



Best Practices for Weibull Analysis of Time-to-Failure Data

1 Introduction

Note: Nothing in this standard supercedes applicable laws and regulations.

Note: In the event of conflict between the English and domestic language, the English language shall take precedence.

1.1 Purpose. This procedure provides guidance to GM and Suppliers on best practices for Weibull analysis of product life (time-to-failure) data. In this procedure, Weibull analysis refers to calculating the parameters of a Weibull distribution used to model time-to-failure data, and from that, to determine the reliability of the product.

Whenever test data is fit to a Weibull (or any other) distribution, that distribution is only an estimate of the distribution of the underlying population of parts from which the test parts were taken. The best practices given here are intended to improve the chance of getting a Reliability number that is as close as possible to the population's true Reliability.

There are many different "settings" that can be used when doing Weibull analyses. Unfortunately, each decision typically leads to a different result. GM has preferred or required choices for some settings, and no official position (at this time) for others. The preferred/required settings are based on theory, simulations and other studies made by experts in the field, and observations from past analyses and studies within GM.

1.2 Applicability. The best practices given in this procedure are applicable when the following situations are both true:

- The data being analyzed is observed times-to-failure for a product on a durability test, where the failures are due to a "wear out" type of failure mechanism (where the "damage" to the product builds up over time, eventually leading to a failure). The data may also include:
 - a. The amount of testing performed on products that have not yet failed – these are referred to as right censored or suspended items; and
 - b. A range of times within which a part failed when the exact failure time is not known – this is called interval censoring, or left censoring if the left-most point is at the start of test.
- The portion of the Weibull curve of greatest interest is at the low end of the distribution, where there is a low percentage of parts that have failed, which means high part reliability.

The practices given here may not be the best choices under other situations, such as analyzing customer usage data (where the upper end of the distribution is usually of greater interest), or strength data such as material yield strength (which usually has much higher Weibull slopes, or which may be better fit using a different distribution). Therefore, the practices given here should not be used blindly for other situations.

1.3 Remarks. It is assumed the reader has at least a basic knowledge of how to enter data and to navigate within the Weibull analysis software that their organization uses. The locations of some of the lesser known settings within Weibull++ and Minitab are provided here. Appendix C shows settings in Weibull++ under Application Setup that conform to these best practices. There are no equivalent overall settings in Minitab; instead, the settings are selected within each analysis. Table D1 in Appendix D gives nomenclature and feature availability differences between Weibull++ and Minitab.

2 References

Note: Only the latest approved standards are applicable unless otherwise specified.

2.1 External Standards/Specifications.

None

2.2 GM Standards/Specifications.

GMW17162

2.3 Additional References.

- Reliability Engineering SharePoint - "Weibull Analysis of Test-to-Failure Data.xlsx", located in the "Reliability Tools" folder at: <https://link.gm.com/ReliabilityTools> (ID)
- Information on ways to compare datasets, at:
http://www.reliawiki.org/index.php/Comparing_Life_Data_Sets

(ID) = This website is limited to internal access within General Motors and cannot be accessed outside this company. Contact the GM Validation Engineer for further instructions.

3 Resources

3.1 Facilities.

3.1.1 Calibration. The test facilities and equipment shall be in good working order and shall have a valid calibration label.

3.1.2 Alternatives. Alternative test facilities and equipment may also be used. However, all measuring variables as specified in this standard shall be determined correctly with respect to their physical definition.

3.2 Equipment. Not applicable.

3.3 Test Vehicle/Test Piece. Not applicable.

3.4 Test Time. Not applicable.

3.5 Test Required Information. Not applicable.

3.6 Personnel/Skills. Not applicable.

4 Procedure

4.1 Preparation. Not applicable.

4.2 Conditions.

4.2.1 Environmental Conditions. Not applicable.

4.2.2 Test Conditions. Deviations from the requirements of this standard shall have been agreed upon. Such requirements shall be specified on component drawings, test certificates, reports, etc.

4.3 Instructions. Not applicable.

5 Data

5.1 Calculations. Use of the Best Practices given in Section 5.1 is required when performing Weibull analysis of product life data for General Motors, for validation to the product's reliability requirements.

5.1.1 Minimum Number of Failures. Although Weibull analysis can be done with fewer failures, it should be avoided if there are fewer than three failures, or if the fraction of failed parts to total parts on test is less than about one-third ($\frac{1}{3}$). Testing should strive to run long enough to get this minimum number of failures (running long enough to get a majority of the parts to fail, so the bottom half of the curve is covered, is preferred). Note that if done appropriately, degradation analysis, covered in GMW17162, may be used to predict future failure times – past the end of testing – and these times can count toward this minimum number of failures.

A variant of the Success Test Formula that allows for a limited number of failures is available in the Tools folder of the Reliability Engineering SharePoint site at <https://link.gm.com/ReliabilityTools>. It is on the "Success Test with Failures" tab of the "Weibull Analysis of Test-to-Failure Data.xlsx" file. This tool should be used when there is only one failure, and is an alternative to using Weibull analysis when there are two failures. This is a GM-Restricted tool – suppliers should ask their contact in the GM Validation department to analyze such data. Use the Weibull slope provided by the Validation Technical Integration Engineer (TIE).

5.1.2 Removal of Failure Points. Part failures may not be removed from the analysis without approval from the GM Validation TIE. Proof that a part was tested incorrectly (e.g., due to a test fixture problem) is generally an acceptable reason to remove the part from the analysis, but "the part was built wrong" is not.

5.1.3 Inclusion of Suspensions. All censored data (including left, interval and right censoring) should be included, with the possible exception of unfailed parts removed from test very early. These add little real

information, and any suspensions earlier than the first failure can be omitted. The analyses will take all entered points, including suspensions, into account. If the software is capable of showing the censoring times on the plot (Weibull++ can, Minitab cannot at this time), they should be shown. Note that if analyzing data from customer-owned or fleet vehicles, the current times or mileages on all unfailed vehicles must be included as suspensions (right censored points).

5.1.4 Two-Parameter Only. Always use a 2-parameter (2-P) Weibull distribution, never 3-parameter (3-P). This means that the minimum life value (or location parameter in some software) is set to zero. Use of a 3-P analysis would require all of the following:

- A sample size greater than about 50 with more than half tested to failure;
- The physics/chemistry behind the failure mechanism proves there is a quantifiable period within which it is absolutely impossible to have a failure (simply being a wear out failure mechanism is not proof); and
- A 3-P analysis provides a significantly better fit to the data, with many points (not just a couple) all suggesting a downturn in the curve.

This situation is so unlikely that GM does not plan to accept any 3-P analyses of product life data.

5.1.5 Least Squares Method. When doing a Rank Regression (RR) analysis (also called a Least Squares (LS) analysis), always use the method that minimizes the errors (actually, the sum of the errors squared) between the line and the data points in the x (time) direction. In Weibull++, this is called Rank Regression on X (RRX) and in Minitab is called Least Squares X on Y (LSXY). The alternative, which is not to be used, minimizes the error in the y (percent failed) direction, and is called Rank Regression on Y (RRY) in Weibull++. At this time, Minitab does not have a comparable option.

5.1.6 Least Squares vs. Maximum Likelihood Estimation. *This Best Practice is for selecting which type of analysis is the default method to try first. Neither of the two methods is guaranteed to work the best for a given set of data, so if the calculated Weibull slope using the default method is outside the range of 1.5 to 6.0, the Best Practice in 5.1.7 must also be followed.*

A Rank Regression/Least Squares analysis should generally be used initially.

However, Maximum Likelihood Estimation (MLE) should be tried first under either of the following two situations. *This is because a Rank Regression analysis uses the suspension times only to determine the ranking (y-axis plotting location) of the failed parts. MLE uses the actual suspension times.*

- There are any suspensions that are significantly (20% or more) further out in time than the last failure time. MLE accounts for the actual suspension times, RRX only accounts for its order.
- There are multiple (three or more) different times at which unfailed parts were suspended. This is called progressive right censoring. MLE makes better use of multiple suspension points than RRX can. Note that customer-owned and fleet vehicles will have mostly suspensions, at many different suspension times (typically based on vehicle time-in-service or mileage), so a MLE analysis is the default method for that data.

Also try MLE first if both of the following two conditions apply. These are, individually, weak indicators to try MLE first, but together they make the case for using MLE stronger.

- Any of the suspensions are before the last failure.
- There are multiple (two or more) failed parts whose times-to-failure have a relatively large amount of uncertainty (e.g., where a part was okay when last checked at 100 h, but had failed at an unknown point before the next inspection at 150 h), and the data is being entered as left or interval censored data.

In addition, certain analysis options in Weibull++ are only available by using MLE. Examples include the option to enter a slope and recalculate the characteristic life given that slope; the life comparison tool; and the use of contour plots of the Weibull parameters. These techniques will be discussed later. Note that some of these are available in Minitab using Rank Regression/Least Squares. However, MLE should be used.

Note that a Weibull line generated with an MLE analysis often will not look like it fits the data points very well when plotted. This by itself does not mean that MLE is a worse choice. The poor visual fit is because the y-axis locations of the plotted data points are not used in an MLE analysis. An MLE analysis selects the Weibull parameters to maximize the chance that the observed lives came from the selected distribution, but it does this without using the y-axis values used in a Rank Regression analysis. It would be more accurate to say that the plot is wrong, than the analysis.

Figure 1 summarizes the above decisions.

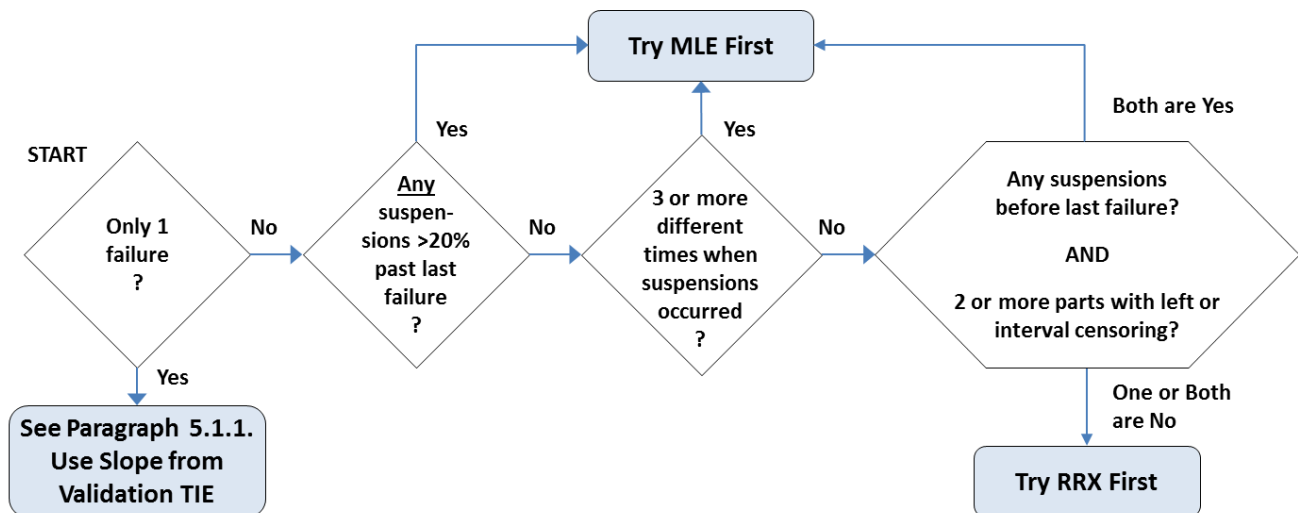


Figure 1: Flowchart for Picking Initial Weibull Analysis Method

5.1.7 Adjusting Slopes for Test-to-Failure Data. *This Best Practice is for adjusting the calculated Weibull slope when analyzing failure times from a Test-to-Failure plan, and need to determine the demonstrated reliability. See 5.1.8 for developing a Weibull slope to be used to set up a Success Testing validation plan.*

Several factors conspire to produce unrealistic Weibull slopes:

- With smaller sample sizes, the slope determined in a Weibull analysis of test data can vary significantly from the slope of the full population of parts. The calculated slope can be either too high or too low, and is due to the fact that when parts can have large differences in their lives (an order of magnitude or greater difference is common in a population), one can, just by chance, get several test parts that are unusually close (resulting in a high slope) or unusually far apart (giving a low slope) in their times to failure.
- If there are suspensions in the dataset, this increases the uncertainty, and almost always raises the calculated slope.
- In addition, the MLE analysis method is inherently biased toward determining a Weibull slope that is too high, meaning the slope for an individual data set will usually have a higher slope than the underlying population of parts. To a much lesser extent, Rank Regression will also tend to give slopes that are too high on average.

Whenever a slope less than 1.5 or greater than 6.0 has been calculated when using the initial calculation method (5.1.6), the process (Figure A1) and table (Table A1) given in Appendix A must be used. The purpose is to use a slope that is closer to what the overall population's slope would be. Studies performed by GM (see Appendix B) show that the accuracy of the estimated reliability at one test life will be improved in many cases using this process.

Note: Weibull++ has an analysis option called "unbias parameters" which attempts to correct the slope in an MLE analysis, to address point (c). However, work conducted by GM shows that the correction factors in Weibull++ (at least through version 10.1.1) are incorrect for low sample sizes, especially when there is a high proportion of suspensions. ReliaSoft has been notified of this issue. Until the factors used in Weibull++ are changed, GM does not approve using this feature. Use the factors in Table A1 of Appendix A instead.

5.1.8 Slopes for Success Testing. *Use this Best Practice if setting up a Success Test validation plan and need a Weibull slope to determine possible trade-offs between sample size and test length.*

Due to inaccuracies in the slope from small sample sizes, the slopes from a prior phase of testing should not generally be used when calculating the sample size and test length options for later success testing. The GM Validation TIE for each subsystem has default slopes to use for success test planning, and has final approval on the use of any alternative slope value. The TIE has access to a tool in the Reliability Engineering SharePoint site for deciding on alternative slope values.

This tool uses a process similar to the one given in Appendix A, but also accounts for the existing default slope value. In addition, slopes outside the range of 1.5 to 3.0 will not be considered for setting up a success test plan,

as data on very large, complete data sets (with no suspensions, which greatly improves the accuracy) show that true population slopes for wear out failure mechanisms will generally be within this range.

An alternative slope should only be considered if life data exists for a very large number of failures. This may be from a single group of parts, or could be from several slightly smaller groups. If averaging the slopes from multiple datasets, the geometric average, not an arithmetic average, should be used (multiply the n slopes together, then take the n^{th} root, or use the GEOMEAN function in Excel). Prior to doing that, however, the Best Practices in 5.1.6 and 5.1.7 should be followed for each individual dataset.

5.1.9 Multiple Failure Mechanisms. When there are multiple failure mechanisms, or widely-varying locations of failure of the same mechanism, there is a choice between analyzing all the times-to-failure together without differentiating the failure mechanisms or locations, or to analyze each mechanism or location separately. Unless there were enough parts tested to failure to give at least three failures of each mechanism or location, all the failures should simply be analyzed together.

If there are at least three failures of each, then a judgment should be made as to which method gives the best fit to the data. If a single analysis of all the data gives a good-looking fit, but one or more of the separate failures gives a poor fit, use the single analysis. On the other hand, if each separate fit looks good but the single analysis looks worse, use the separate fits.

When using the separate fits, the results must then be combined to get an overall reliability at the time of interest. To do this, determine the reliability at that point in time for each separate failure mechanism or location, and multiply them together. That is,

$$R_{\text{Overall}} = R_1 \times R_2 \times \dots \times R_N$$

Where:

R is the Reliability and

N is the total number of failure mechanisms or locations analyzed.

In Weibull++, one can easily analyze separate mechanisms or failure locations using a “Competing Failure Modes” analysis. The Subset ID column is used to label the different failure mechanisms or locations. Each line will be fit individually, and the overall reliability line will also be generated. The overall reliability at one life can be found using the Quick Calculation Pad. Minitab has a similar function, although it does not calculate an overall reliability that combines the results from the individual analyses. To analyze all the mechanisms at once in Minitab, use the FMode settings, select the column that gives the labels, and select Use All Failure Modes.

Note that if any of the individual analyses gives a slope outside the range of 1.5 and 6.0, that particular set of data should go through the process in 5.1.7 before combining all the analyses to give an overall reliability. This means having to ignore the overall curve in Weibull++’s Competing Failure Modes analysis.

5.1.10 Grouped Data. If “grouped” life data is input (multiple parts with the same life or suspension point entered on one line, with the number of parts entered in a separate column), always select to have it ungrouped for purposes of the analysis and plotting. This will make the program include all the points in the analysis, as opposed to treating them as a single point. Ungrouping reduces the uncertainty in the calculated Weibull parameters. This is done with the “ungroup on regression” and “ungroup on MLE” checkboxes in the Weibull++ (and the Accelerated Life Testing Analysis (ALTA) program) settings. In Minitab, under Graphs settings, select Handle Tied Failure Times by Plotting All Points.

5.1.11 Two-Sided vs. One-Sided Confidence Bounds. Two-sided confidence intervals (bounds) are used to compare multiple distributions to see if they are statistically different from each other. A one-sided lower confidence bound on reliability is used to compare a part’s demonstrated reliability to its required minimum reliability (i.e., for validating the part). Two-sided confidence intervals are made up of the area above a one-sided lower bound and below a one-sided upper bound. Be careful of the confidence numbers: a 90% two-sided interval is between a 95th percentile lower bound (sometimes also called a 5th percentile bound, when the word “lower” is omitted) and a 95th percentile upper bound. In this example, there is a 5% chance the true Weibull line is below the lower bound, another 5% chance it is above the upper bound, leaving a 90% chance it is within the interval formed by the two bounds.

When comparing the results of the Weibull analysis to a reliability requirement that is stated with a confidence of 50% (more correctly stated, with a 50% one-sided lower bound on the reliability estimate, and often

abbreviated as C50), that bound is the estimate (best-fit) line. Confidence bounds do not need to be calculated or shown for this purpose.

5.1.12 Confidence Bounds Algorithms. There are many algorithms for computing confidence bounds on the Weibull line. Two of the most commonly used are the Fisher Matrix (FM) and the Likelihood Ratio (LR) methods. Additional methods that may be available in some Weibull analysis programs, are not approved for GM use. The following information is from ReliaSoft:

“The Fisher Matrix method is the standard way of calculating confidence bounds, and it is the default method in the software [Weibull++]. For data sets with a large number of data points, there is no significant difference in the results between the Fisher Matrix method and the Likelihood Ratio method; however, when working with very small sample sizes, the Likelihood Ratio method may be more accurate.”

Weibull++ has both methods available, while Minitab currently has only Fisher Matrix. Because GM states its reliability requirements at 50% confidence, GM generally only uses confidence bounds for comparing two distributions (see 5.1.14, part (a)). For that purpose, it is not critical which method is used since both distributions will have their confidence bounds determined the same way.

5.1.13 Bounds on Life vs. Bounds on Reliability. Another choice when showing confidence bounds is whether to use bounds on the Reliability (to see uncertainty in the y-axis direction) or bounds on the Time (to see uncertainty in the x-axis direction). In the plot legend in Weibull++, the “[R]” in “CB@90% 2-Sided [R]” indicates the confidence bounds are on the Reliability, while “[T]” would be on Time. These generally give different results when using the Fisher Matrix method, but the same result when using the Likelihood Ratio method. Since GM is typically measuring life and calculating reliability, bounds on the reliability would be the appropriate choice. Weibull++ will do either. Minitab only gives bounds on the life in its plots, although one can get bounds on percent failed (which is 1-R) in a table by selecting “Confidence limits for cumulative failure probabilities” under the Storage settings. For comparing distributions, it is not critical which one is used, but bounds on reliability are preferred if available.

5.1.14 Comparing Products. Comparing two distributions is a common use of Weibull analysis. The goal of this is to determine if one distribution is statistically better or worse than the other, or if they are similar enough (given the uncertainty in Weibull fits from small sample sizes) that they cannot legitimately be distinguished from each other. There are three main ways to do this in Weibull++ (Method A, B and C). Minitab can do Method A.

- A. Preferred method: Plot the estimate line and two-sided confidence bounds of both distributions together on one plot (an overlay multiplot in Weibull++), then look to see if each estimate line is inside the confidence bounds of the other distribution. The most important part of the plot to look at is the area at one life and below, as this represents the usage and exposure the parts will see in most customer vehicles. If each estimate line is inside the confidence bands of the other distribution, then one cannot say the distributions are different, to the level of confidence selected. Typically, 70% or 90% two-sided bounds on the reliability (per 5.1.11 through 5.1.13) are used. See Figure 2 for an example. In this example, the confidence bounds were set at 80%, which was the highest setting that still showed both distributions to be outside the bounds of the other distribution below one life. In other words, with 80% confidence, the two distributions are different below one life, and that in the range of most customer usage, Data1 (in blue) is better than Data2 (in red).
- B. Plotting the contour plots of the two distributions' Weibull parameters on the same plot (using an overlay multiplot and selecting Contour as the plot type). Again, 70% and/or 90% confidence are typically selected, but here, the confidence is on the parameter values, not on the estimate lines. If the contour plots partially overlap, then, to the confidence selected, there is no distinction between the two distributions. This method requires using an MLE analysis. See Figure 3 for an example. This is for the same failure data as in Figure 2, but note that when using the same 80% confidence setting, the lines overlap. So at 80% confidence, it cannot be said that the two distributions are different. In contrast to method A, where the focus is on the area to the left of one life, method B looks at the entire distribution.
- C. A Life Comparison Test or Stress-Strength Interference analysis, both available in Weibull++. Since they give the same result, any tool that does Stress-Strength Interference can be used. One calculation method is to randomly select a data point out of each distribution and determine if the point from one distribution is higher. This is repeated thousands of times, and the results summarized. Most programs, however, perform Stress-Strength Interference using numerical integration instead of randomly selecting points. With either method, if the chance that a random point from one distribution is higher than a random point from the other distribution is near 50%, then it cannot be said that the two distributions are significantly different. If the

© Copyright 2016 General Motors Company All Rights Reserved