We All Have to Use Less Than Perfect Data

And process behavior charts work better than you know with imperfect data

Donald J. Wheeler

In their recent article "We Do Need Good Measurements," Professors Stefan H. Steiner and R. Jock MacKay take exception to two of my *Quality Digest* articles: "Don't We Need Good Measurements?" and "The Intraclass Correlation Coefficient." While we all want good measurements, the trick is in learning to live with imperfect measurements.

There seem to be two major points to Steiner and MacKay's critique. The first pertains to Figure 1 below, and the second concerns my interpretation of what the curves in Figure 1 mean in practice. As we investigate their criticisms we will discover some divergent world views which I will discuss in the latter part of this paper.

The curves in Figure 1 show the effect of increasing measurement error upon the ability of a process behavior chart to detect process changes. The vertical axis shows the probability that a chart for location will detect a sustained shift within ten subgroups of when that shift occurs. The horizontal axis uses the standard measure of relative utility for a measurement procedure—the intraclass correlation coefficient (*ICC*). As we move from left to right measurement error will contribute a larger share of the total variation, and the *ICC* value will drop.
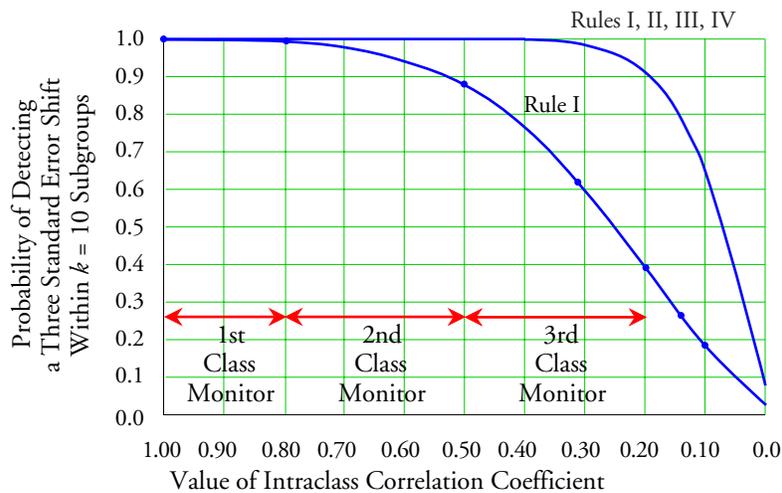

Figure 1: The Probabilities of Detecting Shifts of Three Standard Errors

The two curves in Figure 1 are for a sustained three standard error shift in location. The upper curve is for the case where all four of the Western Electric zone tests are used, and the lower curve is for the use of Detection Rule One alone.

In my articles I did not explain the derivation of the curves in Figure 1. These curves come from tables given in Chapter 15 of *EMP III: Evaluating the Measurement Process and Using*

Imperfect Data. A portion of these tables is reproduced below as Figure 3. These tables are based on my tables of the power function from ***Journal of Quality Technology v. 15, no. 4***, October, 1983, pp.155-169.

It would appear that Steiner and MacKay did not believe the curves of Figure 1. They definitely misconstrued how I computed those curves. So they chose to compute their own curves using Detection Rule One and looking at shifts of one, two, and three standard errors. Their resulting graph is shown in Figure 2. The reader should note that it is the top curve in Figure 2 that is analogous to the bottom curve in Figure 1. The other curves in both figures are for different scenarios and are not directly comparable.
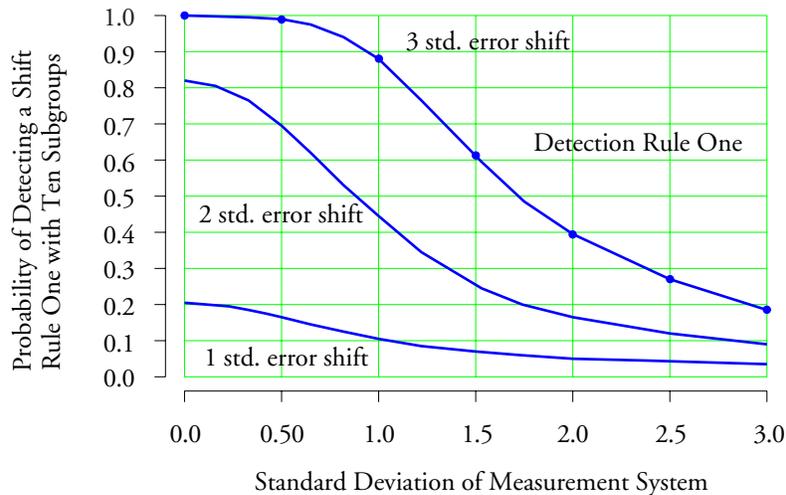

Figure 2: Steiner and MacKay's Probabilities of Detecting Shifts with Rule One

Steiner and MacKay claim that Figure 2 shows that the process behavior chart will become insensitive to small shifts as measurement error increases. This argument is disingenuous for two reasons. First, they are looking at small shifts but using Detection Rule One alone.   (Rule One was never intended to be sensitive to small shifts.) Consequently Figure 2 does not actually describe the sensitivity of a process behavior chart to small shifts.

And the second reason that Figure 2 is disingenuous is that, for equivalent amounts of measurement error, Figures 1 and 2 are plotting *exactly the same probabilities.*

The curves in Figures 1 and 2 look different because the horizontal scales are different. To convert an *ICC* value from Figure 1 into a point on Steiner and MacKay's scale you would invert the *ICC*, subtract 1.000, and take the square root. This inverse, non-linear conversion changes the shape of the curves.

To show the correspondence between Figures 1 and 2, seven points have been placed on the corresponding curves. While these seven points are evenly spaced on the horizontal scale of Figure 2, they are not equally spaced on the horizontal scale of Figure 1. This difference in the horizontal scales creates the differently-shaped curves.

Once we remove this difference in choice of scales, we find that Steiner and MacKay have essentially verified the probabilities given in Figure 1. Moreover, by doing so, they have also validated the approach I used to transform the traditional power functions into the power

functions for different *ICC* values. Since independent confirmation is always welcome, I must thank Professors Steiner and MacKay for taking the time and trouble to verify and confirm my results.

Interpreting Figure 2 Steiner and MacKay write: "So, in our view, substantial measurement variation does adversely affect the performance of the *X* chart." While this vague statement is true, their use of the wrong power function curve undermines their position. To determine just how much measurement error affects performance with small shifts we need to use the detection rules that are designed and used to detect small shifts.

Intraclass Correlation Coefficient	Rule I Two Standard Error Shift	Rule I Three Standard Error Shift	Rules I, II, III, IV Two Standard Error Shift	Rules I, II, III, IV Three Standard Error Shift	Measurement Std. Dev. when Process Std. Dev. = 1.00
1.00	0.821	0.999	1.000	1.000	0.000
0.90	0.762	0.997	0.999	1.000	0.333
0.80	0.697	0.991	0.995	1.000	0.500
0.70	0.620	0.977	0.985	1.000	0.654
0.60	0.531	0.938	0.969	1.000	0.817
0.50	0.443	0.880	0.940	1.000	1.000
0.40	0.345	0.764	0.884	0.999	1.225
0.30	0.245	0.596	0.773	0.982	1.527
0.20	0.163	0.395	0.596	0.917	2.000
0.10	0.088	0.184	0.352	0.649	3.000
0.00	0.027	0.027	0.080	0.080	

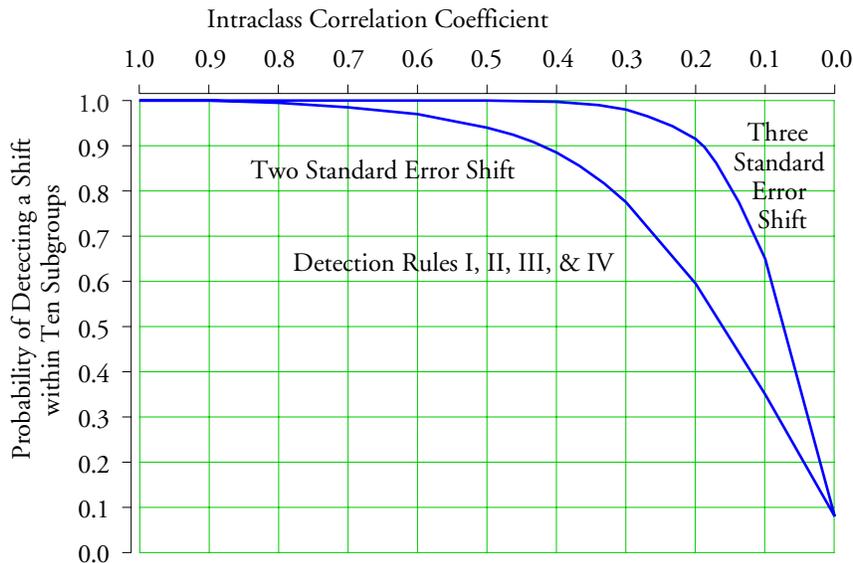**Figure 3: The Probabilities of Detecting Shifts Within *k* = 10 Subgroups**

Figure 4: Probabilities of Detection for Two and Three Standard Error Shifts

Figure 3 gives the probabilities of detecting two and three standard error shifts in location using all four of the Western Electric Zone tests. The first column gives the *ICC* values while the

last column gives the values for Steiner and MacKay's scale. Figure 4 shows these curves.

So what is the effect of using all four detection rules? Even when the *ICC* is 0.50, where measurement error is equal to the process variation, a process behavior chart will still have a 94 percent chance of detecting a sustained two standard error shift in location. This more than doubles the value of 0.44 given by Steiner and MacKay for Rule One alone. So, the "substantial error" here belongs to Steiner and MacKay. The rigorous assessment contained in Figure 4 completely undermines their interpretation of Figure 2.

THE PRACTICAL INTERPRETATION OF FIGURE ONE

Process behavior charts are much more robust to the presence of measurement error than Steiner and MacKay ever dreamed they would be. This robustness was the point of the example given in my articles. This example was not manufactured to "boost" my "contention that even if there is substantial variation due to the measurement system, a process performance chart can signal the actions of assignable causes." The example is real (although the data were coded for confidentiality). Also, the probabilities listed in Figure 3 and plotted in Figure 4 are not a "contention." They are established mathematical facts.
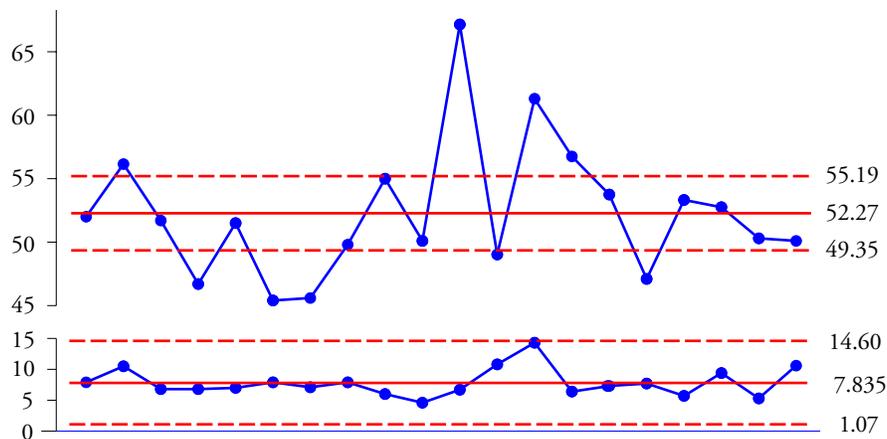

Figure 5: Average and Range Chart for Product 2131

When Steiner and MacKay finally get around to interpreting the average chart in Figure 5 they end up agreeing with me: "there are plenty of signals (we count 11 using the four Western Electric rules as on figure 1). So Wheeler is correct in his assertion that the performance chart can signal the action of an assignable cause even if the intra-class correlation is small. But as shown in figure 2, a noisy measurement system makes the detection of smaller shifts less likely."

In their disclaimer at the end of the passage above Steiner and MacKay say that Figure 2 shows something that it does *not* show. They just used all four detection rules in their analysis. Appealing to a graph based on Detection Rule One alone is more of a case of wishful thinking than an example of rigorous analysis.

At this point Steiner and MacKay have replicated the probabilities of Figure 1 and have validated my approach to calculating the effects of measurement error upon those probabilities. In addition they have agreed with my analysis of a real example. So where's the beef? Why have

they taken the time and trouble to write, revise, and publish their article? Evidently my articles contain something that disturbs them. If it is not the mathematics, which they verified, and if it is not the conclusions, with which they agreed, then it can only be the point of view.

TWO VERY DIFFERENT POINTS OF VIEW

Many statisticians have trouble understanding process behavior charts. Based on my own experience, I suspect that this comes from our mathematical backgrounds. We look at the technique and see opportunities to estimate several different parameters. My friends and colleagues have repeatedly focused on the problems of estimation involved in the computations rather than using the chart to characterize the process behavior. After all, when looking at a process behavior chart as a statistical procedure what else is there but a little estimation? The technique is simple, so let's work on how to sharpen up the estimates.

However, when process behavior charts are considered from the perspective of listening to, and learning from, the process, they become something quite different. It is not a matter of sharpening up the estimates but rather one of recognizing that a process behavior chart is a statistical axe that works whether the estimates are sharp or dull. In each of the following sections I will try to contrast these two very different viewpoints. Henry Neave and I have written more on these different in viewpoints in "Two Routes to Process Improvement" and "Shewhart and the Probability Approach."

INTERPRETING THE CHARTS

Qualitatively, the elephant in the room in Figure 5 is the unpredictable nature of the average chart. This process is screaming for help. Yet Steiner and MacKay take their time getting around to interpreting the average chart.

First they do something that is not part of any traditional statistical analysis procedure. They ignore the screaming process in Figure 5 and focus instead on the *XmR* chart (not given here) for repeated measurements of a standard to see if the measurement system is "stable."

Next they wander back to the range chart in Figure 5 and decide that the within-day variation also appears to be stable.

Then they return to the *XmR* chart to estimate the standard deviation for the measurement system. Following this they use the range chart to estimate the standard deviation for the production process. With these two estimates in hand they compute the intraclass correlation.

Finally, only after fussing over all the details above, do Steiner and MacKay turn to the screaming process shown on the average chart and discover 11 signals of a process change within 20 subgroups!

Contrast the approach above with simply starting with the fact that this process is very unpredictable. Once we have Figure 5 we do not need to worry about all the details above. All the king's computations and all the king's estimates cannot put this process together and make it predictable. Action is required: Assignable causes have to be identified and then made part of the set of control factors for this process. If this is done, all of the estimates above become moot. If this is not done the process will continue to change in unpredictable ways, and all of the estimates above will become moot.

On a personal note, I have used this example in my classes for twenty years, and I did not

know there were 11 signals of change on the average chart. I have never needed to count them. It is not about exactly how many signals we have. Rather it is about whether the process is behaving predictably or unpredictably. Thus, the overall interpretation is more qualitative than quantitative. When a substantial proportion of the averages are outside the limits, we no longer need to count points or invoke additional detection rules. The process is being operated unpredictably, and the only question of interest is "What can be done to fix it?"

BUT WE REALLY NEED GOOD MEASUREMENTS

Next Steiner and MacKay complain that we cannot know if the measurement system is adequate until we have an estimate of measurement error. They ask what would happen if someone "followed Wheeler's advice." From their viewpoint it would be a disaster: estimation is the first step in statistical analysis, and without estimates we have no knowledge. They imply that without the knowledge that we have a good measurement system we cannot make sense of our data.

Fortunately things are not so grim as Steiner and MacKay make them sound. The limits of a process behavior chart automatically take measurement error into account. (Since we know that Figure 5 has an intraclass correlation is 52 percent we can take the square root and invert it to discover that the limits are 39 percent wider than they would be with a perfect measurement system.) *However, we do not need to know this precise value in order to use the chart.* The inflation of the limits is automatic regardless of whether or not we estimate it. If we construct a process behavior chart and find signals on that chart, then we know that no matter how imperfect the measurement system may be, it is adequate for the job at hand, and we need to fish or cut bait rather than worry about the quality of our fishing line.

COMMENT ONE:

Finally, Steiner and MacKay make several additional comments that further reveal their viewpoint that estimation is paramount. In the first of these comments Steiner and MacKay leave out a key word in their summary of my position. I argue that it is not necessary to **start** with an assessment of the measurement system, not that such an assessment is "unnecessary." My first book was **Evaluating the Measurement Process** which came out in 1984. The current (third) edition has over 300 pages. So I have been helping people to assess their measurement processes for 33 years.

COMMENT TWO: PHASE I AND PHASE II

In comment two Steiner and MacKay further reveal their point of view when they write: "In phase I, we collect data to establish limits for phase II. We may omit some data if they correspond to instability. If we treat the June data [Figure 5] as phase I, it is difficult to decide which subgroups to delete since the process is so unstable. We cannot trust the centerline of the average chart and hence the control limits to apply in the future."

This perhaps is the most revealing passage in the article. The whole focus is upon creating a statistical model for the process by obtaining the best possible estimates of process location and process dispersion. Only then we can compute "good" limits for the charts that can be used

happily ever after. The objective here is the computation of the right limits. Many statisticians think this way. I remember a close colleague who argued that "you cannot compute limits for the average chart unless the range chart is predictable."

Contrast this "phase I and phase II" data manipulation approach with simply listening to the process. The running record shows the *process performance*. The limits, when calculated according to the usual formulas, approximate the *process potential*. So when the process performance does not fall within the lines defining the process potential, the process is not performing up to its full potential.   Here there are only two outcomes, the process is predictable and nothing needs to be done, or the process is unpredictable and it needs to be fixed. The calculations are simply a means to characterize the process behavior, not the end in themselves. When judging process behavior, exactness is not required—approximate will be close enough.
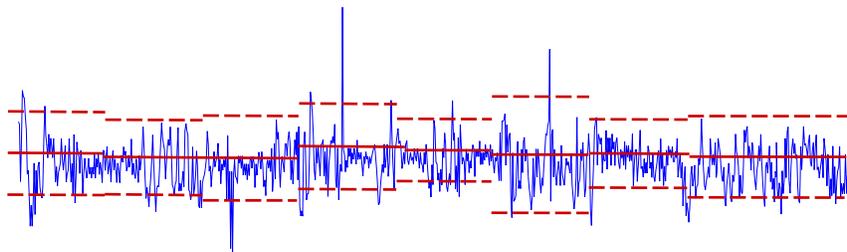

Figure 6: An *X*-Chart for 852 Individual Values

An example of this is provided by the *X* chart in Figure 6. They computed new limits every 100 values. These limits are all over the chart. Yet each set of limits tells the same story—this process is being operated unpredictably. Since they did not look for the assignable causes of this unpredictable behavior nothing happened here. Nevertheless, this chart shows how inexact limits based on an unpredictable process can still be good enough to tell you that the process is being operated unpredictably. Until action is taken nothing will change for the better. Worrying about how to estimate the process location and process dispersion for an unpredictable process is misplaced effort. As you learn how to operate your process more predictably, you will find that your estimates will automatically get better at the same time.

COMMENT THREE: CAPABILITY

In comment three Steiner and MacKay write: "In phase I, if we have a period of stability, we can assess the potential capability of the process…. Based on the capability, we may reconsider proceeding with the process behavior chart." Here we see the expression of an idea promoted by some statisticians that a process behavior chart is just a process adjustment mechanism intended to maintain the status quo for a process. This viewpoint looks at a process behavior chart as a manual form of "process-control." While a process behavior chart may be used this way, this is a far cry from what it was intended to be. (For more on this see "Process Monitor Charts.")

Contrast this idea of "maintaining the status quo" with the proven use of process behavior charts for continual improvement. When we find an assignable cause and make it part of the set of control factors for our process two things happen: The new control factor helps us operate the process on target; and the overall process variation is reduced. These reductions increase quality,

increase productivity, and improve competitive position. (For an example of this see "How Do you Get the Most Out of Any Process.")

COMMENT SEVEN: CONFUSION

Steiner and MacKay write "If the goal is continuous improvement of the process, then using a process behavior (i.e. control) chart may be an inefficient way to proceed." Given the baggage that they have attached to the technique, I can see how their version of SPC might be inefficient. It certainly will be less effective than listening to the voice of the process. When people "follow Wheeler's advice" and use process behavior charts as intended, they find them to be very efficient. I have clients all over the world who have found process behavior charts to be simple and effective tools for continual improvement.

Steiner and MacKay then go on to write: "Control charts play a minor role in the 'Holding the Gains' step of our algorithm." Once again they reveal their thinking that the process behavior chart is merely a process monitoring technique intended for maintaining the status quo. Just because they do not know how to use process behavior charts for continual improvement does not mean that you cannot do so.

But wait—there's more! In the very next paragraph Steiner and MacKay write: "We think that knowing the measurement system variability and some idea about its stability over time is valuable information to isolate assignable causes and improve the process."

So which is it? Do we use process behavior charts to "improve the process" or do they play a "minor role" as a process monitor? By their own testimony Steiner and MacKay do not know the answer to this question.

SUMMARY

Clearly Professors Steiner and MacKay *believe* that we need good measurement systems. This idea seems to be essential to their world view. But they can muster neither the mathematics nor the examples to prove this assertion in practice. The need for good measurements is simply taken as an article of faith. However, every day we all collect, use, and interpret imperfect data.

Fortunately, there is an alternative world view that works with imperfect data and which does have both theory and practice behind it. It is not the world view of the mathematical statisticians, but the practical approach of listening to the voice of the process that was championed by Shewhart. The process behavior chart is robust to problems of measurement error, and because of this it allows us to characterize our process behavior as the first step in knowing what to do next. As seen in the flowchart from "The Four Questions of Data Analysis" in Figure 7, this characterization of process behavior is the primary question of data analysis.
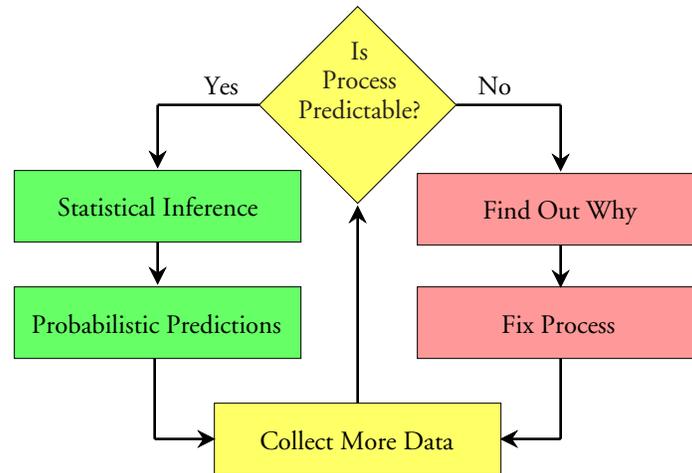
Figure 7: Predictability is the Primary Question of Data Analysis

If the process appears to be reasonably predictable then we can use statistical inference to estimate parameters and probability theory to make predictions. But if the process shows evidence of unpredictable operation none of the statistical computations make any sense. An unpredictable process is subject to the effects of one or more assignable causes, and action is required, first to identify these dominant cause-and-effect relationships, and then to make them part of the set of controlled inputs for the process. In either case, we collect more imperfect data and repeat the cycles above. This repeated cycle is how the process behavior chart becomes the locomotive for continual improvement. Simple. Easy. Proven.

REFERENCES

1. Steiner, S. H. & MacKay, R.J., "We Do Need Good Measurements," *Quality Digest 2017*, May 24.
2. Wheeler, D.J., "Don't We Need Good Measurements?" *Quality Digest 2017*, Feb. 6.
3. Wheeler, D.J., "The Intraclass Correlation Coefficient," *Quality Digest Daily 2010*, Dec. 2.
4. Wheeler, D.J., *EMP III: Evaluating the Measurement Process and Using Imperfect Data*, SPC Press, Knoxville, 2006.
5. Wheeler, D.J., "Detecting a Shift in Process Average: Tables of the Power Function for X-bar Charts," *Journal of Quality Technology, v. 15, no.4*, October, 1983, pp.155-169.
6. "Two Routes to Process Improvement," *Quality Digest Daily 2010,* May 5 and May 6, 2010.
7. Wheeler, D.J. & Neave, H.R., "Shewhart and the Probability Approach," *Quality Digest Daily 2015*, November 2.
8. Wheeler, D.J., "Process Monitor Charts," *Quality Digest Daily 2016*, July 11.
9. Wheeler, D.J., "How Do You Get the Most Out of Any Process?," *Quality Digest Daily 2016*, November 7.
10. Wheeler, D.J., "The Four Questions of Data Analysis," *Quality Digest Daily 2009,* Dec. 1, 2009.