3). While this effect may be somewhat rare in mechanical assemblies, it can be a major issue with different circuit elements on a given integrated circuit, and in the manufacturing of chemical products where positive autocorrelations are common.

Failure Mode 5. Component Dimensions May Not Be Normally Distributed.

The Central Limit Theorem shows that in "regular" cases the distribution of a weighted sum of a "large number" of independent random variables (e.g., characteristics of components in a product) will be approximately normally distributed, even if the distributions of the individual components are not (Serfling (1980); Skovgaard (1981); Graves (1983, pp. 226-255)). Of course, no general statement can be made about what constitutes a "large number" of component characteristics / independent random variables. However, as we saw in the example of Figure 1, $n = 2$ was almost large in that case! Moreover, in many other cases n does not need to be very large before the normal distribution provides a very good approximation.

To evaluate the utility of the Central Limit Theorem and a normal approximation for distributions commonly encountered in manufacturing, we recall the many histograms and probability plots of industrial data that we have seen. In most cases these are "regular." However, there are two specific scenarios that we have seen that would limit the value of the normal distribution as a reasonable engineering approximation to a production distribution. These are bimodality and exceptionally long tails.

Let us first discuss bimodal distributions. The extreme case of bimodality is a 2-point distribution hugging the specification limits, similar to the example described above where 2 percent resistors were what remained in a batch after the 1 percent resistors had been removed. If virtually all components follow such a distribution, an n as large as 5 or 10 might be required to obtain reasonable engineering accuracy from assuming a normal distribution for assembly characteristics.

The second "non-regular" case is extreme long distributional tails. Admittedly we have, in practice, seen production distributions with long tails, but have not checked to see if they were so extreme that the Central Limit Theorem would not apply. In many cases, long-tailed produc-

tion might be accompanied by difficulties that would lead to 100 percent inspection, thereby eliminating the tails. Even without the truncation, however, most long tailed distributions are still subject to the Central Limit effect.

For completeness, however, let us consider a special case with tails so long that the Central Limit Theorem does not apply. Again we will use the example of tolerancing the height of a stack of n washers. Specifically suppose that the thickness of an individual washer followed a Student's t distribution with 1 degree of freedom, also called the Cauchy distribution (Johnson and Kotz (1970a, ch. 16; 1970b, ch. 27)). It is well-known that this distribution does not satisfy the regularity conditions of the Central Limit Theorem because its standard deviation is infinite. In regular cases, the spread of the distribution of an average of n independent, identically distributed random variables declines as $1/\sqrt{n}$. However, for Student's t with 1 degree of freedom, the distribution of an average is the same as that of an individual summand and the spread does not shrink at all! Thus, the spread of the distribution of a sum (rather than an average) of n such random variables is n times the spread of an individual component. For tolerancing, this means that if the distributions of characteristics of components were all Student's t with 1 degree of freedom, statistical tolerancing using (3), would be grossly misleading.

To conclude this sub-section on normality, the Central Limit Theorem seems to apply to most manufacturing distributions to assure us that the normal distribution will generally provide reasonable engineering approximations to manufacturing distributions. Certainly non-normality may be a problem in some cases but we doubt that it creates difficulties as often as Failure Modes 1-4, described above.

How to Avoid Tolerancing Problems Without the Extra Cost of Worst Case Tolerancing

As we have shown above, there are many ways in which the assumptions required for statistical tolerancing based on (3) can be violated. Many of these violations occur in practice and can cause serious difficulties. Thus, (3) is highly non-robust. One response to this would be simply to avoid using statistical tolerancing, and instead use what is called worst case tolerancing. Undeniably this may be the best strategy for a company with low volume production, limited competition, and sales that depend on engineering innovation. However, with higher volume production, modest per-unit savings can add up and quickly justify the nominal effort required to set more balanced tolerances.

We have therefore three recommendations for organizations with higher volume production: (a) Support im-

proved tolerancing with shop floor education and informed management of production to establish stable and capable processes, centered at the nominal, except for the rare cases where that is not appropriate. (b) Set tolerances using actual process capability data and expressions (9) and (10). (c) If that can not be done adequately, use inflated statistical tolerancing, described below. We will now discuss each of these strategies.

Improved Process Management: Leading manufacturers the world over today routinely maintain statistical control charts to identify opportunities for process improvement. As a result, many manufacturing processes are in statistical control. For the many processes that are not in statistical control, there are people available who know how to seek out the causes of problems and change the processes so the difficulties are less likely to occur. In many cases, this involves educating and training of managers and others in the relevant concepts of process control and improvement as well as the issues discussed in this article. In fact, the discussion in this article merely provides other reasons for continuing to pursue stable and capable processes, centered at the nominal.

We admit that situations exist where it may be desired to bias the process: If an industrial customer only looks at the percent within specifications, and our cost structure is similar to that discussed with Figure 2, then our best strategy may be to bias the process as described in that example. Even in that case, however, it is still wise to pursue process control and improvement. In this vein, Deming (1984, e.g., p. 339) commented, "Once a process [is] in statistical control, ... the specifications that it can meet are predictable." Without process control, there is no basis for prediction and no basis for estimating defect rates in tolerancing.

Tolerance Using Process Capability Data: If sufficient information about all component distributions is available, then it is best to compute the process mean from the expectation of expression (9):

$$\mu_a = \theta_a + b'(\mu - \theta),$$

and σ_a from the more general expression (10). This will avoid problems of bias and correlation as well as the often unchecked assumption that all the component process capabilities are the same. From (10) we can then set $(L_a, U_a) = \theta_a \pm (|\mu_a - \theta_a| + Z_a \sigma_a)$, as recommended by Harry and Stewart (1988), where, e.g., $Z_a = 3$.

Inflated Statistical Tolerancing: With Failure Modes 1-4, we described a number of problems that seem to occur with disturbing regularity and often create problems in production or warranty for the unwary user of traditional statistical tolerancing. Most of these problems can be fixed

with additional data and appropriate modifications of (1)-(3). However, in many cases, this detailed information may not be available when tolerances must be established. This will especially be the case, as already discussed, when tolerances for new products are chosen. An alternative procedure is "inflated statistical tolerancing," advocated by Gilson (1951) and Bender (1962) who found from their industrial experience that

$$T_a = f \sqrt{\sum a_i^2 T_i^2} \quad (11)$$

worked pretty well with $f = 1.5$ or 1.6 . Graves (1994) showed how this expression can be derived by adding a discrepancy term to (1), representing, e.g., measurement error, and by making other assumptions. Graves (1997) further noted that this expression might work better for some types of assemblies than others. He also suggested developing different inflation factors, f , for initial vs. mature production (to shorten time to market in concurrent engineering), high vs. low volume production, and mature vs. developing technology and metrology. Alternatives to inflated statistical tolerancing, (11), are also reviewed by Graves (1994) and Scholz (1995a, b).

Discussion and Conclusion

In this article we outlined several different ways that traditional statistical tolerancing can fail and thus lead to unexpected high rates of problems in production. These problems may be so common that most people who have tried to use traditional statistical tolerancing may have encountered unexpected problems in production. In fact, there has been speculation for years that statistical tolerancing may not be widely used in industry (e.g., Tipnis (1992, p. 11)), in spite of its apparent intuitive appeal. The problems outlined here may provide a partial explanation for the alleged non-use. The discussion of potential causes was followed by suggestions for how to avoid them, including improved process management, more careful analysis of process capability, and inflated statistical tolerancing.

Acknowledgments

This research was supported by grant number DMI 9500140 from the National Science Foundation.

References

- Altschul, R. E. and Scholz, F. W., "Case Study in Statistical Tolerancing," *Manufacturing Review*, 7, 52-56, 1994.
- Bender, A., "Benderizing Tolerances," *Graphic Science*, December, 17-21, 1962.
- Bisgaard, S. and Ankenman, B., "Analytic Parameter Design," *Quality Engineering*, 8, 75-91, 1995.
- Bisgaard, S. and Graves, S. B., "A Negative Process Capability Index from Assembling Good Components? A Problem in Statistical Tolerancing," *Quality Engineering*, 10(2), 409-414, 1997.
- Brandenberger, H., *Toleranzen, Passung und Konstruktion*. Schweizer Druck- und Verlagshaus, Zürich, 1946.
- Bruyevich, N. G., *Tochnost Mekhanizmov* (in Russian) Mashgiz, 1946, cited from G. H. Sutherland and B. Roth, "Mechanism Design: Accounting for Manufacturing Tolerances and Costs in Function Generation Problems," *Transactions of the ASME: Journal of Engineering for Industry*, February, 283-286, 1975.
- Deming, W. E., *Out of the Crisis*, MIT Center for Advanced Engineering Study, Cambridge, 1984.
- Gilson, J., *A New Approach to Engineering Tolerances*, The Machinery Publishing Co., London, 1951.
- Grant, E. L., *Statistical Quality Control*, McGraw-Hill, New York, 1946.
- Graves, S. B., *Edgeworth Expansions for Discrete Sums and Logistic Regression*, Ph.D. diss., University of Wisconsin-Madison, 1983.
- , "Tolerance Analysis Considering Manufacturing Variability and the Cost of Deviating from the Nominal," *CQPI Technical Report No. 117*, May 1994.
- , "How to Reduce Costs Using a Tolerance Analysis Formula Tailored to Your Organization," *CQPI Technical Report No. 157*, April 1997.
- Harry, M. J., "The Vision of Six Sigma Quality: Producibility Measurement and Analysis," in *Tipnis* (1992b, 41-64). 1992.
- Harry, M. J. and Stewart, R., *Six Sigma Mechanical Design Tolerancing*, Motorola University Press, Schaumburg, 1988.
- Hounshell, D. A., *From the American System to Mass Production, 1800-1932*, Johns Hopkins U. Press, Baltimore, 1984.
- Johnson, N. L. and Kotz, S., *Continuous Univariate Distributions - 1*, Houghton Mifflin, New York, 1970a.
- , *Continuous Univariate Distributions - 2*, Houghton Mifflin, New York, 1970b.
- Juran, J. M., *Quality Control Handbook*, McGraw-Hill, New York, 1951.

- Juran, J. M. and Gryna, F. M., *Quality Planning and Analysis*, 3rd ed., McGraw-Hill, New York, 1980.
- Kotz, S. and Johnson, N. L., *Process Capability Indices*, Chapman and Hall, New York, 1993.
- Ku, H. H., "Notes on the Use of Propagation of Error Formulas," *Journal of Research at the National Bureau of Standards*, 70(2), 263-273; reprinted in H. H. Ku, ed., *Precision Measurements and Calibration, National Bureau of Standards Special Publication 300-vol. 1.*, U.S. Government Printing Office, Washington, DC, 1966.
- Li, W. W. and Wu, C. F. J., "An Integrated Method of Parameter Design and Tolerance Design," Technical Report 271, Department of Statistics, University of Michigan, Ann Arbor, 1996.
- Liggett, J. V., *Dimensional Variation Management Handbook*, Prentice Hall, Englewood Cliffs, 1993.
- Odeh, R. E. and Owen, D. B., *Tables for Normal Tolerance Limits, Sampling Plans and Screening*, Marcel Dekker, New York, 1980.
- Parratt, S. W., *A Theory of One-Dimensional Tolerancing for Assembly*, Ph.D. diss., Cornell University-Ithaca, 1994.
- Rao, C. R., *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.
- Rüdenberg, R., "Die Beurteilung elektrischer Maschinen und Apparate durch Toleranzen auf Grund statistischer Methoden," *Zeitschrift für angewandte Mathematik und Mechanik*, 9, 318-334, August 1929; reprinted as 40-56 in H. C. Plaut, *Fabrikationskontrolle auf Grund statistischer Methoden*. VDI-Verlag, Berlin, 1930.
- Scholz, F., "Tolerance Stack Analysis Methods: A Critical Review," ISSTECH-95-021, Boeing, Seattle, 1995a.
- , "Tolerance Stack Analysis Methods," ISSTECH-95-030, Boeing, Seattle, 1995b.
- , "Hole Pinning Clearance," ISSTECH-96-028, Boeing, Seattle, 1996d.
- Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
- Shewhart, W. A., *Economic Control of Quality of Manufactured Product*, Van Nostrand, 1931.
- , *Statistical Method from the Viewpoint of Quality Control*, Graduate School, Department of Agriculture, Washington, DC, 1939; Reprinted, Dover, New York, 1986.
- Shore, H., "Process Capability Analysis When Data Are Autocorrelated," *Quality Engineering*, 9, 615-626, 1997.
- Skovgaard, Ib M., "Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions," *Scandinavian Journal of Statistics*, 8, 207-217, 1981.
- Spiring, F. A., "A Unifying Approach to Process Capability Indices," *Journal of Quality Technology*, 29, 49-58, 1997.
- Taguchi, G., *Introduction to Quality Engineering*, Asian Productivity Organization, Tokyo, 1986.
- Taguchi, G., Elsayed, E. A. and Hsiang, T. C., *Quality Engineering in Production Systems*, McGraw-Hill, New York, 1989.
- Tipnis, V. A., *Tolerance and Deviation Information*, American Society of Mechanical Engineers, New York, 1992.
- Tukey, J., "The Propagation of Errors, Fluctuations and Tolerances," Technical Reports 10-12, Princeton U. Statistics Dept., Princeton, 1957.