Field Failure Data – the Good, the Bad and the Ugly

Dr. William M. Goble, CFSE
exida
Sellersville, PA 18960, USA
wgoble@exida.com

Revision 2.5 – February 2017

Keywords: safety instrumented system, component failure rates, field failure studies.

Abstract
There are many benefits to a company when they have access to realistic field failure data for their site. Most of the benefits are categorized as saving money. At the same time, most of the expenditure to collect and analyze failure data is already being spent in other maintenance activities. Given an incremental cost of improving data collection quality and better data analysis, a nice return on investment could be achieved.

Good high quality field failure data has often been described as the ultimate source of estimated failure data. However, not all field failure studies produce data suitable for a specific industry/site. Some field studies simply do not have the needed information. Some field studies make unrealistic assumptions. Some data collection procedures are badly flawed. The results can be quite different depending on methods and assumptions. Some methods produce optimistic results that can result in bad designs and unsafe processes.

This paper presents some common field failure data collection and analysis techniques, shows some of the problems and limitations of the methods and describes important attributes of a good field failure data collection system that can provide high return on investment.

Introduction
The benefits of having quality product failure rate data include:

- The ability to achieve less process downtime [ISO16],
- identification of efficiency/reliability improvement opportunities[Skwe08],
- optimized (lower cost) maintenance programs,
- opportunities to optimize the timing of equipment overhauls and inspections [ISO16],
- reduced environmental impact [ISO16],
- proper application fit of instrumentation products, and
- getting data that can be used to optimize safety instrumented system verification calculations for a particular application [Skwe08, Gobl17].

All of these reasons can be categorized as simply reducing expenses. Most companies already spend most of the money required to get good data. What is missing in some cases is a small amount of incremental data and better data organization and statistical analysis.
Quality field failure data has often been described as the ultimate source of reliability information. However, not all field failure data studies produce the needed quality of data. In comparing the results for the apparently same product in a similar application/environment, the numbers vary considerably. Consider the variations in failure rate data presented in Table 1 electronic / mechanical pressure transmitters.

<table>
<thead>
<tr>
<th>Source</th>
<th>Product Type</th>
<th>Failure Rate per hour</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCPS89</td>
<td>Diff. Pressure Transmitter, mean</td>
<td>6.56E-05</td>
<td>Highest Number</td>
</tr>
<tr>
<td>NRPD95</td>
<td>Pressure Transducer</td>
<td>8.31E-06</td>
<td></td>
</tr>
<tr>
<td>Refinery Data [She100]</td>
<td>Analog Pressure Transducer</td>
<td>2.71E-06</td>
<td></td>
</tr>
<tr>
<td>Refinery Data [She100]</td>
<td>Microprocessor Based Pressure Transmitter</td>
<td>7.19E-06</td>
<td></td>
</tr>
<tr>
<td>OREDA [ORED09]</td>
<td>Pressure Transmitter</td>
<td>4.10E-07</td>
<td></td>
</tr>
<tr>
<td>DOW Plant Study [Skwe08]</td>
<td>Pressure Transmitter</td>
<td>4.96E-07</td>
<td></td>
</tr>
<tr>
<td>Manufacturer Study [Moor98]</td>
<td>Microprocessor Based Pressure Transmitter</td>
<td>3.57E-07</td>
<td>Lowest Number</td>
</tr>
</tbody>
</table>

Table 1: Field Failure rate results from different sources for electronic pressure sensors.

The ratio of the highest (most pessimistic number) to the lowest (most optimistic number) is over 100X. Similar results occur for mechanical products as shown in Table 2.

<table>
<thead>
<tr>
<th>Source</th>
<th>Product Type</th>
<th>Failure Rate per hour</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCPS89</td>
<td>Solenoid Valve-Operated, mean</td>
<td>4.87E-05</td>
<td>Highest Number</td>
</tr>
<tr>
<td>NRPD95</td>
<td>Valve, Pneumatic Solenoid</td>
<td>1.67E-05</td>
<td></td>
</tr>
<tr>
<td>OREDA Derived [exid15]</td>
<td>Solenoid Valve</td>
<td>3.86E-07</td>
<td></td>
</tr>
<tr>
<td>Refinery Data [She100]</td>
<td>Solenoid Valve</td>
<td>Not Available</td>
<td></td>
</tr>
<tr>
<td>DOW Plant Study [Skwe08]</td>
<td>Solenoid Valve</td>
<td>7.02E-07</td>
<td></td>
</tr>
<tr>
<td>Manufacturer Study [IMI15]</td>
<td>Solenoid Valve</td>
<td>8.40E-08</td>
<td></td>
</tr>
<tr>
<td>Manufacturer Study [AEAT05]</td>
<td>Solenoid Valve</td>
<td>1.70E-08</td>
<td>Lowest Number</td>
</tr>
</tbody>
</table>

Table 2: Field Failure rate results from different sources for pneumatic solenoid valves.

Which number is “correct”? Why the different results? There are many possible reasons including differences in the:

1. Category Grouping: what is included in the failure rate group?
2. Date of Design/Manufacture: old technology devices versus new devices with improved reliability
3. Application environment of the site: different stress conditions
4. Site Safety Culture – operations and maintenance capability
5. Definition of what is counted as a failure
   a. random versus systematic – Are many failure reports thrown away?
   b. initial probability of failure included?
   c. wear-out failures included?
6. Failure recognition method
   a. automatic diagnostic
   b. process upset
   c. manual proof test
7. Failure data recording and collection policy
   a. simply not recording the data
   b. product categories
   c. product identification
   d. record only if process trip occurred
   e. record only if outside service required
   f. record only if external failure cause
   g. records retained only until repair completed
8. Assumptions used to calculate operating hours
9. Assumptions used to calculate number of failures
10. Statistical analysis techniques with invalid assumptions

1. Category Grouping
In field failure data collection systems, there are a number of “bins” into which the field failure reports are placed. Each bin represents a product category where data is analyzed and failure rate data is generated. The number of product categories varies substantially between different systems. Systems with fewer categories require the aggregation of data from products with highly different failure rates. This reduces the usability of the results. For example, what is included in the category “pressure transmitter”? Does the category include simply the pressure transmitter body? Are impulse lines included? Are manifolds included? Are remote seals included? The failure rate results would certainly vary by an order of magnitude or more depending on the answers to those questions.

2. Date of design/manufacture
Many manufacturers have strong programs to improve the quality and reliability of their products. The failure rate of a product designed and manufactured in 1990 is substantially higher than a similar product designed and manufactured in 2010. Tables 1 and 2 show results from an older analysis [CCPS89, NRPD95] that are high when compared to newer results [Skwe08, ORED09].

3. Application environment
Failures occur when stress exceeds the strength of a product [Gobl17]. Some sites have higher temperatures, more vibration, higher corrosive atmospheres, etc. While many of these factors do not change failure rates significantly, others are more critical depending on the product. If failure reports from different application environments are lumped together, only an average value results.

4. Site Safety Culture
Several studies have shown the failure rate of a specific manufacturer and model have varied from site to site [Buko16a, Buko16b] with failure rates changing by 2-4X even when application environments are similar. Therefore failure reports must record the specific site and perhaps the specific operating unit.

5. Definition of a failure
Different failure recording and analysis systems have used very different definitions of what is a failure. Many field failure recording systems do not count “systematic” failures as a failure. Studies have shown that the definition of systematic varies considerably from person to person [Gobl16]. If “systematic failures” are discarded, the failure rate results will be highly optimistic. Systems operated by manufacturers often review all product returns and perform incoming testing of each returned device. Those devices passing the test are marked “no problem found” and not counted as a failure. However, the tests may be incomplete and may not be performed under field operating conditions. These results are often highly optimistic as well.

6. Failure recognition method
Failures are discovered by a number of methods:
   a. Automatic diagnostics
   b. Manual proof testing
   c. Process upsets
Some failure recording systems do not include all sources of failures. Yet they do count all operating hours. This will result in optimistic failure rates.

7. Failure data collection
Some systems are set up to collect failure data only at a specific point in the maintenance process. What happens when an out of calibration transmitter is re-calibrated during the proof test? Some do not count this as a failure yet it may not indicate a dangerous condition if the calibration is too far out. “As-found” conditions should always be recorded and if the calibration is too far out, a failure event must be recorded for accurate results. Many other examples exist.

8. Operating hours calculation
In some products a power on hours counter keeps accurate track of operating hours for that device. However, that is rare. In all other cases, some estimate must be made. If the operating plant has a good inventory database the estimates can be pretty good. Manufacturers do not have this level of detail. They often use shipping records to estimate unit operating hours and unless the assumptions are very conservative, the results are optimistic.

9. Number of failure assumptions
In addition to the failure definition issue, the number of failures reported can be highly variable. One of the biggest issues is in manufacturer field failure data systems. As assumption must be made regarding the percentage of failures returned to the manufacturer. Some systems assume all failures are returned. That can be the case but often few failures are returned. Surveys indicate as few as 10% are returned for inexpensive products like solenoid valves and trip relays. The 100% assumption versus a 10% number results in an order of magnitude change in the failure rate.

10. Statistical analysis techniques
The statistical analysis part of a field failure system is usually the least significant in terms of failure rate variability. Some assumptions are important especially the confidence factor when few (or no) failures are recorded for a given category. Confidence factors of 70% or 90% are suggested in IEC 61508. The impact is important only when few failures are recorded.

There are so many reasons why failure rate analysis results can change that it is totally useless and highly misleading to compare results without a complete understanding of the data and how it is collected. These variations can result in numbers that change by orders of magnitude. It may seem as if good field failure rate data is not possible after understanding all the reasons why failure data collections systems fail. But there are good systems. The OREDA failure data collection system based on ISO 14224 is one of the best in the world. The analysts explain how data is collected and analyzed. That is an example of a good system.

The Good – A good data collection program
A good field failure data collection program needs to be founded on a good reliability/safety culture where all failures are reported, where all failure detection mechanisms are included, where the field failure data is valued and where procedures for gathering the data are well defined and followed. The system must allow for easy capture of the data right after a failure is resolved not the next day, the next week or the next month. Generally, a good system is tied to the plant work order system. A good system tracks work to be done and any overdue activities such as proof testing. Figure 1 shows the dash board task tracking screen from the exSILentia SILStat™ software.
Data should be collected for each device rather than a collection of devices. For example, failures recorded for a “valve assembly” are of value but failures recorded for each solenoid, each actuator type and each valve type are much better. A good description of exactly what is included in each device category is needed. Are the failure records kept for each type of pressure transmitter or do the failure records include the transmitter, manifold, impulse line, remote seals, and other devices?

All failures must be recorded. It is misleading to exclude some failures because they were classified as “systematic” [Gobl16]. The collected data must be carefully analyzed and actually used to identify opportunities for cost savings. There is nothing like a documented cost saving to generate continued management support.

**Inventory Database – information about each instrumentation product**

A good field failure data collection program has a number of important attributes. It specifies a site inventory database that includes for each instrument:

- Product manufacturer and model number
- Serial number
- Design version level
- Product classification
- Commissioning date and decommissioning date
- Operational duty cycle (e.g. 24/7 or other?)
- Installation site
- Operating environment as installed – ambient and process
- Failure event records and proof test records – dates, methods, results
- Calibration information

This information clearly allows the technology vintage to be determined. Knowing that manufacturer’s have been working hard for many years to improve the reliability and quality of their products, averaging the results from products designed years ago with products made with newer improved designs misses an opportunity to see the improvements and utilize the lower failure rates to reduce cost.
The information in the site inventory database also allows sorting of wear-out failures from random useful life failures. This also is important as predictable failures due to wear-out should not be included in the random failure rate. It is also very important to understand when wear-out occurs so that preventative maintenance programs can reduce unanticipated downtime and improve safety by replacing the products before failure occurs.

Whenever practical, the product classification scheme in the site inventory database should separate products into groups with similar failure rates and modes. For example, a poppet type solenoid valve will have different failure rate data than a spool type solenoid valve. A butterfly valve will have different failure rates than a gate valve. A product grouping where apples, oranges, and bananas are grouped and called “fruit” gives failure data of reduced accuracy for all elements of the group.

The commissioning date, decommissioning date and operational duty cycle should enable accurate calculation of operating hours for each instrument. If the product automatically keeps track of power-on hours as some smart instruments do, the power-on-hours information can be extracted and recorded as well.

The actual installation location and the operating environment for each product at that site must be recorded in the database. The site description is also used to assign a maintenance capability rating.

**Failure Events – information about each event in a product history**

For each product in the site inventory database ALL failure related events are recorded. This includes all failure events detected by any automatic diagnostics, any process disruptions or false trips, failures caused during maintenance activities and any failures detected during manual proof tests. For each event, the date/time should be recorded. The person responsible for the repair must be recorded along with failure type, ideally from a short pick list of defined failure events. Figure 1 shows the event capture screen from the exSIListia SILStat™ tool.
Any suspected causes should be recorded even when caused by a maintenance mistake. The failure rate database should NEVER be used to reprimand or even evaluate the maintenance team. Any correlating events that might be related to the failure should be recorded. For example “Failure occurred during lightning storm” or “Failure occurred during product calibration” would be important information. For proof test results, the proof test method should be recorded so that proof test effectiveness can be evaluated and considered during failure rate analysis.

The data should be analyzed by an experienced reliability engineer with specialty knowledge in field failure data analysis and statistics. The data analyst must understand the technology of the instrumentation products as this allows faster and more accurate separation of the data between random, systematic and wear-out. Data analysis requires careful review and validation. When a system has this kind of data and analysis, good high quality failure rates can be calculated when enough data is gathered.

A good field failure data collection system also involves full management support and evergreen training of those involved (Engineering, Maintenance and Operations) [Skwe08].
“The Bad” – A less than good data collection program

When the kind of data described in “The Good” field failure data system is not practical, assumptions must be made to simplify the process. Even without all the information, very useful data can be obtained [Buko07]. However, depending on several variables, the quality of the data can vary considerably.

If the application environment and site information is not recorded for each product it is possible that a wide variety of different stress conditions can be mixed together giving useful results but not quite right for any environment.

Perhaps the biggest issue is the recording of all failure related events. The definition of “failure” varies substantially. In one site “failure records” were kept only when a product had to be sent off site for repair [Sieb07]. All failures were not recorded. If the data analyst did not know that, resulting numbers would be very optimistic. In another study, many failures were classified as “systematic” and not counted in the failure rate.

Some field failure data systems record only failures that cause a process upset/trip. Any “failures/issues” discovered during manual proof testing are not recorded. In some systems all items “replaced” during a repair are considered a failure. Often multiple items were replaced because it is faster and cheaper to replace them all rather than troubleshoot to the point of identifying the actual failure.

Data analysis work can be done in several different ways depending primarily on the assumptions used and the review process used to sort the data. In one field failure study [Buko07], it was discovered that many failures were being caused by faults in newly manufactured product or an error made sometime during installation and commissioning. This resulted in many units having what appeared to be random failures during the next manual proof test. A “failure rate” calculated without recognizing the pattern in the data would be quite pessimistic. This situation is more correctly modeled with “initial probability of failure,” a random failure rate during useful life and a limit on useful life.

In other studies it is common to discover that product serial numbers are not recorded, commissioning dates are not recorded and there is little information available to calculate unit operating hours. So assumptions are made. These assumptions can, of course, have a major impact on the result.
The Ugly
One of the most abused and potentially misleading field failure study methods is the manufacturer’s warranty field return study. Problems are caused not by the calculation method itself but rather the lack of information. Without information assumptions must be made. Often the assumptions used result in highly optimistic failure rates being published. Consider Table 2: the most optimistic results are from manufacturer's warranty return data systems. That result is over 100X smaller than the most pessimistic result.

A typical analysis method includes an estimate of operational hours in the field and a count of “failures” reported. The failure rate calculation is simply the number of failures divided by the total operational hours.

\[ \lambda = \frac{\text{# of failures}}{\text{# of field operational hours}} \]

Quality training during the 1980s and 1990s [Jura91] taught this simple calculation but never explained the assumptions needed or the limitations of the method. Many of the analysts working for product manufacturers do not understand that while these methods are useful for internal quality comparison purposes, they should not be used to publish absolute failure rates for safety critical system design.

Consider what information does an original equipment manufacturer really know? The number of units manufactured is known. The shipping dates are known. Therefore the estimate of operational hours usually assumes some period of time between shipment and operation, some percentage used for spares, and some assumption about time in service per day. Many manufacturers estimate a period of two to four weeks between shipment and operation. Often it is assumed that all units shipped are put into service. Another assumption is that the devices are operated 24 hours a day, seven days a week. Knowing the total number shipped each week along with all the assumptions allows the "operational hours in the field" to be calculated [vanB98]. While this sounds reasonable the assumption is typically quite optimistic. The time it takes from shipment to usage in the process industries will likely be much longer, six months or even one year.

But the failure count itself is the big variable. The failure count is typically based upon field return records where the failure is verified by testing the returned product. As products are returned they are tested. Many times the test shows "no problem found." Therefore it is assumed that those returns must not be real failures and are not counted. There is also another optimistic assumption that the test can detect all possible failures. What if the failure occurs only under specific field conditions not duplicated in the manufacturing test? What if the test simply is not complete and does not detect all failures? Very optimistic results can occur.

Perhaps the biggest level of optimism comes with the assumption that all field failures are reported. A survey of end users done by exida found that often it is cheaper to throw away the product than send it back to the manufacturer. Sometimes repairs are made by the end users and the failure is not reported back to the manufacturer. This is especially true of mechanical devices like valves and actuators where "rebuild" kits are sold. Overall end user survey results indicated that roughly 10% of failed items were actually sent back to the manufacturer for inexpensive devices. The percentage returned can be much higher depending on the device cost and the manufacturer repair policies.

Overall this collection of assumptions results in very optimistic and very low (unsafe) failure rate estimates as can be seen in the last line of Tables 1 and the last two lines of Table 2.
Conclusion

Ultimately field failure data is the source of all reliability information. Therefore the importance and value of high quality field failure data cannot be argued. However, high quality data depends on accurate recording of enough information so that assumptions are minimized. Since this does not always occur, field failure data must be provided along with a full explanation of the failure event definitions, methods for data collection and methods for data analysis including all assumptions. Without this, the results can be dangerous.

Standards [ISO16] exist to help define field failure databases and collection methods. The Process Equipment Reliability Database (PERD) initiative of AIChE (http://www.aiche.org/CCPS/ActiveProjects/PERD/index.aspx) is developing taxonomies, methods and software to help promote higher quality field failure data [CCPS98]. As these methods mature, the data will get better.

Tools are also being developed like the exida SILStat™ software. This program is fully compliant with standards [ISO16], and PERD. SILStat™ comes fully populated with failure modes per product type and makes it easy to set up a data collection system.

REFERENCES


[NPRD95] Non-electronic Parts Reliability Data, 1995, Reliability Analysis Center, Rome, NY


