FEATURE
STATISTICS

RELIABILITY

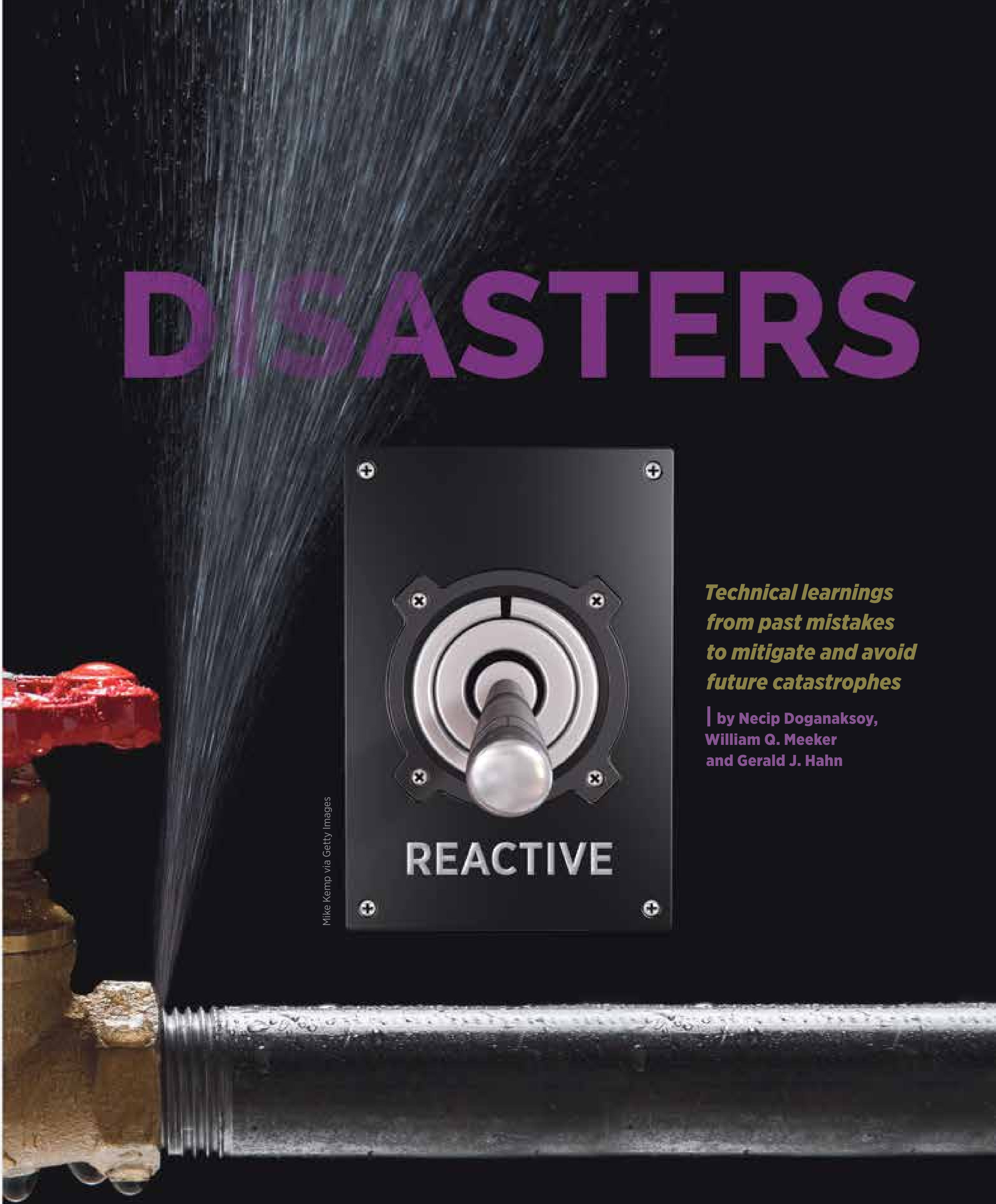
Just the Facts

When products fail in the field, disasters can result. To head off problems, manufacturers must build reliability into the design of products and processes.

Statistics can be used proactively to help improve reliability during product design and development, and enable manufacturers to "do it right the first time."

The authors describe some technical and statistical problems from four specific reliability disasters to highlight lessons learned.

DISASTERS



Mike Kemp via Getty Images

Technical learnings from past mistakes to mitigate and avoid future catastrophes

by Necip Doganaksoy, William Q. Meeker and Gerald J. Hahn

Reliability evaluations present a challenge for manufacturers and consumers because there is an elapsed time between when the product is designed and built, and when the reliability information is forthcoming.

For instance, the manufacturer of a newly designed car might have only three years to demonstrate its reliability before it is brought to market. But you might want to know, for example, whether—if the car is properly cared for—it will provide trouble-free service for the next 12 years or if it is going to lead to mounting problems after just a few years.

The desire for high reliability is not new. Throughout history, people learned from their successes and mistakes, for example, in building durable wheels, larger domes, stronger ships and longer bridges. The learnings from disasters, such as the Titanic sinking, the de Havilland DH 106 aircraft crash and the Tacoma Narrows Bridge collapse, can help achieve improved reliability and safety of subsequent generations of products.[1]

In the past, reliability assurance was often an afterthought—even in organizations that emphasized quality. This is sometimes referred to as the "design, build, test and fix" approach. Using this approach meant that manufacturers discovered and addressed reliability issues through extensive testing after the product already had been designed.

Sometimes, the time allotted for such testing did not allow all important failure modes to be identified and fixed, and the product was released before its reliability was fully validated. There was a heavy reliance on end-of-line product testing and fixing problems in the field after they occurred. As a result, much effort was spent responding to crises and fixing problems after they already had created damage to the customer and the manufacturer's reputation.

In large part in response to increased competition in the world markets, starting about 30 years ago, forward-looking business and industry leaders have realized that achieving reliability by reactive measures is unacceptably expensive and potentially disastrous to retaining customer confidence. If reliability problems arise after a product has been released for production and, especially if units in the field must be recalled for retrofitting, costs can be severe and may rapidly dwarf a product's profit margin.

There is now general agreement that reliability must be built into the design of products and processes proactively. Problems discovered in design—though often more difficult to identify, are usually less costly and much easier to fix. Problems found after the design has been frozen—and especially after significant quantities have been built, although easy to identify, are often difficult and expensive to fix.

As a result, the traditional design, build, test and fix approach has been replaced—at least in the minds of most reliability practitioners—by a "do it right the first time" mindset or proactive reliability assurance.

Focus on proactive reliability assurance has led to using statistics to help improve reliability during product design and development. This requires quantitative methods for predicting and assessing product reliability, and for providing early information on failure causes, as well as—and perhaps most importantly—careful planning to ensure the most meaningful information for analysis is obtained.

Many companies, indeed, have incorporated proactive approaches and methods into their reliability assurance efforts. Despite all of this, however, reliability disasters still occur, as evidenced by the frequent media reports concerning reliability problems affecting such products as laptop computers, electronic tablets, automobiles and batteries. Recently, the Boeing 737 Max airplane disaster has made the headlines.

So what are we still doing wrong?

The problem is surely not with the concept of proactive reliability assurance, but in the manner in which it has been applied, or, in some cases, not applied. What can we learn from various past reliability disasters to avoid similar ones in the future?

The past shortcomings have been managerial and technical. The former, such as the Boeing 737 Max reliability disaster, once uncovered, are typically reported in the mainstream press.[2] Reliability disasters frequently are related to problems in communication, especially to management, and are often brought on by pressure to expedite product launch—even if the needed information for reliability assurance is not forthcoming. In contrast, technical deficiencies that contributed to reliability disasters, such as inadequate testing, typically receive much less press.

This article focuses on such technical (and statistical) problems and highlights key technical lessons learned from four specific reliability disasters—as gathered from the publicly available sources referenced. Although some examples deal with events that preceded the acceptance of proactive reliability assurance, the basic lesson learned still applies. In particular, the focus will be on the bulwark of modern proactive reliability assurance, the accelerated life test (ALT) and the ALT process.

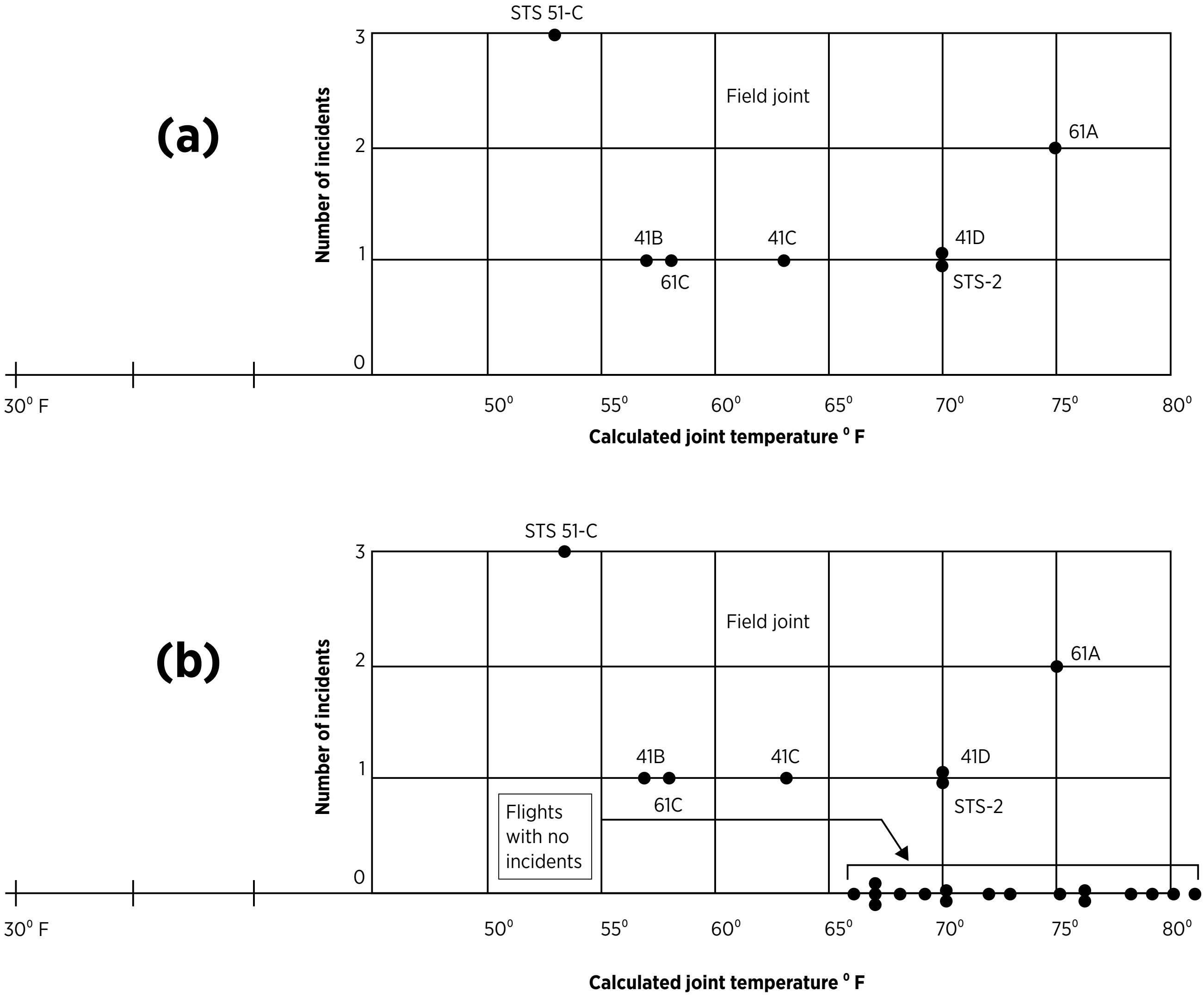
Accelerated life testing

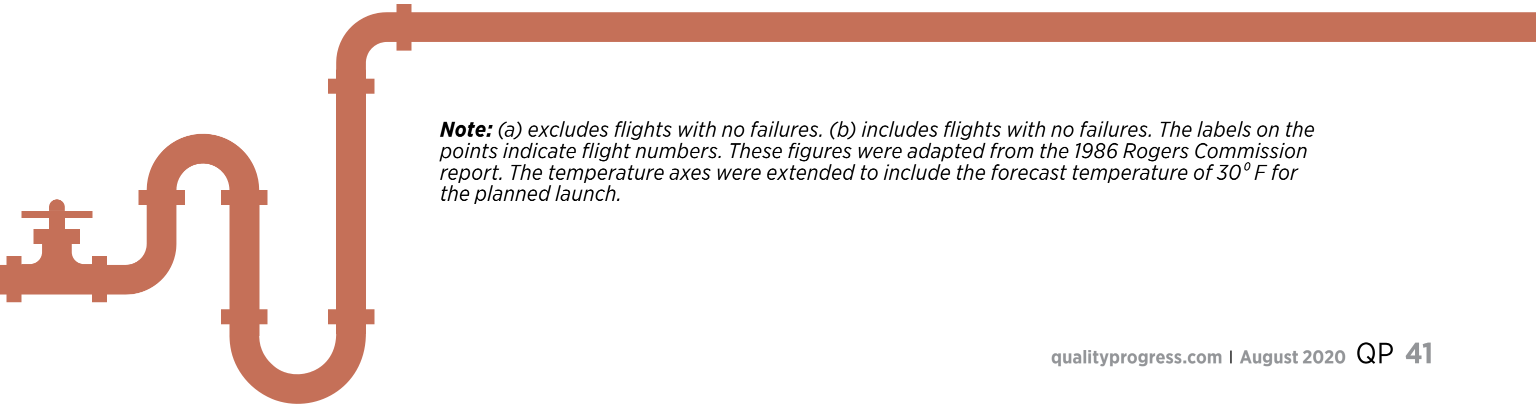
Statistically planned investigations often can be used during product design to assess the reliability that can be expected

iStock.com/sumkinn

FIGURE 1

Plots of number of field joint O-ring failures per flight vs. launch temperature on previous flights





Note: *(a) excludes flights with no failures. (b) includes flights with no failures. The labels on the points indicate flight numbers. These figures were adapted from the 1986 Rogers Commission report. The temperature axes were extended to include the forecast temperature of 30° F for the planned launch.*

for components, assemblies, subsystems and, eventually, the final product or system when there is insufficient time to make such evaluations at normal operating conditions. An ALT involves running sample product under carefully chosen, more severe environments, such as higher temperature or humidity, than would be experienced under normal field operating conditions. It also may involve exposing test units to increased stress, such as higher voltage or pressure. Such variables accelerate the physical and chemical degradation processes that cause certain failure modes and result in failures sooner than in normal operation.

When there is a well-understood failure mode (or modes), a carefully planned ALT often can provide useful information about how that mode will affect product life in the field. In particular, a physically appropriate model that relates stress to lifetime is fitted statistically to the data from the ALT and used to extrapolate to estimate expected lifetime under the conditions encountered in normal operation.[3]

ALTs have been used for a long time. They have achieved a particularly prominent role with the emergence of proactive reliability assurance. They can, however, also lead to incorrect conclusions if not used with great care.

Identify key failure modes ASAP

The AT&T round cell (a.k.a. "Bell Cell") was a radical new design of a lead-acid battery developed at AT&T Bell Laboratories in the late 1960s and deployed in the field in the early 1970s.[4] The round cell was designed as part of a backup power source for uninterruptible power supply applications. Plate growth is the well-known life-limiting failure mode of a lead-acid battery cell.

An ALT used elevated temperature to increase the rate of plate growth and predicted extremely long lifetimes (hundreds of years) for that part of the system.[5] Another ALT for the same cell used elevated voltage to accelerate a known corrosion mechanism associated with the post seal and predicted the post seal would last at least 40 years in service.[6] Encouraged by these findings, hundreds of thousands of units were installed in the field during the 1970s.

Within a few years after the product was introduced into the field, however, a serious blister corrosion failure mode arose and caused many installed cells to fail. Subsequent investigations revealed that the root cause for these failures was incompatibility between an epoxy used in a seal of the positive post and the lead-acid chemistry. The epoxy was redesigned for future production, but a large proportion of the previously installed cells needed to be replaced.

The blister corrosion phenomenon was not observed in the ALT because the elevated voltage actually inhibited this failure mode.[7] Blister corrosion was, in fact, an unanticipated failure mode. As a result, it had not been properly excited in the ALT.

The AT&T round cell thus demonstrates it is crucial that all relevant failure modes are identified early in the design process and thoroughly studied by an appropriately planned ALT.

Obtain and report all relevant data

The introduction of rotary compressors in General Electric (GE) refrigerators provides an enlightening example in which highly relevant additional findings from an ALT, although recorded, were not given the attention they merited.[8]

GE Appliances was losing market share to competitors.[9] The corporation was under intense pressure to bring new products to market to turn business around.

To meet the challenge, GE engineers proposed building a new refrigerator with a rotary compressor to replace the reciprocating compressor in use at that time. Rotary compressors had been used successfully in air conditioners but not in refrigerators. The new refrigerator would have higher efficiency and lower cost than the current refrigerators.

A sample of 600 rotary compressors was run continuously in an ALT at various elevated temperatures. None of the units failed after one year of testing. Thus, it was decided to proceed to launch the product. By 1987, there were more than a million units in service. The first compressor failures occurred after 1.5 years in field service, and many more failures took place shortly thereafter.

It soon became evident that all of the refrigerators that had been sold would have premature compressor failures. As a result, GE replaced the compressors in all refrigerators that it could locate at an estimated total cost of more than $450 million.

What is the explanation for the inconsistency in the results from the ALT and the actual field performance? The responsible engineers took the prudent step of disassembling some of the units that had not yet failed from the ALT.[10] In so doing, they uncovered multiple early signs of problems with the new compressors, such as unexpected discoloration, providing evidence of a lubrication issue and suggesting these units were well on their way to failure.

iStock.com/posteriori

However, this information did not seem to have been properly communicated through the management chain—perhaps at least in part—because of the pressure felt throughout the business to adhere to a tight schedule for product launch.

Systems must be in place to carefully monitor field data by appropriate segments for timely detection of potential issues that may affect a subset of the product population.

Ensure proper data analysis

The Challenger Space Shuttle was scheduled for launch early Jan. 28, 1986. According to the Rogers Commission that subsequently investigated the failed launch, on the evening of January 27, engineers and managers from NASA and NASA's contractor, Morton Thiokol, discussed during a three-hour teleconference the risk posed to the next morning's planned launch.[11] The temperature at launch was predicted to be 30° F. This low of a temperature was believed by some to increase the risk of failure of the O-rings that were used in critical joints of the solid rocket motor during launch. Most of the 24 previous launches had been held at temperatures between 65° F and 77° F, with the lowest at 54° F.

According to the Rogers report, based on analysis of the data that was discussed in the meeting on O-ring failures during past launches (see subsequent discussion), NASA management concluded that the probability of launch failure was about one in 100,000 (engineering estimates were one in 100). Based on this evaluation, and despite a strong recommendation not to launch from some of its engineers, Morton Thiokol acquiesced, and the launch proceeded the next morning.

A subsequent review, described in the Roger Commission report, determined that the prelaunch analysis of the available O-ring data, as presented to management, was inadequate and erroneous. There was only a hand-written list of the dates of O-ring failures, the number of such failures and the temperature. No plots of the data were presented.

Most importantly, the analysis ignored information from the 18 previous launches for which there were no O-ring failures. In particular, analyses omitting this information provided no clear evidence of a relationship between temperature and O-ring failure. However, when the data on the launches without any O-ring failures are correctly included in the analysis, a strong association between temperature and O-ring failure probability was evident. See Figure 1 (p. 41), adapted from plots in the Rogers Commission report.

In particular, Figure 1 (a) is a plot of the number of failures per flight versus launch temperature for previous launches excluding the flights with no failures. This limited data plot shows no clear evidence of a relationship between temperature and O-ring failure. Figure 1 (b) shows all the data, including the flights with no failures. This plot suggests a strong association between temperature and number of O-ring failures, with low temperatures being particularly risky.

There were two O-rings at each of six field joints in the space shuttle's solid rocket motors. If only one of these O-rings failed at any location, it would not be a problem. If both failed at one location, it would be catastrophic. In the Jan. 28, 1986 Challenger launch, both O-rings failed in one of the field joints. A subsequent careful statistical analysis of all the available data,[12] taking the preceding system structure into consideration, estimated the risk of failure for a 31° F launch to have been at least one in eight.

The Challenger Space Shuttle disaster was, at least in part, attributable to obtaining insufficient data and not paying enough attention to the data that were available. It also illustrated the usefulness of appropriate plots of the data and the fact that statistical analyses of reliability data typically call for the use of advanced statistical methods that generally are not taught in introductory statistics courses.

Finally, this application highlights the importance of the skills engineers and statisticians must possess to communicate complex technical matters to upper-level management.

Conduct up-front experimentation

The background for this example comes from the U.S. National Highway Traffic Safety Administration (NHTSA)[13] report addressing this disaster. In the late 1990s, reports of tire tread separation of Firestone tires on Ford Explorer SUVs started to arise and rapidly multiplied. This was of major concern to Ford and Firestone because tread and belt separation (TBS) failures at high speeds often cause vehicle roll-over accidents, resulting in injuries and fatalities.

In August 2000, 14.4 million potentially vulnerable Ford Explorer Firestone tires were recalled. A team of Ford and Firestone engineers and statisticians was commissioned to scrupulously investigate the problem and recommend action to ensure its elimination in future tires.

A combination of field failure data and physical understanding of TBS failures led the investigatory team to conjecture that the failures were affected by such factors as tire inflation pressure, radial load, ambient air temperature, the adhesion strength between belts and the thickness of the rubber wedge between a tire's belts.

However, unequivocal conclusions could not be drawn from the available observational data. Thus, Ford conducted a large experiment to study the effect of these and other factors. The experimenters were able to reproduce the field failures in the lab, estimate the effects of the different experimental factors and identify the root causes: The Firestone tires produced at several of their plants had—as a result of a change in specifications—a narrower inter-belt gauge and less adhesion strength than tires from other manufacturers.

According to the NHTSA report, this knowledge resulted in Firestone resetting its manufacturing specifications to what they had been originally to avoid such failures in the future. Moreover, a key lesson learned from this experience is the importance of conducting probing up-front experimentation to understand the effect of proposed design changes—and, for that matter, original design factors—on product reliability and performance.

If such experimentation had been conducted in this application before relaxing specifications (instead of after the problems had been discovered), the large direct and indirect associated costs would, most likely, have been identified and avoided.

Further key points to remember

The preceding examples illustrate only a few key points to remember in proactive reliability assurance. But there are other major considerations and common pitfalls of ALTs to keep in mind.[14,15]

Strive to obtain and use degradation data. Meaningful product degradation data can provide important clues into failure mechanisms even when few—and sometimes even no—failures are observed in product life testing.[16,17]

Search for the best possible physical model. Many reliability problems are caused by unanticipated failure modes or known failure modes that are accelerated in an ALT by unanticipated or poorly understood environmental conditions. Thus, statistical analyses of reliability data must be well-grounded on physical considerations.

Use different credible approaches for analysis and compare the results. A useful practice is to analyze the same data under different assumptions and credible models, and perhaps using different methods. This is particularly relevant for new technologies lacking well-established reliability models.

Beware of too much extrapolation. It is nearly impossible to avoid some extrapolation in making decisions leading to a new product launch. Therefore, manufacturers must continue to collect and monitor reliability data through in-house testing and exposure of early production units to the harshest field conditions.

Carefully track field reliability data. The field provides the ultimate testing ground and yields the most realistic, though not the timeliest, information about reliability and product performance. The worst situation is when a customer is the first to bring a problem to a manufacturer's attention.

Systems must be in place to carefully monitor field data by appropriate segments (for example, manufacturing period, product age, use conditions and component vendor) for timely (especially important) detection of potential issues that may affect a subset of the product population.

EDITOR'S NOTE

The references listed in this article can be found on the article's webpage at qualityprogress.com.

Necip Doganaksoy is the Douglas T. Hickey chair in business and associate professor at Siena College School of Business in Loudonville, NY, following a 26-year career in industry, mostly at General Electric (GE). He has a doctorate in administrative and engineering systems from Union College in Schenectady, NY. Doganaksoy is a fellow of ASQ and the American Statistical Association.

William Q. Meeker is professor of statistics and distinguished professor of liberal arts and sciences at Iowa State University in Ames. He has a doctorate in administrative and engineering systems from Union College. Meeker is a fellow of ASQ and the American Statistical Association.

Gerald J. Hahn is a retired manager of statistics at the GE Global Research Center in Schenectady. He has a doctorate in statistics and operations research from Rensselaer Polytechnic Institute in Troy, NY. Hahn is a fellow of ASQ and the American Statistical Association.

iStock.com/posteriori