Never assume that the underlying cause of error is appropriately labeled “human” or “operator.” Instead, look at the design of your systems and processes. This article provides examples of good and bad design and summarizes how to track quality using charts and graphs.

E. Deming’s research shows that far fewer than 20 percent—and more likely fewer than 10 percent—of errors have as their root cause a human element [1]. Rather, it’s system variation that drives the true error rate. That idea is encapsulated in a quotation attributed to but never uttered by Deming: “A bad system will defeat a good person every time.” A report from the US Naval Safety Center states that, “Human error has been implicated in 70 to 80 percent of all civil and military aviation accidents” [2]. Yet the report also notes that the real story is counterintuitive and complicated, and much of what we consider across many industries to be human error is not. Likewise, some of what we consider not to be human error truly is.

Don’t pick a red bead

Consider Deming’s well-known red-bead experiment. In this thought exercise, a bin contains 80 percent white beads and 20 percent red beads. The white beads represent a job well done, and the red beads represent defects or errors. Participants are asked to scoop out some beads and submit them for appraisal to the person designated as the manager. Clearly, each scoop of beads will likely contain “errors,” and that’s the point of the experiment: The number of errors each worker “commits” is beyond his/her control. It’s built into the system and does not measure actual performance. In this example, no matter how hard you work, with every scoop of say, 25 beads, you can expect about five red beads, and that rate will not decrease until the percentage of red beads is reduced or eliminated from the system.

Where fault lies

Are you skeptical about this take on human error? That’s understandable, because humans are indeed fallible. That’s why we must take preventive steps to insulate our systems...
and processes from them. In fact, not erecting barriers is—to paraphrase quality guru Shigeo Shingo—a management failure. Consider, for example, a hypothetical group of emergency room physicians who don’t have ready access to nitrile gloves, sleeves, and other protective gear and who become infected by the blood of a patient. Is it the physicians’ fault that their skin had pores and fissures that made infection possible? Of course not. In the same way, when managers don’t institute adequate measures to reduce the probability of human mistakes from contaminating our systems, they contribute to the propagation of errors. These “opportunities” for errors to occur shouldn’t be deemed “human errors,” but failures to mistake-proof systems and processes.

Look at the three errors shown in Figure 1. Setting up a machine incorrectly doesn’t necessarily implicate humans as the root cause. Rather, the machine was allowed to be set up incorrectly. Have you been to the airport lately? If so, you probably passed through a full-body scanner. Even if it was your first time through, you probably knew exactly how to stand because there was an outline of shoes painted on the floor. That allows travelers to understand quickly how to stand without further instruction. Compare the simplicity of that approach with the other inspection systems at airports, especially at the conveyor belt that transports carry-on luggage and personal items through a scanner. Almost without failure, a bottleneck forms as people are given verbal instructions about removing shoes and belts and placing liquids and laptop computers into the plastic bins.

Likewise in the second case: The error in record-keeping likely stemmed from a poorly designed document or user interface—one that didn’t apply the principles of human recognition sciences. If it was an electronic interface, it could have been programmed to disallow progress until an incorrect value was corrected.

The last error likely occurred because there wasn’t a system or standard for cleaning the line: how it must be done and measured before the next batch setup. Cardiac surgeons, for example, have a checklist of what to have ready in order to reduce complications and/or the probability of a failure. I advised a hospital on this topic, and much of the system was adopted directly from the aviation industry’s use of a pre-flight checklist [3].

### Quality control tools

Kaoru Ishikawa, the inventor of the “fishbone” diagram for problem-solving, said that the skillful use of seven quality control tools will resolve 95 percent of workplace problems [4]. Deming’s favorites among these tools were cause-and-effect (C&E) diagrams, Pareto charts, flow charts, histograms, run charts, and control charts. Here’s a quick review.

**C&E diagrams.** Root-cause analysis is really just a set of steps that demonstrate causality. In a C&E diagram, the problem is the head of the fish and the causal factors form the ribs (Figure 2a). For more detail, you can add probabilities to the ribs and thus begin to attribute causality to each element (Figure 2b). But look at everything carefully. In this case, under “machine” there is the causal

<table>
<thead>
<tr>
<th>Figure 1</th>
<th>Examples of error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine setup error</td>
<td>Record filled out wrong</td>
</tr>
</tbody>
</table>
Pareto charts. Named for Italian theoretician Vilfredo Pareto, Pareto charts show the degree to which different factors affect an outcome. They do so using a bar chart with a line plot superimposed on the bars that corresponds to a second y-axis to show what percentage each element contributes to the total set of variables. For example, of 133 people surveyed about their medication compliance, say 47 reported non-compliance due to unclear dosing instructions. To show that, the first bar would show a count of 47 respondents, and the line plot at that point would correspond to 35 percent on the second y-axis ($133 \div 47 = 35.3\%$)

does that stem from a lapse in an automatic cleaning cycle (equipment-related error) or from a cleaning protocol that was not written correctly and clearly (method-related error)? Or perhaps there was no precise definition of what acceptably "clean" is (measurement-related error). But in making the chart, don't devote too much time to such questions. It's more important to understand your process and to recognize what is working correctly and incorrectly (defects). During the initial search for causal factors, just list as many potential causes of the problem as you can.

**Figure 2**

*Cause-and-effect ("fishbone") diagrams*

a. The “head” is the problem, and the “ribs” are causal factors

b. Assigning probabilities to causal factors indicates where to focus corrective action

**Figure 3**

*Pareto chart*

<table>
<thead>
<tr>
<th>Reasons for refusal</th>
<th>Number of respondents</th>
<th>Percentage</th>
<th>Cumulative percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dosing instructions unclear</td>
<td>47</td>
<td>35.3</td>
<td>35.3</td>
</tr>
<tr>
<td>Tablet too large</td>
<td>32</td>
<td>24.1</td>
<td>59.4</td>
</tr>
<tr>
<td>Unclear benefits</td>
<td>21</td>
<td>15.8</td>
<td>75.2</td>
</tr>
<tr>
<td>Ordering/delivery too complex</td>
<td>12</td>
<td>9.0</td>
<td>84.2</td>
</tr>
<tr>
<td>Side effects</td>
<td>8</td>
<td>6.0</td>
<td>90.2</td>
</tr>
<tr>
<td>Too many other meds</td>
<td>5</td>
<td>3.8</td>
<td>94.0</td>
</tr>
<tr>
<td>Forget</td>
<td>3</td>
<td>2.3</td>
<td>96.2</td>
</tr>
<tr>
<td>Lost prescription</td>
<td>2</td>
<td>1.5</td>
<td>97.7</td>
</tr>
<tr>
<td>Didn't fill prescription</td>
<td>2</td>
<td>1.5</td>
<td>99.2</td>
</tr>
<tr>
<td>Didn't want to buy too expensive</td>
<td>1</td>
<td>0.8</td>
<td>100.0</td>
</tr>
</tbody>
</table>
percent), as Figure 3 shows. Many Pareto charts show only the few factors that account for the bulk of the effects in order to separate the "vital few" from the "trivial many," as Joseph Juran said [5]. Some call it the "80-20 rule."

**Flow charts.** These are useful in showing how a process should run. Note that I didn't say how it actually runs, although knowing both how it should and how it does run are important to understanding what could (or did) go wrong. "If you can't describe what you are doing as a process, you don't know what you are doing," Deming said. Figure 4a shows a flow chart of what is supposed to happen, and Figure 4b shows what happens in practice. Use flow charts like these to find differences and fix them.

**Histograms.** These are powerful tools that enable you to visualize the overall shape of the distribution of your data. The x-axis is ordered into so-called bins, and the y-axis tallies the observations that fall into each bin (Figure 5a). There should be no gaps between the bars of your plot, unless you intend it to indicate a zero value. Figure 5b shows Microsoft's example of a histogram but, because of the gaps, it's a bad example [6]. Also, there is no definitive answer to how many bins a histogram should include, but a statistical rule of thumb is to either divide the total number of observations by your data range (e.g., 500 observations with a data range of 30 units = 500 ÷ 30 ≈ 17 bins) or to use the square root of the number of observations and round up (i.e., \(\sqrt{500} ≈ 23\) bins).

Figure 6 shows a histogram of weights of tablets taken from two different processing lines over 10 days. Because the analysis is bi-modal, I'd suggest looking at data using a stratified histogram of each processing line to see what each contributes to the total output in tablet weights (figures 7a and 7b).

**Run charts.** These are line plots (fever charts) that show the change to a variable over time, with time plotted along the x-axis and progressing left to right. Humans are adept at noticing deviations from the horizontal, and
run charts help us identify positive and negative deflections in the data as shown in Figure 8. In this case, I'd probably look at months of results to see whether the peaks in tablet defects relate to other systemic issues.

Control charts. No self-respecting data analyst would be comfortable assessing a process without a control chart. Developed by Walter Shewhart, control charts take many different forms depending on the type of data you wish to analyze and to what effects size you expect to find [7]. Generally, a control chart's data are ordered by time, must include a center line (typically the mean of the data series), and upper and lower control limits (Figure 9). What sets control charts apart from other charts is that the control limits are calculated based on
historical data. This means that your data set is the foundation for the width of the control limits. Once those are established, you look for any departures from the so-called “run rules.” Run rules—commonly the Western Electric or Nelson rules—help you distinguish “common-cause” variation from “special-cause” variation. The difference is critical because common-cause variation determines your process capability: whether it can do what it’s designed to do.

While there is some overlap in what each of these charts expresses, they all tell a different part of the whole story. For example, the data in Figure 10 about tablet weights may look pretty good to you, depending on your process. It shows that the mean tablet weight is 135.8 milligrams; the confidence level is 95 percent; and the standard deviation is 2.37.

Now look at the same data in a run chart that shows the tablet weights over time (Figure 11), and the story changes: Weights have been creeping up. What else might be happening? Are the dosages getting more variable? Have the densities changed? What are the potential effects on safety, identity, strength, purity, and quality?

To divine the cause of unwanted variation, begin with the presumption that willful ignorance or neglect is rare. That means taking a sober approach to systems design, one that recognizes that most errors at the interface of people and equipment likely stem from a lack of understanding or clarity.

Sometimes it’s a failure to measure what’s important or even obvious. At Stanford University’s hospitals and clinics, for example, wait times in the emergency room decreased once managers started measuring and reporting them [8].

\[
\begin{align*}
\text{UCL} &= 22.01 \\
\text{Avg} &= 20.40 \\
\text{LCL} &= 18.78
\end{align*}
\]

\[
\begin{align*}
\text{Mean} &= 135.81 \\
\text{Standard deviation} &= 2.37 \\
\text{Variance} &= 5.63 \\
\text{Skewness} &= 0.6222829 \\
\text{Kurtosis} &= -0.027441 \\
\text{N} &= 33
\end{align*}
\]

\[
\begin{align*}
\text{Minimum} &= 132.00 \\
\text{1st quartile} &= 134.00 \\
\text{Median} &= 135.00 \\
\text{3rd quartile} &= 137.00 \\
\text{Maximum} &= 141.50
\end{align*}
\]

\[
\begin{align*}
\text{95\% confidence interval for mean} &= 134.97 \, \text{to} \, 136.65 \\
\text{95\% confidence interval for median} &= 134.45 \, \text{to} \, 136.90 \\
\text{95\% confidence interval for standard deviation} &= 1.91 \, \text{to} \, 3.14
\end{align*}
\]
Such visible accountability mirrors what is known as the Hawthorne effect: Performance improves solely because the subject knows that he or she is being studied [9]. In addition, reporting wait times should be a sustainable improvement step because what’s measured (incoming outpatients) is different nearly every time. Thus it introduces what’s called the novelty effect, in which performance tends to improve because the arrival of something new increases interest; it also works because the rolling measurement of the data prompts the staff to act on at least one component of best practice: expeditious treatment.

Implementing similar systems requires more than just taking measurements. Staff members need an explanation of why they do what they do. Indeed, when people understand the importance of their work elements and how their work contributes to the success or failure of the process, they perform better. In 21 CFR 211.25, it states that “Each person engaged in the manufacture, processing, packing or holding of a drug product shall have education, training and experience, or any combination thereof, to enable that person to perform the assigned functions.” That only makes sense, and the more that we succeed in delivering the right training and tools at the right place and time, the more likely employees are to notice or prevent defects before they manifest. As a result, the process will generate fewer errors and produce better outcomes.

In advising leaders about how to apply different training evaluation measures, I stress the importance of creating an appropriate feedback loop because that is critical to achieving positive outcomes [10]. On-the-job education is never static and determining what will reduce defect probabilities the most will help you design the right systems. From the start, reassess the health of your systems to ensure they are designed to deliver what you intend, that they address the few mission-critical aspects and are not caught up in addressing the many trivial factors that can obscure what’s truly important.

Root causes

Not long ago, I was talking with an engineering professor and asked how his students learned about root cause analysis. He said it’s not addressed much in his courses. I was disappointed because it’s an essential strategy. But even more troubling, was the professor’s own misunderstanding of root causes, which came to light when we talked about the 2010 explosion of the Deepwater Horizon oil rig that killed 11 people and polluted the Gulf of Mexico.

Many causal factors contributed to that disaster, but two principal factors were the rig’s use of a single blowout preventer, a mechanism that is supposed to seal the well’s pipe and stop a leak. In this case, the device failed and that’s why most oil rigs now include at least two blowout preventers connected in series. I’d wager that the Deepwater Horizon used only one because there was little real-world evidence that more were needed. In fact, the CEO of the oil company that contracted the oil rig testified to Congress that the likelihood of failure was “about one in a hundred thousand to one in a million” [11]. But a subsequent report pegged the failure rate at 45 percent [12]. Needless to say, that’s an enormous difference.
In any case, as we talked, the engineering professor said the root causes of the disaster were well known. They included, he said, operators who missed many telltale signs, such as the presence of seal elastomer in the drill mud effluent. But that is not a root cause. Nor is failure to report seeing the seal debris a root cause. Those are only symptoms of deficient systems. Perhaps the seal material wasn’t suited to the task; maybe preventive maintenance was insufficient; and maybe the rig lacked contingency blowout preventers. Acknowledging those factors will get you closer to determining the root cause(s), even though each may not itself be a root cause. But there is one factor that we can rule out as a root cause: the rig’s operators.

References


Ben Locwin, PhD, MBA, is president of Healthcare Science Advisors, 959 Maplewood Ave., Suite 2, Portsmouth, NH 03801. Tel. 603 397 7304. Website: www.healthcarescienceadvisors.com. He provides expertise in quality control and employee training and behavior to the pharmaceutical and food and nutrition industries, as well as to people who work in the psychological and academic fields. He last wrote for Tablets & Capsules’ October 2013 issue.