Paper 424-2013

Good as New or Bad as Old? Analyzing Recurring Failures with the RELIABILITY Procedure

Roberto G. Gutierrez, SAS Institute Inc.

ABSTRACT

You can encounter repeated failure events in settings ranging from repairs of equipment under warranty to treatment of recurrent heart attacks. When a system fails repeatedly, the risk of failure can change with each subsequent failure—a unit once repaired or a patient once treated might not be “good as new.” Analysis with the RELIABILITY procedure focuses primarily on the mean cumulative function (MCF), which represents either the average number of failures per unit over time or some related cost measure. This paper describes how you can use PROC RELIABILITY to estimate and compare MCFs. You can choose either a nonparametric or a parametric approach. Features added to the procedure in SAS/QC® 12.1 are highlighted.

INTRODUCTION

Statistical models for recurring events are common to various fields and are used to make inference about the number of times an event takes place over a time period. These events are called *failures* because often that is what they represent, but “failure” can be any event of interest. For example, a drug manufacturer might be interested in modeling the number of cardiovascular events for patients in a new drug trial, or an automobile manufacturer might want to know the expected number of warranty repairs per new vehicle sold.

In both cases, the statistical issues are the same, and the statistical science of event counting is directly related to that of modeling the time between events. This science is known by various names, the most common being *survival analysis* in biostatistics and medicine, *reliability analysis* in engineering, and *duration analysis* in economics and social sciences. Regardless of the name and the field in which it is used, this type of analysis is replete with jargon, and the jargon varies across fields. Because the RELIABILITY procedure in SAS/QC focuses on engineering applications, this paper uses terminology from that field. However, what is presented here also applies to other fields (in particular, medicine).

The key to analyzing recurrent-failure data is understanding the process that generates failure. Failure is the result of the accumulation of risk, and the rate at which risk is accumulated is represented by what is known as the *intensity function*. The integral of this function represents the amount of accumulated risk at any given time, and the integral can be interpreted as the number of failures that you can expect to observe before that time. Engineers call the accumulated-risk function the *mean cumulative function* (MCF). Because the MCF represents the expected number of failures, it is ideal for recurrent-failure analysis.

You need to consider two issues when you examine MCFs and the intensity functions that underlie them. First, you should think carefully about what happens when items are repaired and returned to service after they have failed. Are they repaired or replaced so that they are good as new? Or are they repaired minimally (that is, are they bad as old)?

Second, you need to decide whether to take a nonparametric approach or a parametric approach. For the nonparametric approach, the issue of good as new versus bad as old is irrelevant, but that does not hold for the parametric approach.

Meeker and Escobar (1998) gear their general overview of recurrent-failure analysis toward engineers. Discussions specific to recurrent events are given by Nelson (2002) and Rigdon and Basu (2000).

This paper is organized as follows: The next section briefly discusses the intensity function and MCF. The subsequent sections describe data analysis in the bad-as-old and good-as-new scenarios, respectively. Each of those two sections features both a nonparametric approach and a parametric approach to the analysis.

THE INTENSITY AND MEAN CUMULATIVE FUNCTIONS

The intensity function for a particular time t is denoted by $\lambda(t)$. Statistically, $\lambda(t)$ represents the probability that a unit will fail within a very small window of time around time t. However, this interpretation relies heavily on the intricacies of mathematical limits and continuity of functions. For example, consider $\lambda(t)$ as the rate at which an automobile engine runs. Experience tells you that this rate is merely the engine's revolutions per minute (RPM) at a particular point in time; sometimes the engine revs heavily, sometimes it idles, and sometimes it is shut off completely. In other words, $\lambda(t)$ can change over time.

To generalize the concept of $\lambda(t)$ from engines to anything you could possibly imagine, you only need to generalize the units of measurement; rather than use the revolutions per minute, you use the number of failures per unit time, where the unit of time is specific to the problem at hand. In fact, "time" doesn't even have to represent calendar or clock time; it could represent the number of times a television is switched on, for example.

The mean cumulative function at time t, MCF(t), is calculated as

$$\text{MCF}(t) = \int_0^s \lambda(s) ds$$

MCF(t) represents the accumulation of intensity (or risk) from inception (time 0) to the current time t. In the automobile engine example, MCF(t) represents the total revolutions of the engine. Whether the revolutions were the result of constant highway use or alternating periods of idling and racing, it is only the number of revolutions (not how they accumulated) that determines the likelihood of engine failure before time t. Although this is a somewhat simplistic view of how engines wear down, the main point is that total risk promotes failure much more than the risk profile does.

Recall that, in general problems, $\lambda(t)$ represents the number of failures per unit time. Because MCF accumulates $\lambda(t)$, its units are the total number of failures before time t. Statistically you can model either $\lambda(t)$ or MCF(t). You can specify models for $\lambda(t)$ to correspond to increasing risk due to aging, decreasing risk due to the weeding out of defective units, or a combination of both. These models are parametric and highly structured. Although they carry more assumptions, they enable you to predict both future behavior and behavior under different stress conditions. Alternatively, you can estimate MCF(t) directly to determine the expected number of unit failures, without regard for the risk profile that underlies these failures. This nonparametric approach carries fewer assumptions, but it doesn't allow for inference outside the conditions and time period under analysis.

Which strategy you choose can also depend on what you assume takes place when units fail: whether they are minimally repaired to be bad as old, or repaired or replaced to be good as new. Each scenario is described in the following sections.

THE BAD-AS-OLD SCENARIO

More statistical literature and software are available for the bad-as-old scenario. When a system fails and is minimally repaired, its intensity (risk) profile remains the same as if it hadn't failed. Consequently, its risk profile is similar to that of a single-failure system. Because human beings are single-failure systems (the failure is death), a vast statistical literature and a wealth of methodology are available to build on.

The following SAS statements create the data set Diesel, which contains repair histories of 41 diesel engines (Nelson 1995). Failure occurs when an engine valve seat wears out and must be replaced.

```
data Diesel;
   input Id Days Event @@;
   label Days = 'Time of Replacement (Days)';
   datalines;
251 761 -1     252 759 -1     327  98  1     327 667 -1
328 326  1     328 653  1     328 653  1     328 667 -1
329 665 -1     330  84  1     330 667 -1     331  87  1
331 663 -1     389 646  1     389 653 -1     390  92  1
390 653 -1     391 651 -1     392 258  1     392 328  1
392 377  1     392 621  1     392 650 -1     393  61  1
393 539  1     393 648 -1     394 254  1     394 276  1
```

```
394 298  1      394 640  1      394 644 -1      395  76  1
395 538  1      395 642 -1      396 635  1      396 641 -1
397 349  1      397 404  1      397 561  1      397 649 -1
398 631 -1      399 596 -1      400 120  1      400 479  1
400 614 -1      401 323  1      401 449  1      401 582 -1
402 139  1      402 139  1      402 589 -1      403 593 -1
404 573  1      404 589 -1      405 165  1      405 408  1
405 604  1      405 606 -1      406 249  1      406 594 -1
407 344  1      407 497  1      407 613 -1      408 265  1
408 586  1      408 595 -1      409 166  1      409 206  1
409 348  1      409 389 -1      410 601 -1      411 410  1
411 581  1      411 601 -1      412 611 -1      413 608 -1
414 587 -1      415 367  1      415 603 -1      416 202  1
416 563  1      416 570  1      416 585 -1      417 587 -1
418 578 -1      419 578 -1      420 586 -1      421 585 -1
422 582 -1
;
```

The variable **Id** identifies the engine, the variable **Days** measures the engine age in days, and the variable **Event** is equal to 1 if the engine failed at the given age or –1 if **Days** marks the end of the engine's observation history.

Observations for which **Event** is equal to –1 are known as *right-censored*, because time to failure has not been exactly determined but is known only to have not occurred by the observed time. Because waiting for every unit to fail is often infeasible, right-censoring is a common aspect of data about recurrent events.

By treating these data under the bad-as-old scenario, you assume that valve seat replacement leaves an engine in a condition comparable to that of an engine of a similar age that has not yet failed. This assumption might seem strong on the surface, but it often holds true in practice in systems that contain many replaceable parts. Here, the bad-as-old assumption is reasonable because engines have multiple valve seats. After a valve seat is replaced, subsequent failure will most likely occur not in the replaced seat but in a different one. In the next section, output from PROC RELIABILITY demonstrates the validity of the bad-as-old assumption, but it is good practice to think carefully about this assumption before proceeding with your analysis.

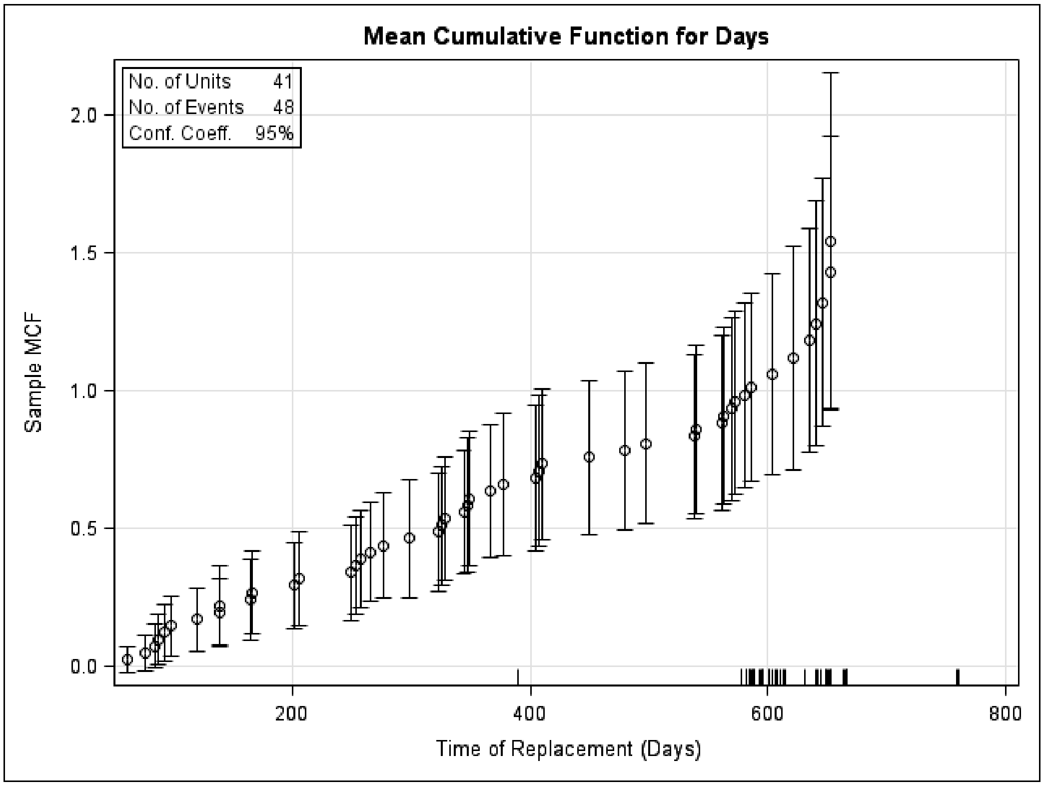
Nonparametric Analysis

A nonparametric analysis estimates the MCF directly by using a sample-based estimator that requires no assumption about the intensity function that underlies the MCF. The following SAS statements produce a plot of the estimated MCF, with 95% confidence bands:

```
proc reliability data=Diesel;
   unitid Id;
   mcfplot Days*Event(-1);
run;
```

Figure 1 shows the resulting SAS® Output Delivery System (ODS) graph, which demonstrates that the number of expected failures grows at an increasing rate.

Figure 1 Estimated Power MCF Function for Engine Data



Providing confidence bands on MCF plots is a feature that is new in SAS/QC 12.1. The expected time to the first engine failure is about 580 days. Also note that the confidence bands widen as time increases. This is a common occurrence in recurrent-failure analysis—the bands become wider because the effective sample size decreases as engines leave the study.

Parametric Analysis

Proceeding with a parametric analysis, note that the previous results indicate that valve seats deteriorate with age, leading to an intensity of failure that is monotonically increasing. That is because the MCF grows at an increasing rate, meaning that its derivative, the intensity function, is an increasing function. A popular parametric model for monotonic risk (whether increasing or decreasing) is the power model, under which

$$\lambda(t) = \left(\frac{\beta}{\eta}\right)\left(\frac{t}{\eta}\right)^{\beta-1}$$

with the corresponding MCF$(t) = (t/\beta)^{\beta}$. The parameter β is a shape parameter and is always positive. When β is greater than 1, intensity increases with age; when β is less than 1, intensity decreases with age. The parameter η is a scale parameter. Both β and η are estimated from the data.

The power model is an example of a nonhomogeneous Poisson process (NHPP) model. The genesis of that name is beyond the scope of this paper, but it suffices to think of an NHPP model as "the parametric model for bad-as-old data."

The following SAS statements fit the power model to the diesel engine data and plot the estimated MCF and intensity functions:

```
proc reliability data=Diesel;
   unitid Id;
   distribution NHPP(Pow);
   model   Days*Event(-1);
   mcfplot Days*Event(-1) / Fit = Model;
run;
```

Note that the syntax is similar to that of the nonparametric analysis, but it contains additional DISTRIBUTION and MODEL statements that specify the NHPP-power model.

The model parameter estimates are shown in Figure 2.

Figure 2 Power Model Parameter Estimates for the Engine Data

The RELIABILITY Procedure

NHPP-Power Parameter Estimates

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
Intercept	553.6430	57.8636	451.0941	679.5048
Shape	1.3996	0.2005	1.0570	1.8533

Because β is estimated to be greater than 1, the intensity (or risk) of failure increases with time. In Figure 2, the estimate of scale η is shown in the row labeled "Intercept" and can be interpreted approximately as a typical time to first failure (in this case, about 554 days).

By default, the MCFPLOT statement produces a plot of the estimated MCF and a plot of the estimated intensity function. Figure 3 shows the model-estimated MCF function (solid line), the confidence bands around this function (shaded area), and the nonparametric estimate of the MCF function (scatter plot). The model fits the data fairly well, lending credibility to the bad-as-old assumption.

Figure 3 Estimated Power MCF Function for Engine Data

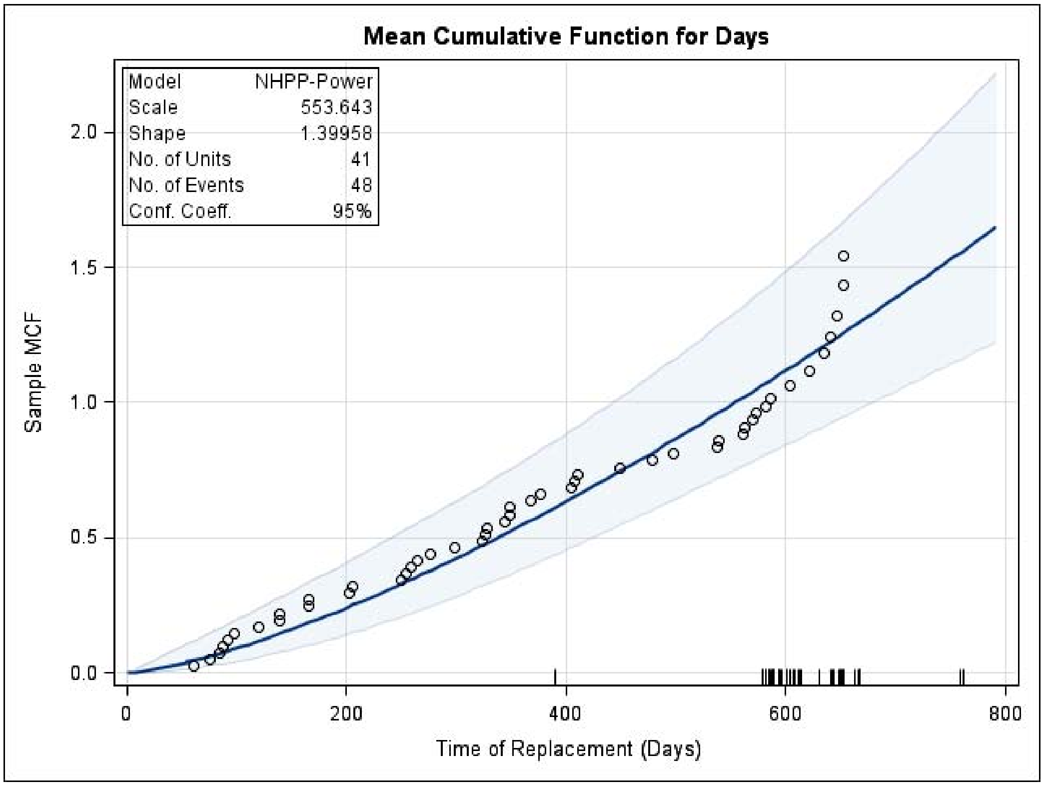
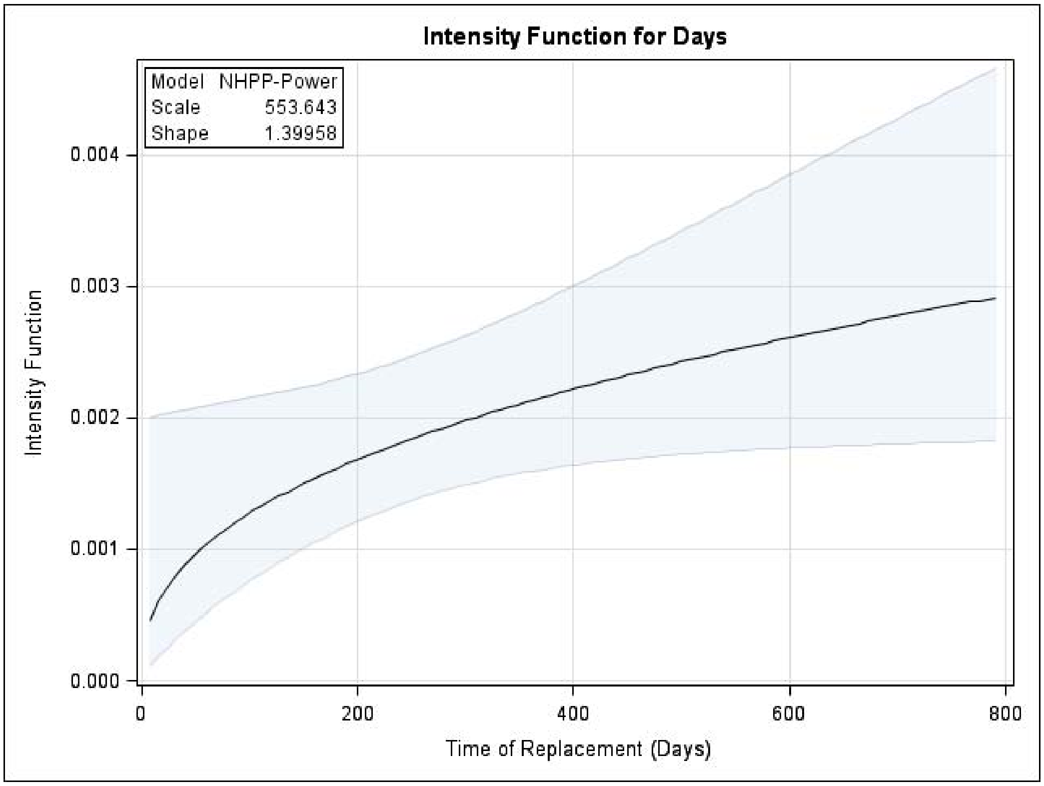


Figure 4 plots the estimated intensity function based on the model fit and, as expected, demonstrates monotonically increasing risk of failure over time.

Figure 4 Estimated Power Intensity Function for Engine Data



The model that is used assumes not only a bad-as-old scenario but also that the intensity can be represented by a power function of time. Both assumptions appear valid for these data. Two advantages of that parametric model are that it yields more precise estimates of both the intensity function and the MCF and that it enables you to predict what happens at times other than the times you observe in the data. But be careful not to predict too far out!

PROC RELIABILITY also provides support for the following:

- four parametric models in addition to the power model: the constant-intensity, Crow-AMSAA, log-linear, and proportional-intensity models
- all censoring patterns, including left-censoring and interval-censoring patterns
- regression models that adjust for covariates such as temperature, voltage, and other forms of stress

For more information, see the chapter about the RELIABILITY procedure in the *SAS/QC 12.1 User's Guide* (SAS Institute Inc. 2012).

THE GOOD-AS-NEW SCENARIO

When units are repaired in such a way as to render them good as new, the intensity function reverts to its behavior at time 0 with each successive failure. Such a process is known in the statistical literature as a *renewal process*.

Admittedly the title of this section is something of a misnomer. It is rare in practice that repaired (not replaced) systems behave just like new, so in practice you do not assume like-new behavior. Instead, you simply assume a behavior that is different from bad-as-old behavior. There is a strategy for bad-as-old data, covered in the previous section, and there is a strategy for *not*-bad-as-old data, of which good-as-new data are merely a special case. This section describes the not-bad-as-old strategy.

Doganaksoy and Nelson (1998) describe data about two samples of locomotives that have braking grids from two different batches. The following SAS statements create the data set Brakes:

```
data Brakes;
   if _N_ < 40 then Batch = 'Batch1';
   else Batch = 'Batch2';
   input ID$ Days Event @@;
   datalines;
S1-01 462  1     S1-01 730 -1     S1-02 364  1     S1-02 391  1
S1-02 548  1     S1-02 724 -1     S1-03 302  1     S1-03 444  1
S1-03 500  1     S1-03 730 -1     S1-04 250  1     S1-04 730 -1
S1-05 500  1     S1-05 724 -1     S1-06  88  1     S1-06 724 -1
S1-07 272  1     S1-07 421  1     S1-07 552  1     S1-07 625  1
S1-07 719 -1     S1-08 481  1     S1-08 710 -1     S1-09 431  1
S1-09 710 -1     S1-10 367  1     S1-10 710 -1     S1-11 635  1
S1-11 650  1     S1-11 708 -1     S1-12 402  1     S1-12 700 -1
S1-13  33  1     S1-13 687 -1     S1-14 287  1     S1-14 687 -1
S1-15 317  1     S1-15 498  1     S1-15 657 -1     S2-01 203  1
S2-01 211  1     S2-01 277  1     S2-01 373  1     S2-01 511 -1
S2-02 293  1     S2-02 503 -1     S2-03 173  1     S2-03 470 -1
S2-04 242  1     S2-04 464 -1     S2-05  39  1     S2-05 464 -1
S2-06  91  1     S2-06 462 -1     S2-07 119  1     S2-07 148  1
S2-07 306  1     S2-07 461 -1     S2-08 382  1     S2-08 460 -1
S2-09 250  1     S2-09 434 -1     S2-10 192  1     S2-10 448 -1
S2-11 369  1     S2-11 448 -1     S2-12  22  1     S2-12 447 -1
S2-13  54  1     S2-13 441 -1     S2-14 194  1     S2-14 432 -1
S2-15  61  1     S2-15 419 -1     S2-16  19  1     S2-16 185  1
S2-16 419 -1     S2-17 187  1     S2-17 416 -1     S2-18  93  1
S2-18 205  1     S2-18 264  1     S2-18 415 -1
;
```

The variable **ID** identifies the locomotive. The variable **Days** provides the locomotive age in days. The variable **Event** is 1 if the age corresponds to brake grid failure and replacement or –1 if the observation is right-censored. The variable **Batch** is a group variable that identifies the source of the brakes.

Nonparametric Analysis

Nonparametric analysis of this case requires no assumption about the condition of repaired locomotives. They could be bad as old, good as new, or anything in between—it simply doesn't matter. In the absence of any substantive knowledge about repaired units, the nonparametric analysis remains your best bet.

You could proceed by estimating the overall MCF nonparametrically. However, because you have data for two separate batches, you might want to detect a difference in the expected number of failures between the two. The following SAS statements plot the difference in MCF between the two batches and perform statistical tests of equality of MCFs across the batches:

```
proc reliability data=Brakes;
   unitid ID;
   mcfplot Days*Event(-1) = Batch / mcfdiff;
run;
```

Figure 5 graphs the difference in the expected number of failures between the two batches, with 95% confidence bands. Because the estimated function is mostly negative, brake grids from Batch 2 tend to fail more often, especially from Day 200 to Day 300. However, given the confidence bands, the difference between batches might not be statistically significant.

Figure 5 MCF Difference for Brakes Data

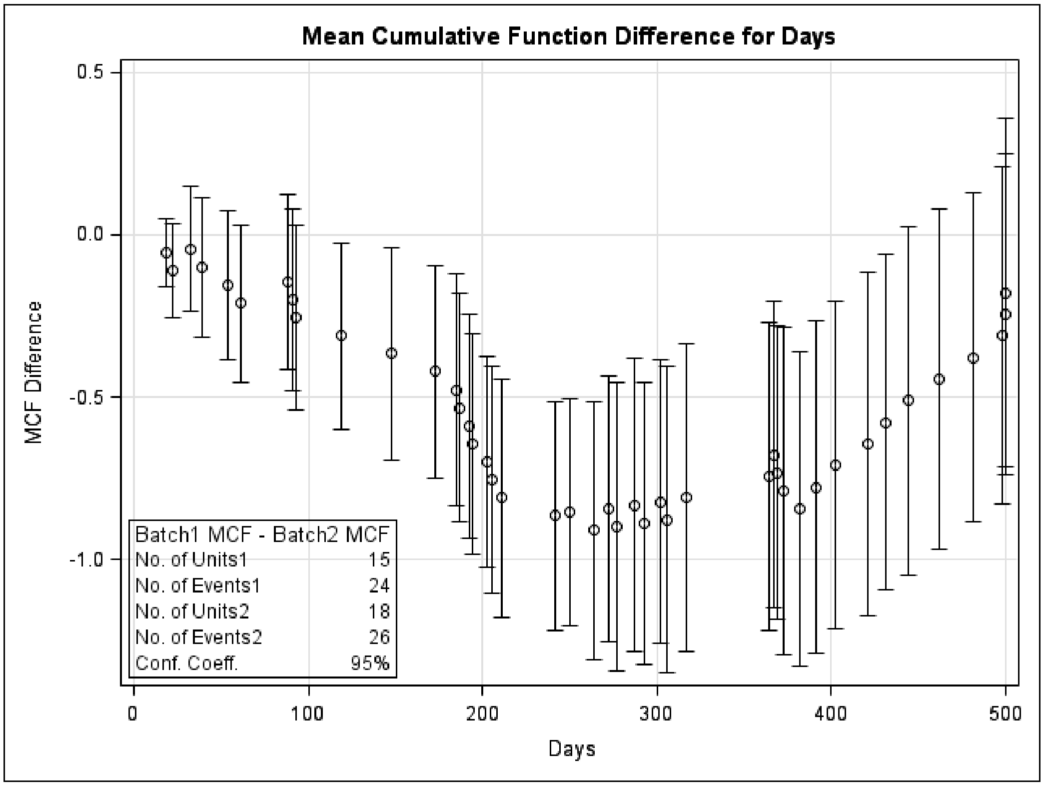


Figure 6 provides two tests of equality of MCFs, but they give conflicting results at the 5% level. At issue is how MCF comparisons are weighted as time progresses and the effective sample size decreases. The “Constant” test treats all comparisons the same, regardless of time; the “Linear” test assigns less weight to later comparisons. Which strategy is optimal is generally difficult to determine, so at best you can say that the results are inconclusive.

Two-sample tests, like the one demonstrated, are an added feature to version 12.1 of SAS/QC.

Figure 6 Test of MCF Difference for Brakes Data

The RELIABILITY Procedure

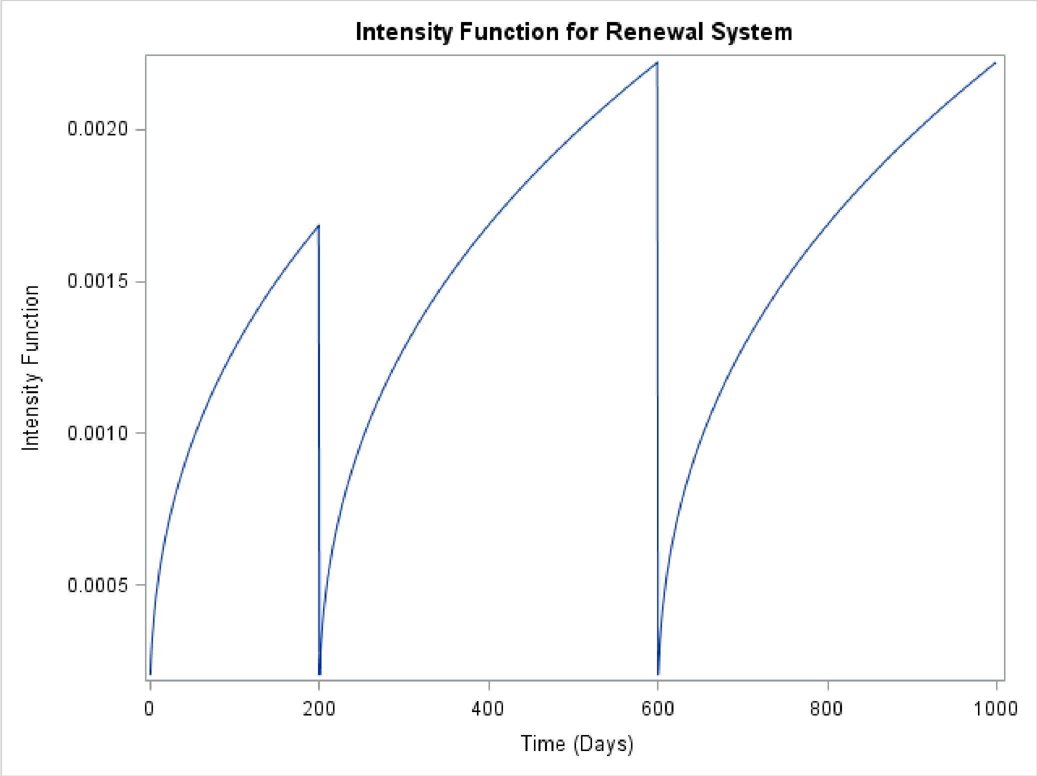
Tests for Equality of Mean Functions

Weight Function	Statistic	Variance	Chi Square	DF	Pr > Chi Square
Constant	-3.673285	4.556053	2.961560	1	0.0853
Linear	-4.435032	1.424770	13.805393	1	0.0002

Parametric Analysis

Parametric analysis of this case requires a model (like the power model in the bad-as-old scenario) for the intensity of failure over time. Refusing to assume bad-as-old behavior and entertaining the notion that risk actually decreases with each repair lead to oddly shaped intensity functions. In the extreme case where repaired systems are good as new, the intensity goes down to 0 at each failure; see Figure 7 for an illustration.

Figure 7 Intensity Function for a Renewal System



These oddly shaped intensity functions (some would say they look like shark fins) are difficult to model directly. Three strategies follow:

1. Perform data manipulation to determine time between failures, and treat each failure as if it were from a distinct unit. This requires accounting for the fact that failure times are no longer statistically independent.
2. Use a flexible piecewise exponential model for the intensity; for more information, see Rigdon and Basu (2000, Section 3.2).
3. Write your own code that implements a customized intensity such as that depicted in Figure 7.

Although these alternative approaches are statistically valid, they are cumbersome to implement and are rarely encountered in the engineering literature. Because the nonparametric approach is always valid, the usual strategy is to forgo a parametric model in situations where bad-as-old repairs cannot be assumed.

SUMMARY

Analyzing data about recurring failures involves a study of the mean cumulative function and possibly the underlying intensity function. When repairs leave units in a bad-as-old condition, you can use either a nonparametric or a parametric analysis. When repairs leave units in a better condition, the nonparametric approach is directly available, whereas a parametric approach is cumbersome.

REFERENCES

Doganaksoy, N. and Nelson, W. (1998), "A Method to Compare Two Samples of Recurrence Data," *Lifetime Data Analysis*, 4, 51–63.

Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: John Wiley & Sons.

Nelson, W. (1995), "Confidence Limits for Recurrence Data—Applied to Cost or Number of Product Repairs," *Technometrics*, 37, 147–157.

Nelson, W. (2002), *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, ASA-SIAM Series on Statistics and Applied Probability, Philadelphia: Society for Industrial and Applied Mathematics.

Rigdon, S. E. and Basu, A. P. (2000), *Statistical Methods for the Reliability of Repairable Systems*, New York: John Wiley & Sons.

SAS Institute Inc. (2012), *SAS/QC 12.1 User's Guide*, Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENT

The author is grateful to Bob Rodriguez of the Advanced Analytics Division at SAS Institute Inc. for his valuable assistance in the preparation of this manuscript.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Roberto G. Gutierrez
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
bobby.gutierrez@sas.com