Use of Censor Data in Reliability Analysis

One is absolutely correct that non-failed engine and or engines that fail by a different failure mode should be included in the Weibull analysis as "censored or suspended" units. This data cannot and should not be ignored. The times or suspended units must be included in the analysis. Life data in general is censored. That is, the exact failure times of some units are unknown; however, one should know the following:

- The failure occurred after a particular time (a suspension or right-censored observation)
- The failure occurred before a particular time (a left censored observation)
- The failure occurred between two particular times (an interval-censored observation)

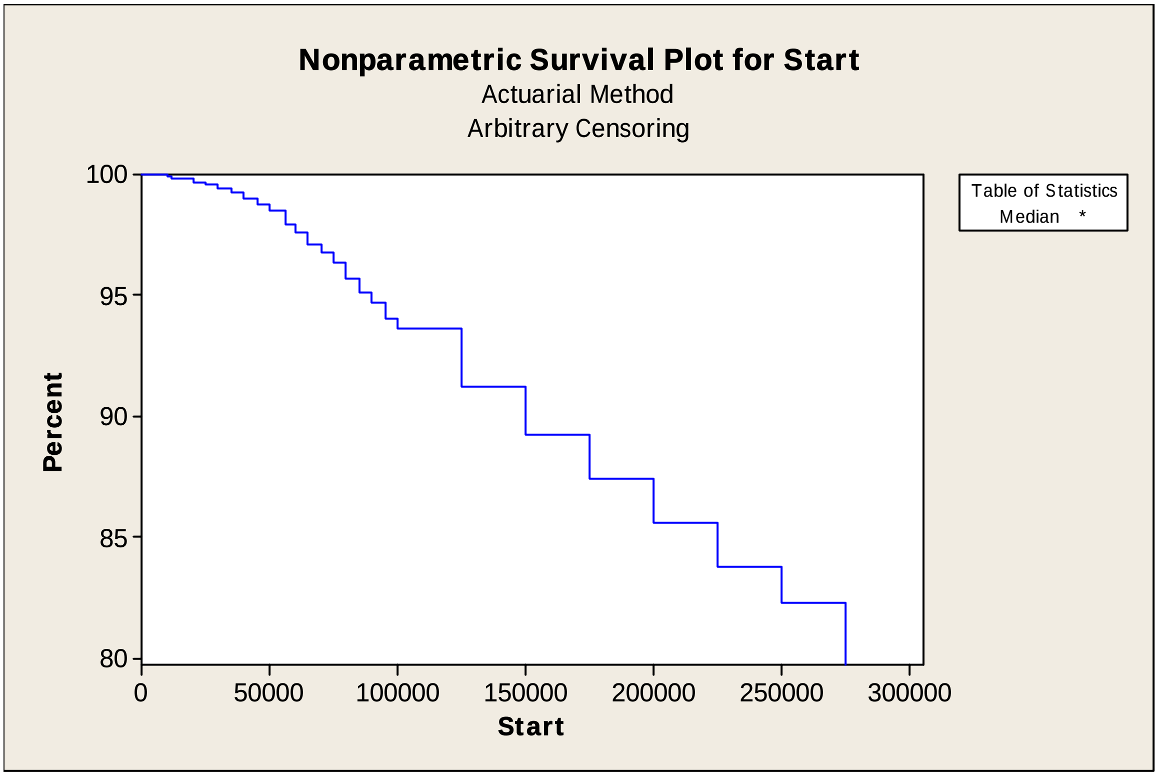
It is very important to include the censored observations in one's analysis because the fact that these units have not failed would greatly impact one's reliability estimates. In order to do that we would need censored data (number of units still functioning in field and no of days each unit has been in field).

For instance, one auto company in India is having problems with the head gaskets. In one of their particular engines, the cylinder head that sits on the engine block has a gasket to prevent the coolant from leaking. Warranty claims indicate that leaky cylinder head gaskets consistently cause failures of these engine models. Engineers were interested in determining what percentage of these gaskets would survive past the 80,000 mile warranty period. They wanted the reliability for the gaskets to be at least 95%.

In the attached appendix, one can see that there are engines that are still in the field and have not yet failed due to leaky gaskets (these would be considered right-censored observations or suspensions). The other question you had raised was how one correlate time in field to actual usage time; hopefully this example would clarify that question also. In this case, though the exact mileage fro these engines are not known, the mileage interval is based on the date of sale.

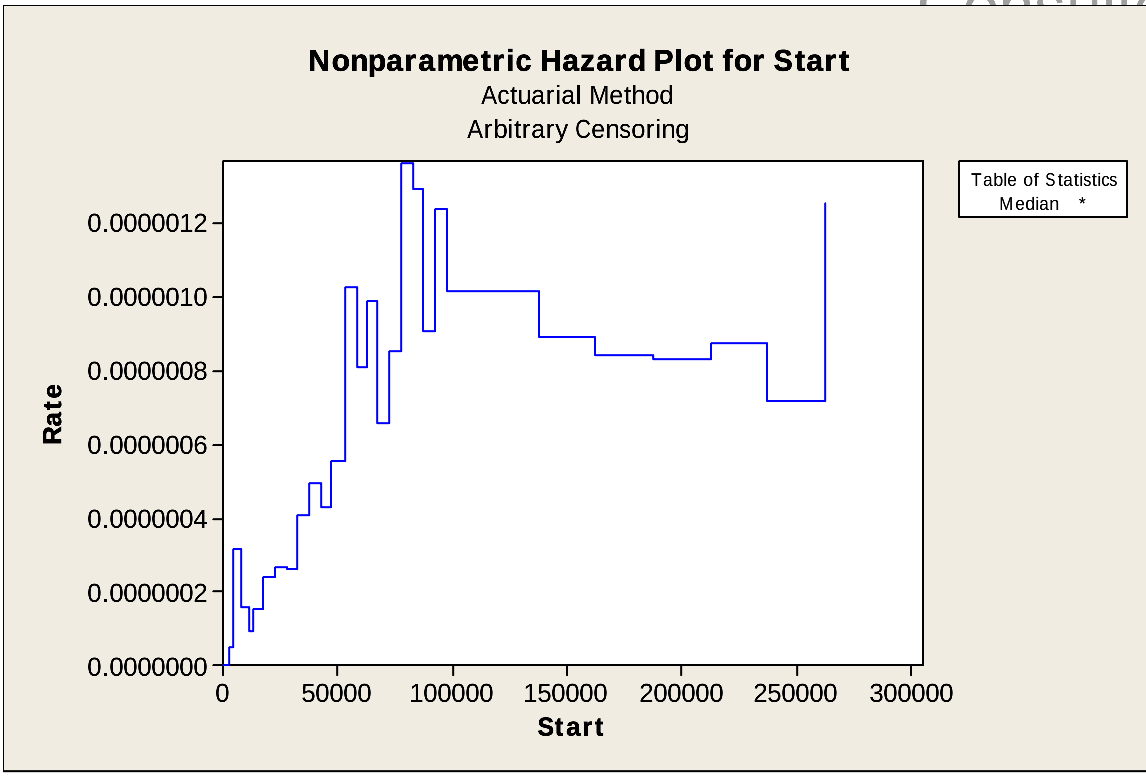
For example, in the 0 to 1000 mile interval, there were 787 suspensions. One does not know the exact mileage for gaskets that have not yet failed. One only knows that the mileage at the time the data were colleted was somewhere between 0 and 1,000 miles. One needs to choose a time within that interval to represent when the suspension actually occurred. Here, we used the midpoint of the mileage interval was used for suspensions occurring within that interval and is recorded in the start column.

To analyze this type of data, I have used actuarial method. Since I have used nonparametric method, survival plot shows a "step function". With non-parametric methods, changes in the survival or the hazard function only occur at actual times or at the midpoints of intervals where failure occurs. Had I used the parametric method, it would have provided smooth estimate of these functions.



I like to use non-parametric survival plot to display the estimate of the reliability across time, without imposing any particular distribution on the data. But, if I had to extrapolate outside the range of this data, I would have used parametric estimation procedure.

Similarly, the nonparametric hazard plot lets one look at the hazard function without depending on chosen distribution. In this case, the rate of failure first increase then levels off. By looking at the shape of this hazard function, I would have used lognormal distribution.



Since the engineers were interested in the 80,000 mile reliability for the gasket, the true reliability of is 95.94%.

Time	Survival Probability	Standard Error	95.0% Normal Bound Lower
2000	1.00000	0.0000000	1.00000
3000	0.99995	0.0000495	0.99987
4000	0.99990	0.0000709	0.99978
5000	0.99959	0.0001463	0.99935
10000	0.99880	0.0002557	0.99838
12000	0.99862	0.0002766	0.99816
15000	0.99817	0.0003250	0.99763
20000	0.99697	0.0004354	0.99625
25000	0.99563	0.0005419	0.99474
30000	0.99432	0.0006382	0.99327
35000	0.99229	0.0007759	0.99101
40000	0.98983	0.0009278	0.98831
45000	0.98771	0.0010518	0.98598
50000	0.98498	0.0012060	0.98300
56000	0.97893	0.0015175	0.97644
60000	0.97577	0.0016690	0.97303
65000	0.97096	0.0018922	0.96785
70000	0.96777	0.0020385	0.96441
75000	0.96365	0.0022284	0.95998
80000	**0.95709**	**0.0025237**	**0.95294**
85000	0.95091	0.0027940	0.94631
90000	0.94659	0.0029823	0.94169
95000	0.94075	0.0032378	0.93542
100000	0.93598	0.0034474	0.93031
125000	0.91244	0.0045226	0.90500

150000	0.89233	0.0056164	0.88309
175000	0.87378	0.0068626	0.86249
200000	0.85576	0.0083442	0.84204
225000	0.83727	0.0101894	0.82051
250000	0.82242	0.0119778	0.80272
275000	0.79705	0.0154506	0.77164

One last point, in this case this was a field data, and since field data is essentially a "snapshot" in time, one should amend the analysis as one gets more data, so as to make sure that the analysis remains substantially remain accurate.

APPENDIX 1

Start	End	Freq	
0	1000	*0*	
1000	2000	*0*	Engines that have already failed (italic)
2000	3000	*1*	
3000	4000	*1*	
4000	5000	*6*	
5000	10000	*14*	
10000	12000	*3*	
12000	15000	*7*	
15000	20000	*17*	
20000	25000	*17*	
25000	30000	*15*	
30000	35000	*21*	
35000	40000	*23*	
40000	45000	*18*	
45000	50000	*21*	
50000	56000	*42*	
56000	60000	*20*	
60000	65000	*28*	
65000	70000	*17*	
70000	75000	*20*	
75000	80000	*29*	
80000	85000	*25*	
85000	90000	*16*	
90000	95000	*20*	
95000	100000	*15*	
100000	125000	*59*	
125000	150000	*33*	
150000	175000	*20*	
175000	200000	*13*	
200000	225000	*9*	
225000	250000	*5*	
250000	275000	*6*	
500*		787	Engines that have not failed
1500*		593	
2500*		559	
3500*		541	
4500*		521	
7500*		2102	
11000*		769	
13500*		1030	
17500*		1542	
22500*		1380	
27500*		1191	
32500*		1031	
37500*		886	

42500*	840
47500*	743
53000*	728
58000*	471
62500*	523
67500*	449
72500*	453
77500*	373
82500*	359
87500*	278
92500*	266
97500*	250
112500*	949
137500*	629
162500*	376
187500*	247
212500*	155
237500*	96
262500*	69
275000*	154