Some Tribulations Of Testing And Truthfulness

Alan Nicol, Executive Member, AlanNicolSolutions

For many industries it is imperative to test and prove that our designs are good before launching into production. The very idea of proof, however, carries risk, which we must prepare our business to assume.

Most of us do some form of testing or other validation of our designs and production systems before we initiate production of our products. For some of us, especially those of us who produce products related to safety, we must prove to a regulatory agency that our products are safe and meet regulations.

I've yet to witness an environment where testing of products is not a delicate balance. Even in high-volume production environments that produce low cost products, a 5 percent error in defect rates or the time delay in production due to testing can be very expensive. Similarly, testing a helicopter isn't cheap or easy.

Testing becomes a gamble for almost all of us. When your business takes that gamble, does it also plan contingencies or insurance in the event that the gamble doesn't prove out? Let me give some examples of what I mean.

Looking at the transportation industry for one of my favorite conundrums for discussion,iIt can be very expensive to test a fully functional automobile. It is vastly more expensive and difficult to test a passenger aircraft or a train. So, how do we prove or otherwise validate that a design is good, and how do we prove that it is safe? We test it, of course.

Here's the challenge. How many should we test? Well, we can't afford to test a statistically significant sample of 30 or more. In many cases we test only one sample for each required test.

In the transportation industry, many of the design firms have sophisticated computer models that are used to do their designs and predict their performance and safety. The test of one sample is more useful to verify the model than it is to prove the design by itself. Demonstrating that the model accurately predicts the test results allows us to have greater confidence that the evidence of performance predicted by the model is more-or-less truthful.

Unfortunately, sophisticated models, particularly of dynamic scenarios such as vehicle crashes, are very expensive to develop. They require an enormous amount of data and huge investment in time and expertise to get them to reasonably predict an outcome. Some of us don't have those things, but our products are just as cost prohibitive to test, at least for our own businesses. What do we do then?

Often the temptation is, again, to test one and, if we pass, we go into production. In many businesses that must meet industry safety or regulatory minimums there is a

bit of a tiger trap set out for us to fall into. We have the option of developing the skills and expertise of self-certifying. We also have the option of developing our own test facilities and protocols. The regulators then simply require that we take responsibility upon ourselves to produce a product that meets regulatory minimums.

It’s really a very reasonable and economic system. This way the regulatory agencies don’t need excessive funds or personnel to police our every decision and we don’t have them camped out in our design centers telling us how to do design or otherwise ordering our business. The trap is, if we make a mistake, we now have only ourselves with whom to share blame and the regulatory agency is free to exact vengeance on behalf of the industry and the public. Our customers can either be drawn into the trap because they trusted us, or part of the vengeance force, or both, depending on the relationship.

It’s not as adversarial as I make it sound. In my experience the various regulatory agencies are reasonably compassionate about our needs to do business, and are very helpful in terms of interpreting the regulations, even if they don’t give an inch in those interpretations. Such is their prerogative.

So, let’s say we are one of these businesses and our method to prove our design of very expensive-to-test equipment is to test one sample. Let’s look at that scenario.

The Tribulation

We have to have some sort of proof that our product is good, but we don’t have the resources, data, or expertise to possess the sophisticated models. We also don’t have the resources to test a large number of our product systems.

The Temptation

The obvious solution is to test one sample. If it fails, then we know we have some more work to do, and it will cost us both time and money to re-design and re-test. If it passes, we go into production. To avoid the former and ensure the ladder, instead of sophisticated models, our best bet is, to carefully and deliberately design as much safety factor as our performance requirements and cost targets will allow.

The Gamble

If we test one and it fails we know we have a poor design. If we test one and it passes, we really don’t know anything except that we might have a good design. We don’t even know if we were to test a second one if it would pass. All we know is that one sample passed.

I’ll caveat the above statement by acknowledging that it is best suited to scenarios where destructive testing yields a simple pass/fail data point. In cases where we can measure some form of stress or other input or output and it significantly exceeds the requirement, we can speak with some more confidence, but still, with only one data point, we don’t really know. The next one could go a little more, or it could go

a lot less.

The Truthfulness Problem

We know that one sample passed. We don't know for certain that a second one would. The best thing we can do to ensure the integrity of our claim that our product is good is to never, ever, test it again. If a second test takes place and it fails, we now have a business problem of addressing our mistake for certifying a product that should not have been so (possibly losing our self-certification privilege) and also dealing with product in the field that might be unsafe.

The Solution Problem

Unfortunately, the various solutions to the whole scenario all cost money. So, we either invest our money to avoid the possibility of going out of business because of a mistake, or we gamble that such a discovery will never be made. If we stay in business long enough, eventually the gamble will bite us. We must be prepared.

The right solution for one business might be right for another, but there are options. Let's discuss a few, briefly.

Perhaps the least popular option is to estimate how much it would cost our business to correct an issue where we discover that our product line in the field must be recalled, and then either pay an insurance company or carry a cash reserve to cover it. A few businesses will do this. For some environments, such as the automotive industry, this strategy is unavoidable, even with robust testing methods. It's not easy though to keep that cash reserve. For smaller businesses it may not be realistic.

The most popular answer is the one I'm trying to advise us against. That is to do nothing and to pray that a problem is never discovered, or that if it is, our business is healthy enough to take the lumps. Please, for the sake of your business, your customers, and your industry in general, don't do this.

If your business can't insure itself against the eventual problem, or if safety is a significant concern and you aren't willing to risk someone's life on uncertainty, then I offer a solution that is often difficult to swallow at first, but in the long run is likely to pay off in multiple ways. Actually, it's a combined strategy.

The first part of the strategy is to endeavor to replace what mathematicians and scientists call "discrete" data with "continuous" data. Discrete data is unquantifiable information such as a pass or fail result. Continuous data is measurable, such as pounds-force, degrees Celsius, or hours of operation.

If you a destructively testing your systems to a pass/fail result, take a small bit of time and investment and apply sensors or equipment to your system and collect data during the test. For example, we can apply load cells to specific attachment points, accelerometers to monitor vibration and modes of deflection, temperature sensors, or any number of other devices to give us measurable readings of our

systems' stresses or responses during the test. There are three advantages to doing so.

The first is that we can begin to construct a sense of what actually happens during the test, and how close or far from minimum performance requirements our system is. This can increase our confidence in our product's ability to meet regulations or requirements, to some degree. If we are well within our performance envelope all of our products are more likely to be inside than if our test sample is on the edge. We still don't really know for sure where they will all fall.

The second advantage to continuous data is that it enables us compare actual performance to our engineering predictions based on analysis and calculation. Doing so can increase our confidence that reality matches our design intent and we can be a little more confident. It also, if we are willing to invest the time and expertise, enables us to start constructing those more sophisticated models that I mentioned earlier. These models and associated data often take a great deal of guesswork out of the design process, further increasing our probability of avoiding and preventing problems.

The third advantage is that if we are willing to adopt part two of the two-part strategy, we can make much more accurate estimates of risk with continuous data than we can with discrete data, which can allow us to reduce our reserves against problems or otherwise lower our insurance if need be.

The best part about adding sensors or modifying tests to produce measurable, continuous data is that, relatively speaking, it's inexpensive. I say relatively because, although the data collection equipment, training, and the sensors can be costly, they are generally re-usable and they are less expensive than performing more tests.

Now for part two of the strategy I recommend. You guessed it already. Take more than one data point! Yes, test several. Now I need to explain one more thing before I explain what to do with more than one data point. Even if you are already taking 30 data points, if all you do is pass or fail your design without any further analysis than a count of results, your risk really isn't much different than taking one data point.

Don't just test a few and count the passes and fails. Unless you are testing every product unit before it ships, you aren't really improving what you know by testing more than one. It's easiest to see when we consider the example of testing 30 units (because we have all heard that most statistical distributions tend to reveal themselves after about 30 samples and so 30 has become a rule of thumb) and then we declare after all 30 pass that our product is good. If we are producing hundreds of thousands of those products a year, how demonstrative is that sample set of 30 really?

Thirty might be more than enough. The best way to use your test samples is to calculate your risk based on the results. One sample will never be enough to do this, though. The calculations to estimate risk, or probability of a failure, are

relatively simple. The hard part is making some decisions and facing the answers.

It’s a funny phenomenon, but it seems easier to take a risk when we don’t know precisely what it is, than it is to deal with a risk when we know exactly what it is.

Case in point, it’s often easier to look at a set of 5 samples that all passed and decide that the product is good to launch, than it is to look at the calculation of risk from those same five test samples and decide what to do when the data tells you there is a 73 percent chance of a test result outside of acceptable limits. Aesop summed it up with the moral, “ignorance is bliss.”

Quickly, let’s discuss what we need to decide and to face when we examine the results of multiple test samples. I won’t torture us with math. If you want to understand the specific calculations, look up “Power and Sample Size” calculations in your favorite resource on statistics.

Whether you are dealing with discrete data or continuous data, the parameters of the equations are the same. However, as I mentioned, we can usually converge on an accurate answer with many fewer data points when the data is continuous. The parameters are as follows.

- Sample size.
- Standard deviation (variation in results between samples) or proportion of samples defective for discrete data.
- Precision (how sensitively we can tell the difference between one outcome and another).
- Confidence level (our risk of being wrong).

To solve the equation we must know at least 3 of the 4 parameters and so we solve for the last one. Unfortunately, we don’t always know three and so we must assume one of them. Usually, we decide on a confidence level, typically 95%, meaning that the next data point is 95% likely to fit our model, or we are allowing a 5% chance of being wrong. Our precision is usually determined by our measurement capabilities, and we can fill it in if we know. Standard deviation must come from our test results or from experience testing similar items similarly.

There are two ways to use this equation. The one the statisticians tell us to do, is to assume a precision, confidence level, or standard deviation based on our best guess from experience and then calculate how many samples we need to test. I’ve done this many times, and usually the recommendation is for more samples than we can afford to test.

I’ve reverted to the opposite strategy and I suggest that you give it a try too. Determine how many samples you can afford to test, using your best abilities to capture continuous data instead of discrete data. Also, it is worth the investment to improve your ability to tell the difference between results. By improving your ability to measure continuous results, we reduce the likelihood of measurement method driving us to test more samples or rob us of confidence.

Let’s say that when we evaluate the risk of having a positive test trick us into launching a non-compliant product or the possibility of a retest leading to more design work and testing, we can afford to test six samples for greater confidence. Now we go ahead and test six and, rather than guessing at our precision and standard deviation, we know them precisely. What we solve for is the confidence level.

If all six samples are tightly grouped with very little variation in measurable results, and our precision is much less than the standard deviation, we might find that six samples gives us a very high confidence. I’ve seen six samples give a confidence of 89 percent. If 89 percent isn’t enough for us, we can play games, using actual numbers for standard deviation and precision, and decide if one or two more samples tested would significantly increase our confidence or if we should skip them.

As I mentioned, when we use this method and we get a confidence number that we don’t like is when we run into problems. If our confidence turns out to be only 57 percent it’s not so easy to decide to launch. Now it’s difficult decision time, but we have options.

We can try to identify some relatively easy design improvements that are both expedient and economical. If beefing up the hardware at an attachment point, or lengthening a welded gusset will improve our safety factor or eliminate a failure mode, we can probably make that change in a day or two and run another very small test to prove it made the difference needed. Another mathematical formula can tell us how likely the second results are to be part of the same population as the first results.

We can have an intelligent conversation about how likely we are to have a problem in the field and whether design work is better or setting aside a contingency reserve is the right thing to do. We can also look at the nature of the failure modes observed and decide if they are unique to the test scenario or if they are indeed going to happen in the field.

At least, armed with more information than a single test sample can provide, or more information than any arbitrary test sample set will provide without understanding the variation of the results, we can move forward intelligently and with some understanding of our risk. It’s not as comfortable as ignorant bliss, but it’s smarter.

When we adopt this routine, something inevitably happens in our design process as well. We begin to pay more attention to reducing the variation between our units and to predicting the outcomes and performances. This is a very good habit to get into. Our designs and our design methods become much more elegant as we seek to improve our probability of supply, manufacture, and assembly, producing a product that is always in the performance envelope.

Take a look at your test methods. If you are not doing enough analysis to predict

the likelihood of a test sample producing an unfavorable result, then you are not using your test information as intelligently as you could, or you are not collecting enough information to know what the truth might be.

Move away from testing an arbitrary sample size and launching if they pass, especially if that arbitrary number is one. Move toward a mindset of collecting as much information as you can afford. That means maximizing samples tested as well as implementing ways of collecting relevant continuous, measurable data instead of pass/fail results.

Make the uncomfortable shift of addressing the risk your limited information indicates instead of gambling on the utter lack of knowledge a few passed samples, or one sample, provides without an analysis of the variance of the output.

Don't gamble your business, your customers' trust, or people's safety on a hope and a prayer. Instead maximize the information available from the limits of your tests, hedge your bets, and work intelligently toward ensuring the integrity of your products and your business.

Stay wise, friends.

Find more of Alan's thoughts at www.bizwizwithin.com *[1].*

Source URL (retrieved on *07/30/2014 - 1:04am*):
http://www.impomag.com/blogs/2012/09/some-tribulations-testing-and-truthfulness?qt-most_popular=0&qt-digital_editions=0

Links:
[1] http://www.bizwizwithin.com