SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853

TECHNICAL REPORT NO. 890

February 1990

(Revised October 1990)

PROBABILITY PLOTTING WITH CENSORED DATA

by

Lance A. Waller and Bruce W. Turnbull

This research was supported in part by grant GM 28364 from the U.S. National Institutes of Health.

AUTHOR FOOTNOTE

Lance A. Waller is a graduate student and Bruce W. Turnbull is Professor, School of Operations Research and Industrial Engineering, 206 E&TC, Cornell University, Ithaca, NY 14853-3801. This research was supported in part by grant GM 28364 from the U.S. National Institutes of Health. The authors are grateful to the referees for their comments and suggestions.

MS 90132R

"Probability Plotting with Censored Data"

ABSTRACT

Graphical methods for testing goodness-of-fit are reviewed and classified as transformations of P-P and Q-Q plots. Survival and cumulative hazard plots for censored data, probability plots, and generalized residual plots are included. Difficulties in interpretations arise when the points are grouped more closely together in particular areas of the plot. This problem commonly occurs in cumulative hazard and Q-Q plots, especially in large data sets, but can occur in P-P plots when censoring is present. An empirical rescaling of the axes is proposed to overcome this problem. The techniques are applied to reliability and clinical trial data sets.

KEY WORDS: Survival Analysis, Graphical Methods, P-P plots, Q-Q plots, Hazard Plots, Reliability, Generalized Residuals

1. INTRODUCTION

The use of plots to assess validity of distributional assumptions and to check goodness-of-fit is common practice among statisticians. Cox (1979, page 188), commenting on the many graphical methods of comparing data with the exponential distribution, stated that: "Most are transformations of one another so that choice is partly a matter of taste." Indeed, most graphical methods are transformations of the theoretical quantile-quantile plots discussed, for example, in Chapter 6 of Chambers, Cleveland, Kleiner, and Tukey (1983), where a table is presented that lists the familiar percentage-percentage (P-P) and quantile-quantile (Q-Q) plots as well as the stabilized probability plot (SPP) of Michael (1983) and others. We expand this table to include the plots described by Elandt-Johnson and Smith (1989) in connection with generalized residuals (Cox and Snell, 1968), as well as the cumulative hazard plots of Nelson (1969, 1972). We also point out a limitation of the current plots when censored data are present. Such a situation suggests a new type of plot, which we add to the list.

Difficulties in interpretation arise in large data sets where plotted points are so dense that individual ones cannot be distinguished (e.g. "hidden" points in SAS graphical output). Such a situation can unduly influence the impression of the plot by giving equal visual weights to separate portions of a plot that contain widely differing proportions of the plotted points. This unbalanced representation can be caused either by the choice of the scales of the axes, or by the presence of censoring in the data. The scaling problem has been noted and various solutions have been suggested by Chambers et al. (1983), Michael (1983), Cox and Oakes (1984), and Elandt-Johnson and Smith (1989). However, these solutions do not compensate for the "bunching" due to censoring and hence lose their effectiveness as the number of censored observations increases. In Section 4 we suggest an empirical rescaling of both axes to give each point equal visual weight in the case of censoring. The methods are developed via a series of examples.

2. REVIEW AND CLASSIFICATION OF PLOTS

Suppose $y_1<y_2<\ldots<y_n$ represent the ordered values of an i.i.d. sample of size n from a continuous distribution F. We wish to test the hypothesis, H_0, that $F(\bullet) = F_0(\bullet,\theta)$ where F_0 is specified completely or up to some finite dimensional parameter θ. For convenience, assume F_0 is strictly monotone so that F_0^{-1} is uniquely defined over its range. Define $\hat{F}_0(y) = F_0(y;\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate (MLE) of θ. For uncensored data, define $\hat{F}$ as the empirical cumulative distribution function; in the presence of multiple right censoring, $\hat{F}$ is the Kaplan–Meier (1958) estimator. Finally, let $\hat{\Lambda}_0$ be the cumulative hazard function associated with $\hat{F}_0$, i.e.

$$\hat{\Lambda}_0(y) = -\ln[1-\hat{F}_0(y)],$$

and $\hat{\Lambda}$ be a nonparametric estimate of the cumulative hazard, either

$$\hat{\Lambda}(y) = -\ln[1-\hat{F}(y)],$$

or the Nelson (1972) estimator

$$\hat{\Lambda}(y) = \sum_{i\,:\,y_i \le y} \frac{\delta_i}{n - i + 1}$$

where δ_i is the censoring indicator corresponding to observation y_i (with $\delta_i = 1$ for y_i exact). Of course, these two nonparametric estimators, $\hat{\Lambda}$, are approximately equivalent (Miller, 1981, page 66.)

If the data do indeed come from the distribution F_0, then plots of y versus $\hat{F}_0^{-1}(\hat{F}(y))$ should approximate a straight line with slope 1 through the origin. This relationship, as well as various transformations of it, provide a method of graphically testing goodness of fit. Miller (1981, page 164) describes a "basic principle" of graphical methods:

> "Select the scales of the coordinate axes so that if the model holds, a plot of the data resembles a straight line, and if the model fails, a plot resembles a curve."

Most graphical methods for testing H_0 compare the plotted points with the straight line through the origin with unit slope, with any qualitative deviation giving reason to reject H_0. Note that with censored data the plotted points correspond to the uncensored observations since the empirical estimates of the distribution function only change at these values (Nelson 1969, 1972, 1982, 1990, Chambers et al. 1983 page 234). Thus the number of plotted points equals the number of uncensored observations, n_u say $\left(n_u = \sum_{i=1}^{n} \delta_i \right)$. Table 1 summarizes many of the graphical probability plotting methods for testing goodness-of-fit and includes the stabilized probability plot (SPP) (Michael 1983, Kimber 1985), and the cumulative hazard plot of Nelson (1969, 1972). It can be seen that any one plot can be simply obtained from any other by the appropriate transformation of the axes. We exclude consideration of those plots which are not simply transformations of the probability plots, such as the total-time-on-test plot (Barlow and Campo, 1975) and the Lorenz curve (Horvath and Yandell, 1988). These two plots do not necessarily have a linear representation of the null hypothesis, and are awkward to use with multiply censored data.

[Table 1 about here.]

In fact, for the plots (b), (f), and (g) which involve $\hat{F}_0^{-1}$ or $\hat{\Lambda}_0^{-1}$, the last point can be off the scale when i = n. Therefore, for exact data, we use plotting positions $\tilde{F}(y_i) = \frac{i}{n+1}$ instead of the sample cumulative distribution function, $\hat{F}(y_i) = \frac{i}{n}$. Other plotting positions have been proposed, including the family:

$$\tilde{F}(y_i) = \frac{i - c}{n - 2c + 1}$$

for particular values of c ($0 \leq c \leq 1$). Popular choices include c = 0 or c = 0.5. Harter (1984) reviews the reasons for the various choices of plotting positions. For censored data, Michael and Schucany (1986) proposed an analogous modification of the Kaplan-Meier estimator, namely

$$p(y_i) = 1 - \frac{n-c+1}{n-2c+1} \prod_{\delta_j=1, j\leq i} \frac{n-j-c+1}{n-j-c+2}$$

However, it should be noted that for the purpose of assessing goodness of fit, the choice of plotting positions makes little qualitative difference in the appearance of any particular plot.

Simultaneous confidence bands can be added to P-P, Q-Q, and cumulative hazard plots to allow more quantitative assessment of the distributional assumptions, see Nair (1981), Michael (1983), and Csorgo and Horvath (1986).

The u-, e-, and w-plots were suggested by Elandt-Johnson and Smith (1989) who defined these for generalized residuals, calling them u-, e-, and w-residual plots, respectively. The u-plot is equivalent to the P-P plot and, in the absence of censoring, the plotted points are evenly spread along the line of unit slope. This follows from the fact that, under H_0, the probability integral transformation is being used for both axes, and so abscissa and ordinate values represent samples from the uniform distribution on [0,1]. The e-plot is equivalent to a Q-Q plot, in the case of generalized residuals, since F_0 is completely specified as the unit exponential, i.e. $\hat{F}_0(y) = F_0(y) = \exp(-y)$. The w–plot will cause increased compression of the plotted points and will not be considered further here.

Note that if our empirical estimators are defined such that $\hat{\Lambda}(y) = -\ln[1-\hat{F}(y)]$ then $\hat{\Lambda}_0^{-1}(\hat{\Lambda}(y)) = \hat{F}_0^{-1}(\hat{F}(y))$ so that the cumulative hazard plot is the same as the Q–Q plot (Nair 1981, Michael and Schucany 1986, Nelson 1990). Note that some authors (e.g.Chambers et al., 1983) interchange the axes on the Q-Q plot. Some texts define the cumulative hazard plot as

$$y \qquad \text{vs.} \qquad \hat{\Lambda}(y).$$

This is the same as plot (f) in the unit exponential case when $\Lambda_0(y) \equiv y$, the identity function. Otherwise, it corresponds to (f) only when the ordinate scale is adjusted appropriately by $\hat{\Lambda}_0^{-1}$. This transformation of the scale is equivalent to Nelson's (1969, 1972, 1990) use of the appropriate probability or hazard paper.

An advantage of plot (b), the Q-Q plot, is that the plot remains linear for location/scale shift distributional families, obviating the explicit need to compute $\hat{\theta}$. However, due to the intrinsic non-linear relationship between quantiles and their associated probabilities for a given distribution, the Q-Q plot has points with abscissa and ordinate values distributed non-uniformly, causing the points to "bunch". The P-P plot, and its parallel confidence band version, the SPP (Michael 1983, Kimber, 1985), will have the points uniformly spread along both axes in the absence of censoring, but can still have "bunched" points when censoring is present. We now illustrate the ideas in this section with several examples.

3. AIRCRAFT AIR-CONDITIONING DATA

We begin by considering a familiar example with no censoring, namely the aircraft air-conditioning data set originally used by Proschan (1963). This data set, consisting of $n = 213$ inter-failure times, has been used extensively in the literature and an exponential distribution with $F_0(y,\theta) = 1-\exp(-y/\theta)$ has often been used as a model. Figure 1(A) gives the Q-Q plot for the air-conditioning data. Of course, from the equivalence of the Q-Q and cumulative hazard plots noted previously, Figure 1(A) is essentially the same as Figure 1.3 of Cox and Lewis (1966) who plotted the log survival function, $\ln[1-\hat{F}(y)] = -\hat{\Lambda}(y)$. Figure 1(A) contains a "reference" line showing the theoretical plotting positions, under H_0, when $\theta = \hat{\theta}$. Here $\hat{\theta} = 93.14$ (Cox and Lewis, 1966, page 7). However, computation of $\hat{\theta}$ is not necessary, and the reference line could be fitted by eye, when justified. As might be expected in a Q-Q plot of a large exponential data set, most points fall in the lower quadrant and many are "hidden". To illustrate this fact we indicate on the graph the cumulative proportions of the plotted points. As a result, over one half of the area of the plot is dominated by approximately 10% of the data (the upper tail). The overemphasis of the tail of the distribution in cumulative hazard plots was noted by Cox and Oakes (1984, pp. 56-57). When we transform both axes by $\hat{F}_0$ to give plot (a), the P–P plot, the result is shown in Figure 1(B). In order to construct this plot we do need to first compute the MLE $\hat{\theta}$. The reference line is now simply a line of unit slope passing through the origin. The points are now more evenly distributed and each plotted point has approximately equal visual emphasis on the interpretation of the plot.

[Figure 1 about here.]

4. SKIN CANCER DATA AND EMPIRICAL RESCALING

Next we consider a data set from an interim analysis of an ongoing clinical trial for the prevention of skin cancer (Abu-Libdeh, Turnbull, and Clark, 1990). In the trial, the incidence patterns of skin lesions are modelled as replicated multi-type point processes with random effects and covariates. To test the fit of the model, generalized residuals were computed and compared with the unit exponential distribution, $F_0(y) = 1\text{-}\exp(-y)$. The 770 patients in the trial gave rise to 1,989 residuals of which only 478 were exact, giving a high proportion (76%) of censored observations. The Q-Q plot (equivalent to Abu-Libdeh, et al., 1990, Figure 2(a)) of the 478 generalized residuals corresponding to the uncensored values is presented in Figure 2(A). The plotted points are dispersed unevenly on the graph, many points are hidden, and again over 90% are compressed into the lower quadrant. As before, we next consider plot (a), the P-P plot, which remedied the situation for the air-conditioning data. In this example there are no parameters to be estimated in order to transform the plot as the hypothesized distribution of the generalized residuals is completely specified as exponential with mean 1. The P-P plot is presented in Figure 2(B). We see that, while the plotted points are more evenly spread than in the Q-Q plot, there is still compression of the points into the lower quadrant. The problem remains because of the high proportion of censored data. The air-conditioner data set had no censored observations, so the P-P transformation was sufficient to spread the plotted points uniformly. In the presence of right censoring, the larger observations are more likely to be censored and therefore less likely to be represented among the set of plotted points.

[Figure 2 about here.]

Elandt-Johnson and Smith (1989) encountered the same situation using their u-residual plot (equivalent to the P-P plot, see Table 1) of data from a follow-up study from the Lipid

Research Clinics with a large sample size and a high proportion of censored data. They enlarged the scale on both axes until each point could be individually distinguished, and separated the plot into 5 sections, each on different graphs. This will reduce the number of "hidden" points but the problem of uneven dispersion of plotted points on the graph remains. Furthermore, by splitting the plot into 5 graphs, the overall pattern of the data becomes more difficult to establish.

To obviate these problems, we suggest rescaling both axes with respect to the empirical distribution of the abscissa values (the ordered generalized residuals corresponding to uncensored observations) so that the quantiles of the abscissa values of the plotted points are evenly positioned giving each plotted point equal visual weight. Specifically, we define the empirically rescaled plot (ERP) to be the plot of

$$\hat{F}_u(y) \qquad \text{vs} \qquad \hat{F}_u(\hat{F}_0^{-1}(\hat{F}(y)))$$

or equivalently for the cumulative hazard plot,

$$\hat{F}_u(y) \qquad \text{vs} \qquad \hat{F}_u(\hat{\Lambda}_0^{-1}(\hat{\Lambda}(y))),$$

where $\hat{F}_u$ is defined to be the empirical cumulative distribution function of the points corresponding to uncensored observations (Miller, 1981, pp. 64-65). The plots are applicable to any choice of F_0 (Λ_0) and are easy to compute.

In order to examine the exponentiality of the residuals for the skin cancer data, we display the ERP in Figure 2(C). It can be seen that there are fewer small residual values than one would expect from an exponential distribution with parameter 1. This information was completely hidden in the Q-Q plot, de-emphasized in the P-P plot, and only readily apparent in the ERP. Such information could be used to compare competing regression models as presented by Abu-Libdeh et al. (1990). With heavy right censoring, the ERP is needed to achieve the same level of "spreading" of the plotted points for equal visual emphasis that the P-P plot accomplishes for uncensored data.

5. BEHAVIOR OF THE ERP UNDER H_0

We now demonstrate the behavior of the ERP under the null hypothesis using simulated reliability data. Our first example consists of a set of 1000 failure times generated from the exponential distribution with mean 1. In all of our simulations we used the random number generator contained in the S-plus statistical software (Becker, Chambers, and Wilks, 1988). A second set of observation times from the exponential distribution with mean 0.25 was also generated. An observation was considered censored if its observation time was less than its failure time. Since the expected observation time is substantially less than the expected failure time we will have a high proportion of censoring. For our example, we have only 199 exact observations from the sample of 1000. The Q-Q plot of this data is presented in Figure 3(A). Note that points are again compressed into the lower quadrant and the fluctuation of the tail points detracts from the linearity of the lower 95% of the points. The tail points of the Q-Q plot are those with the highest variance in plotting position (Cox and Oakes, 1984, page 57). Next, we consider the P-P plot presented in Figure 3(B). Note that again the plotted points are still compressed toward the lower quadrant and that the upper tail fluctuation is still apparent. Such variation in the upper tail is unusual in a P-P plot from exact data, but with censoring the ordinate and abscissa values do not uniformly fill the entire interval (0,1) as they do for the exact case. Finally, the ERP for the simulated exponential data is shown in Figure 3(C). The points are evenly spread, the fluctuation of the upper tail is contained, and the points cross the reference line several times. From the ERP, we see no reason to reject a null hypothesis of exponentiality.

[Figure 3 about here.]

The first examples considered were fit to exponential distributions but censored data can arise from any distribution. As an example we consider a sample of 1000 simulated failure times, normally distributed with a mean of 1000 weeks and a standard deviation of 200, and which were observed in batches of 50. The first 50 units were considered to have been observed for 80 weeks, the next 50 for 160 weeks, and so on, ending with the last 50 being observed for 1600 weeks. If the generated random value was greater than its associated observation time, that particular observation was considered censored. Nelson (1969) used a similar method to create simulated data with multiple censoring. Here 406 exact observations resulted, the remaining 594 were censored. Figure 4(A) gives the Q-Q plot for the normal data. Note that most of the uncensored data are concentrated in the center of the graph so that 80% of the plotted points are compressed into approximately the center half of the plot. For the P-P plot, we take F_0 to be $N(1000, 200^2)$. (Alternatively, to check normality, the mean and variance could have been estimated from the data using maximum likelihood (Miller, 1981, Chapter 2).) The P-P plot is given in Figure 4(B). The quantiles are more evenly spaced, though slightly skewed to the left due to the censoring, since right censoring will cause smaller values to be disproportionately represented in the set of uncensored observations. Figure 4(C) presents the ERP for the normal data. The difference between the ERP and the P-P plot for the normal data is not as dramatic as for the skin cancer data above, primarily because the Q-Q plot for normal data is nearly symmetrically compressed about the median value, whereas in the case of exponential data, the effects of right censoring and the spacing of exponential quantiles reinforce each other in compressing the points into the same lower quadrant.

[Figure 4 about here.]

6. BEHAVIOR OF THE ERP UNDER ALTERNATIVES

We illustrate the behavior of the ERP under alternative hypotheses by considering an example of testing for exponentiality when the data come from an underlying distribution which has an increasing failure rate. A sample of 1000 failure times was generated from a Weibull distribution with scale parameter 1, and shape parameter 1.4. We will first consider the plots for exact data. The Q-Q plot is presented in Figure 5(A). The points clearly deviate from the reference line and moreover, the convexity of the points in the Q-Q plot implies an IFR alternative (Doksum and Yandell, 1984, page 593). Figure 5(B) gives the P-P plot for the Weibull data with no censoring. The 1000 points are evenly spread and one notices an 'S'-shape indicating fewer small observations and more large observations than one would expect from exponentially distributed data. Because there is no censoring, the ERP will be essentially the same as Figure 5(B) and is therefore omitted.

[Figure 5 about here.]

We introduce censoring to the Weibull data as we did for the simulated exponential reliability data above. A vector of 1000 observation times was generated from the exponential distribution with mean 0.5 and if the observation time was smaller than the corresponding failure time, that observation was considered censored. The resulting data contained 270 exact observations. The Q-Q plot for the censored Weibull data is shown in Figure 6(A). The points still deviate from the reference line and the convexity of the plotted points can still be seen though it is not as readily apparent as it was for the Q-Q plot of the exact data (note the change of scale between Figures 5(A) and 6(A)). Figure 6(B) gives the P-P plot for the censored data. Note that the plot begins below the reference line and ends above it, crossing at approximately the 75th percentile of the data, but the 'S'-shape in Figure 6(B) is different from that in Figure 5(B). In fact, the 'S'-shape will change with differing amounts of censoring since it will become more compressed for samples with more censoring and less compressed for samples with fewer censored observations. Figure 6(C) shows the ERP for the censored Weibull data. In fact, comparing the 'S'-shape to that in Figures 5(B) we see that, although the number of plotted points changed, the overall shape of the plot did not. Use of the ERP may enable the methods of Gan and Koehler (1990) for plotting families of alternatives on P-P plots with exact data to be extended to samples of censored data. Use of the traditional P-P plot in this situation could be deceiving as the shapes of families of alternative curves would depend on the degree of censoring as well as on the particular alternative distribution. Of course, features of the distribution not in the range of the uncensored observations can never be revealed by any goodness-of-fit technique.

[Figure 6 about here.]

7. DISCUSSION

The simulated examples in Figures 3-6 represent single illustrative examples that are typical of null and non-null situations. A more thorough grasp of the qualitative behavior of the plots in these situations can only come from repeating the exercise many times and examining the resulting plots, in a manner similar to that of Daniel and Wood (1980, Appendix 3A, pages 33-43).

Any of the methods in Table 1 can be adapted to the two-sample problem simply by replacing F_0 by F_1 and F by F_2, where F_1 and F_2 denote the distributions of each sample. They also can be extended to situations involving left-censoring or double censoring (as can arise, for example, from differences of paired data with both values subject to censoring). An empirical estimate of F (Turnbull, 1976) is then used instead of the Kaplan-Meier (1958) estimate (Michael and Schucany, 1986).

The ERP fills a gap in existing goodness-of-fit graphical methods for large data sets with heavy right-censoring. Its visual appearance is less sensitive to the effects of different censoring patterns than for the other plots considered.

Table 1

Graphical Methods for Testing Goodness-of-Fit

Plot	Abscissa	Ordinate
a) P–P (u–plot)	$\hat{F}_0(y)$	$\hat{F}(y)$
b) Q–Q	y	$\hat{F}_0^{-1}(\hat{F}(y))$
c) SPP	$(\pi/2)\sin^{-1}[\{\hat{F}_0(y)\}^{1/2}]$	$(\pi/2)\sin^{-1}[\{\hat{F}(y)\}^{1/2}]$
d) e–plot	$-\ln[1-\hat{F}_0(y)] = \hat{\Lambda}_0(y)$	$-\ln[1-\hat{F}(y)] = \hat{\Lambda}(y)$
e) w–plot	$\ln[\hat{F}_0(y)]$	$\ln[\hat{F}(y)]$
f) cumulative hazard	y	$\hat{\Lambda}_0^{-1}(\hat{\Lambda}(y))$
g) ERP[1]	$\hat{F}_u(y)$	$\hat{F}_u(\hat{F}_0^{-1}(\hat{F}(y)))$

[1]ERP and $\hat{F}_u$ defined in Section 4.

REFERENCES

Abu-Libdeh, H., Turnbull, B. W., and Clark, L. (1990) "Analysis of Multi-type Recurrent Events in Longitudinal Studies: Application to a Skin Cancer Prevention Trial." Biometrics, **46**, to appear.

Barlow, E. B., and Campo, R. (1975) "Total Time on Test Processes and Applications to Failure Data Analysis", in Reliability and Fault Tree Analysis. SIAM, Philadelphia, 451-481.

Becker, R. A., Chambers, J. M., Wilks, A. R. (1988) The New S Language. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Wadsworth, Monterey, CA.

Cox, D. R. (1979) "A Note on the Graphical Analysis of Survival Data." Biometrika, **66**, 188-190.

Cox, D. R. and Lewis, P. A. (1966) The Statistical Analysis of Series of Events. Methuen, London.

Cox, D. R. and Oakes, D. (1984) Analysis of Survival Data. Chapman and Hall, London.

Cox, D. R. and Snell, E. J. (1968) "A General Definition of Residuals." Journal of the Royal Statistical Society, Series B, **30**, 248-275.

Csorgo, S. and Horvath, L. (1986) "Confidence Bands from Censored Samples." Canadian Journal of Statistics, **14**, 131-144.

Daniel, C. and Wood, F. S. (1980) Fitting Equations to Data. 2nd edition, Wiley, New York.

Doksum, K. and Yandell, B. S. (1984) "Tests for Exponentiality" in Handbook of Statistics 4: Nonparametric Methods. Krishnaiah, P. R. and Sen, P. K. eds., North-Holland, Amsterdam, 579-611.

Elandt-Johnson, R. C. and Smith, F. B. (1989) "Graphical Generalized Residuals in Fitting Distributions: Applications to Epidemiological Follow-up Data." Statistics in Medicine, **8**, 703-723.

Gan, F. F. and Koehler, K. J. (1990) "Goodness-of-Fit Tests Based on P-P Probability Plots." Technometrics, **32**, 289-303.

Harter, H. L. (1984) "Another Look at Plotting Positions." Communications in Statistics -- Theory and Methods, **13**, 1613-1633.

Horvath, L. and Yandell, B. S. (1988) "Convergence Rates for the Bootstrapped Product Limit Process." Annals of Statistics, **15**, 1155-1173.

Kaplan, E. L. and Meier, P. (1958) "Nonparametric Estimation from Incomplete Observations." Journal of the American Statistical Association, **53**, 457-481.

Kimber, A. C. (1985) "Tests for the Exponential, Weibull, and Gumbel Distributions Based on the Stabilized Probability Plot." Biometrika, **72**, 661-663.

Michael, J. R. (1983) "The Stabilized Probability Plot." Biometrika, **70**, 11-17.

Michael, J. R. and Schucany, W. R. (1986) "Analysis of Data from Censored Samples" in Goodness-of-Fit Techniques. D'Agostino, R. B., and Stephens, M. A. eds., Marcel Dekker, Inc., New York, 461-496.

Miller, R. G. (1981) Survival Analysis. John Wiley and Sons, New York.

Nair, V. N. (1981) "Plots and Tests for Goodness of Fit for Randomly Censored Data." Biometrika, **68**, 99-104.

Nelson, W. (1969) "Hazard Plotting for Incomplete Failure Data." Journal of Quality Technology, **1**, 27-52.

Nelson, W. (1972) "Theory and Applications of Hazard Plotting for Censored Failure Data." Technometrics, **14**, 945-966.

Nelson, W. (1982) Applied Life Data Analysis. John Wiley and Sons, New York.

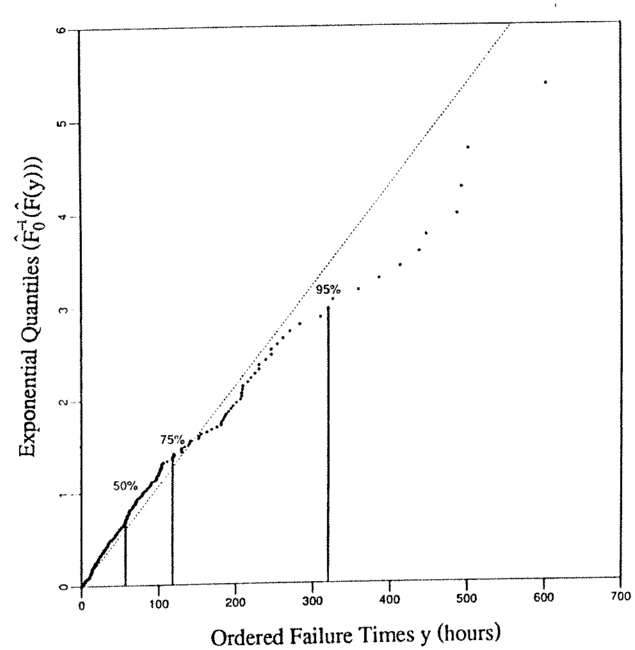
Nelson, W. (1990) Accelerated Testing. John Wiley and Sons, New York.

Proschan, F. (1963) "Theoretical Explanation of Observed Decreasing Failure Rate." Technometrics, **5**, 375-383.

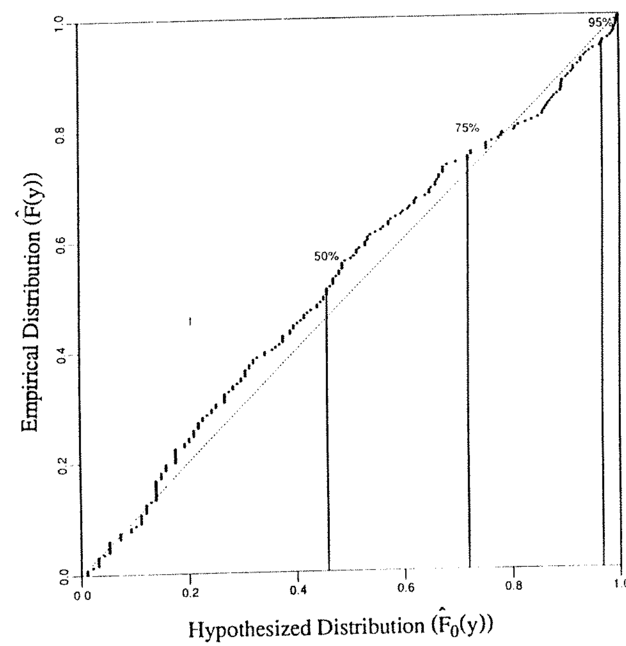
SAS, SAS Institute Inc., Box 8000, Cary, NC, 1982.

Turnbull, B. W. (1976) "The Empirical Distribution Function with Arbitrarily Grouped, Censored, and Truncated Data." Journal of the Royal Statistical Society, Series B, **38**, 290-295.

Figure 1. Probability plots for aircraft air-conditioning data.

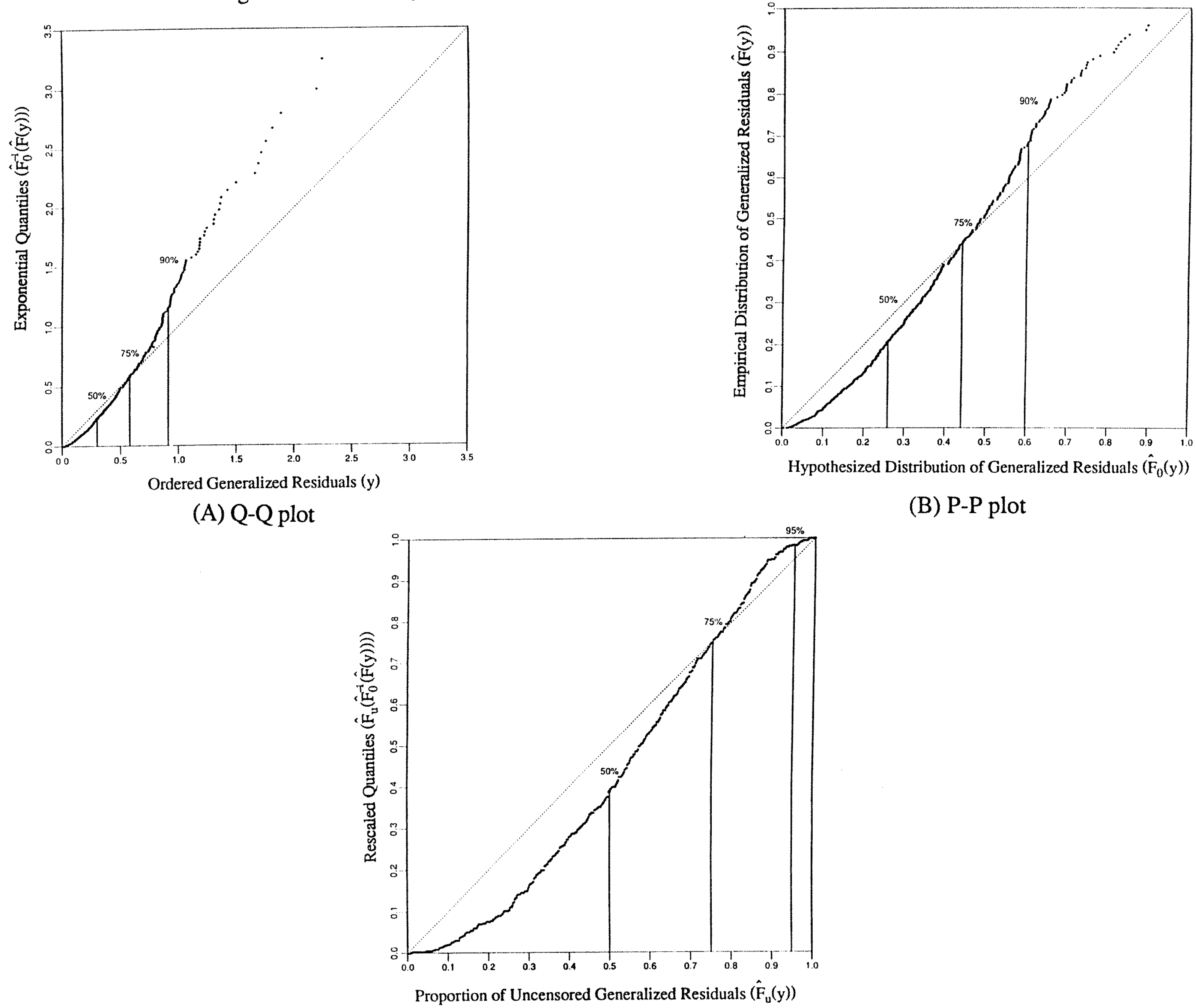


(A) Q-Q plot



(B) P-P plot

Figure 2. Probability plots for generalized residuals from skin cancer data.

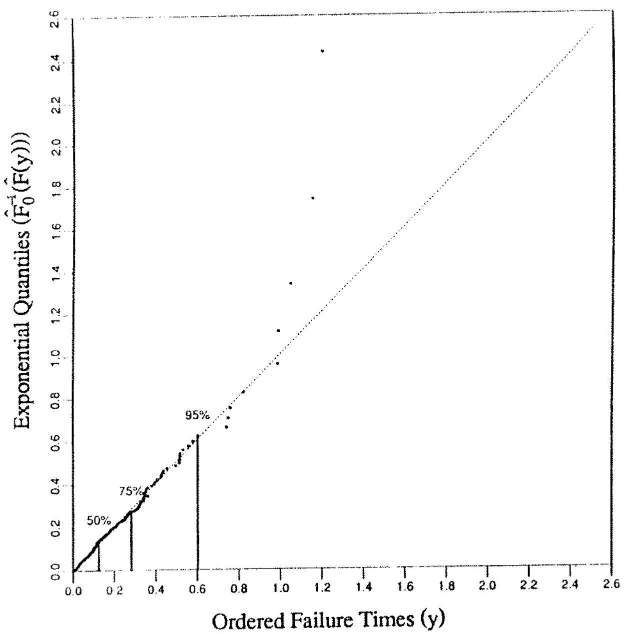


(A) Q-Q plot

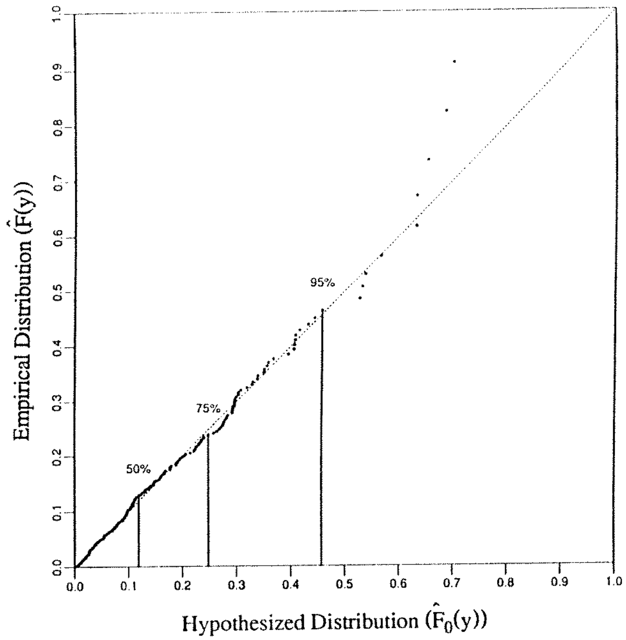
(B) P-P plot

(C) Empirically Rescaled plot

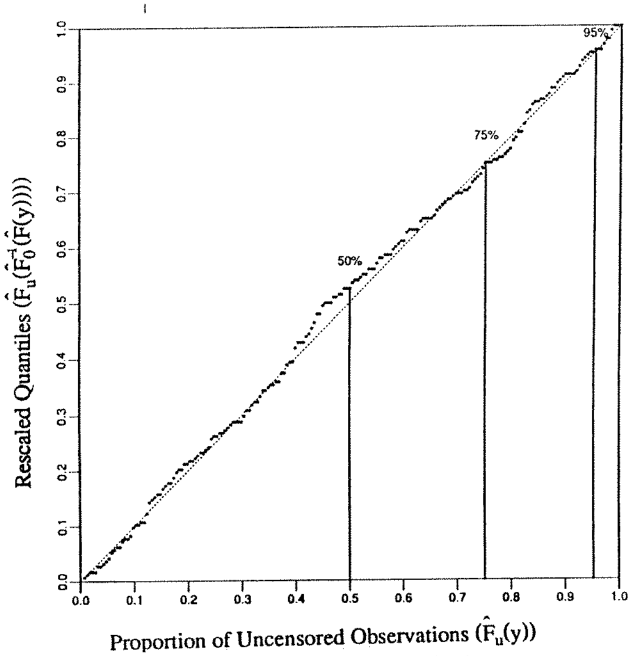
Figure 3. Probability plots for simulated censored exponential failure time data.



(A) Q-Q plot

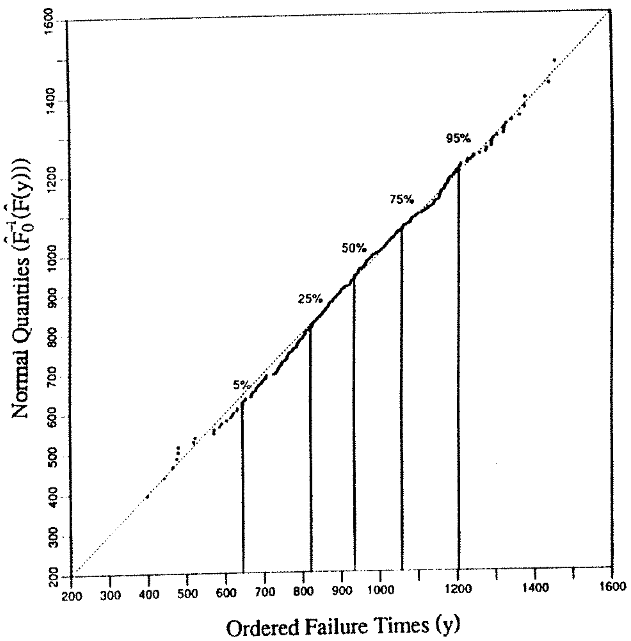


(B) P-P plot

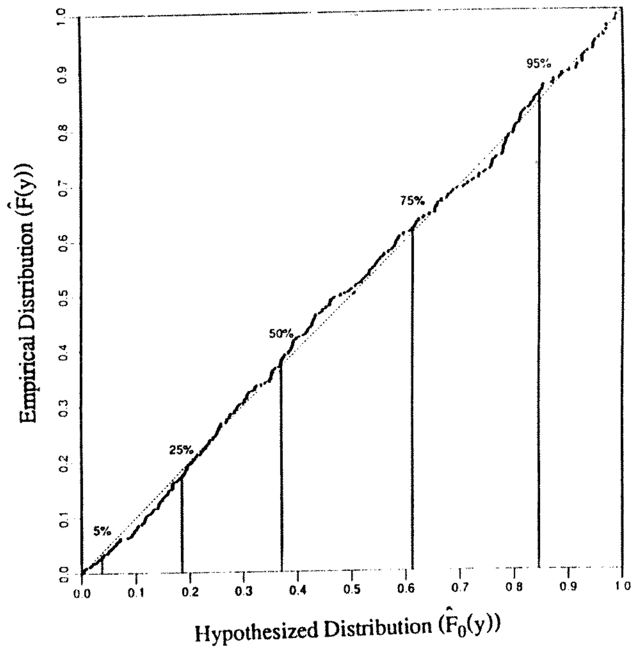


(C) Empirically Rescaled plot

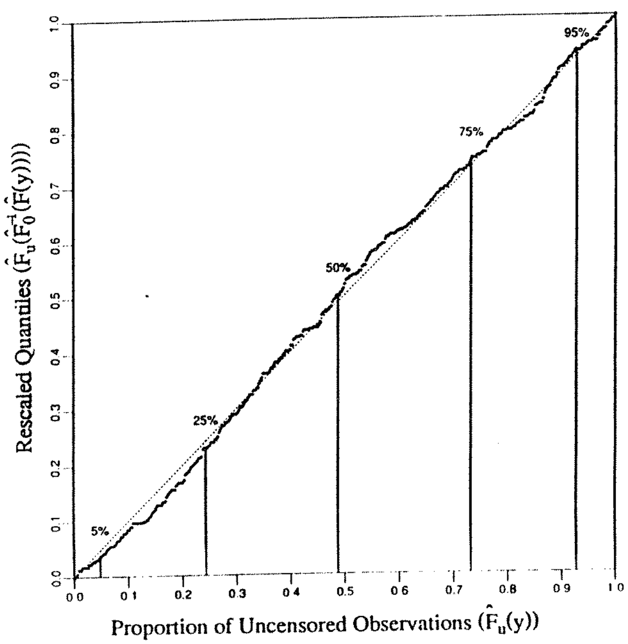
Figure 4. Probability plots for simulated censored normal data.



(A) Q-Q plot

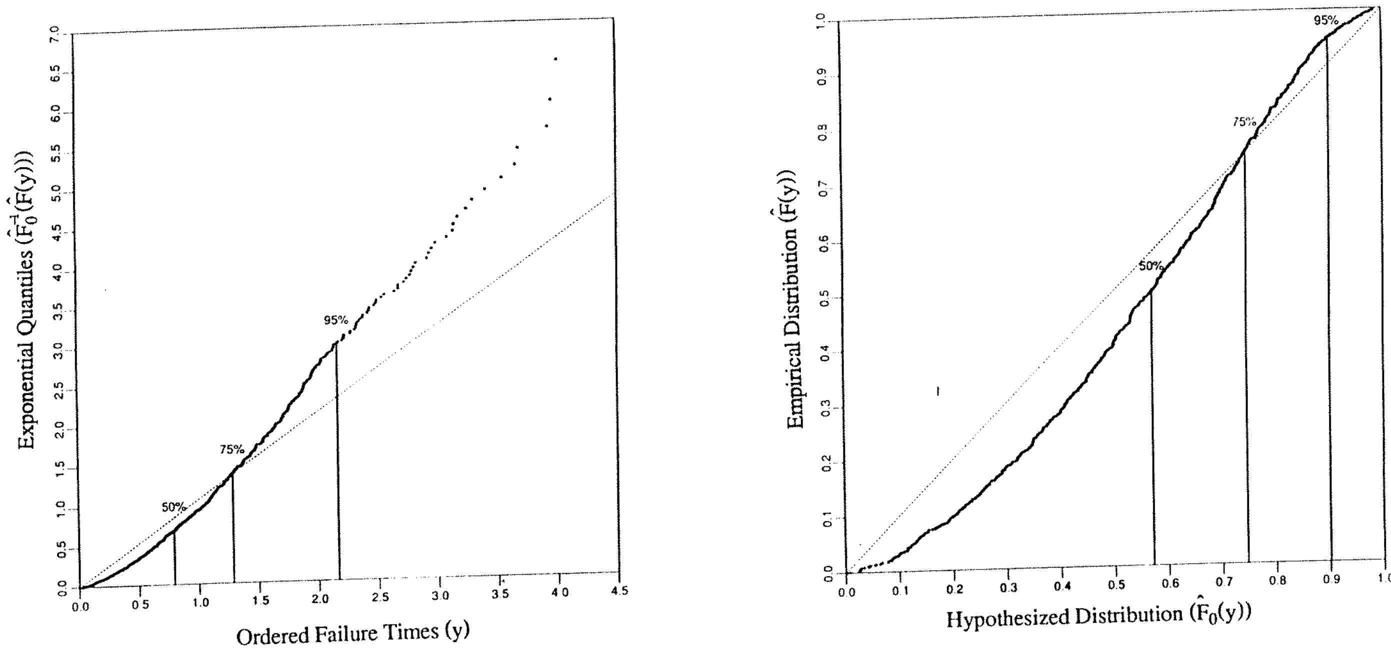


(B) P-P plot



(C) Empirically Rescaled plot

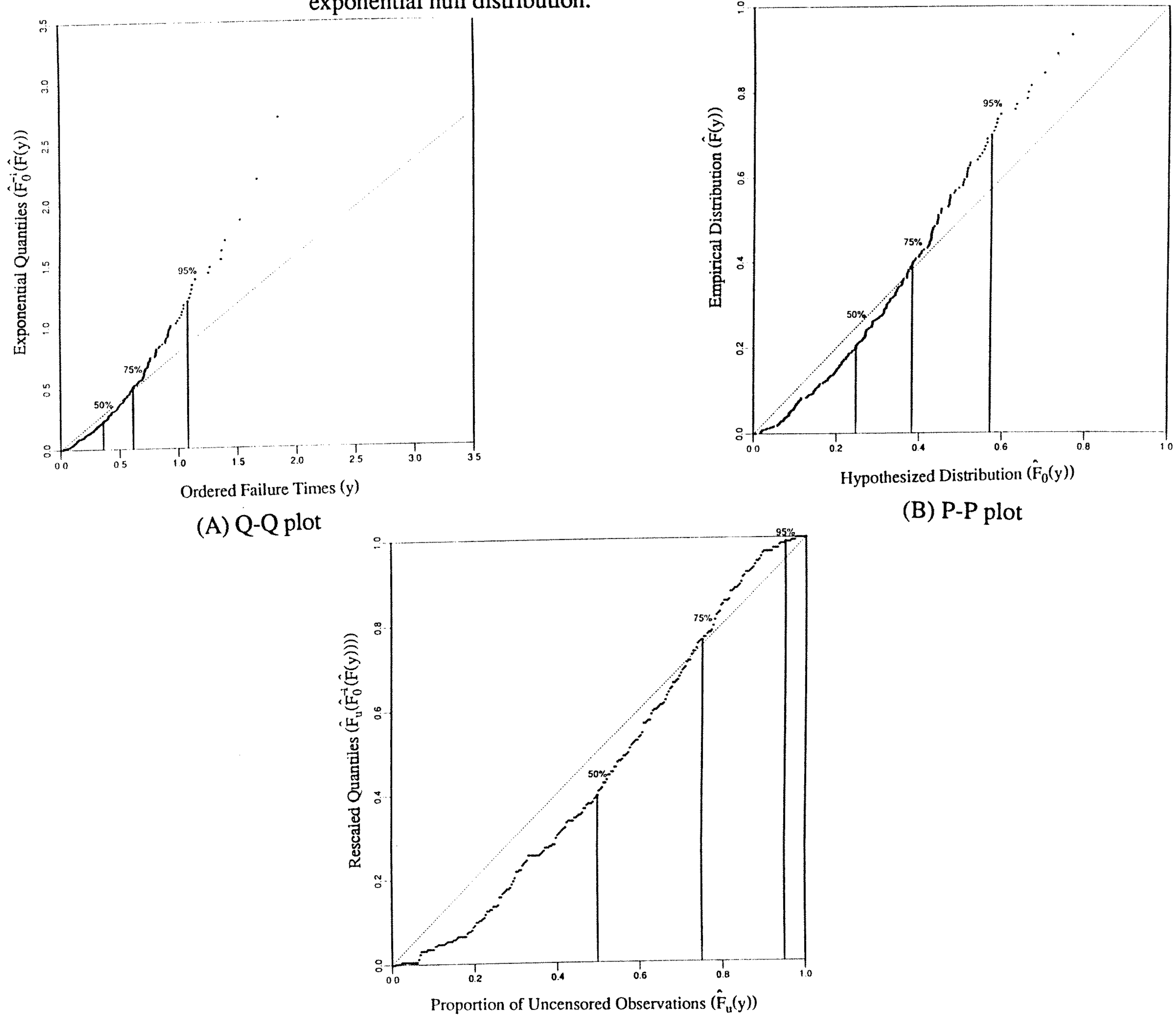
Figure 5. Probability plots for simulated uncensored Weibull data with hypothesized exponential null distribution.



(A) Q-Q plot

(B) P-P plot

Figure 6. Probability plots for simulated censored Weibull data with hypothesized exponential null distribution.

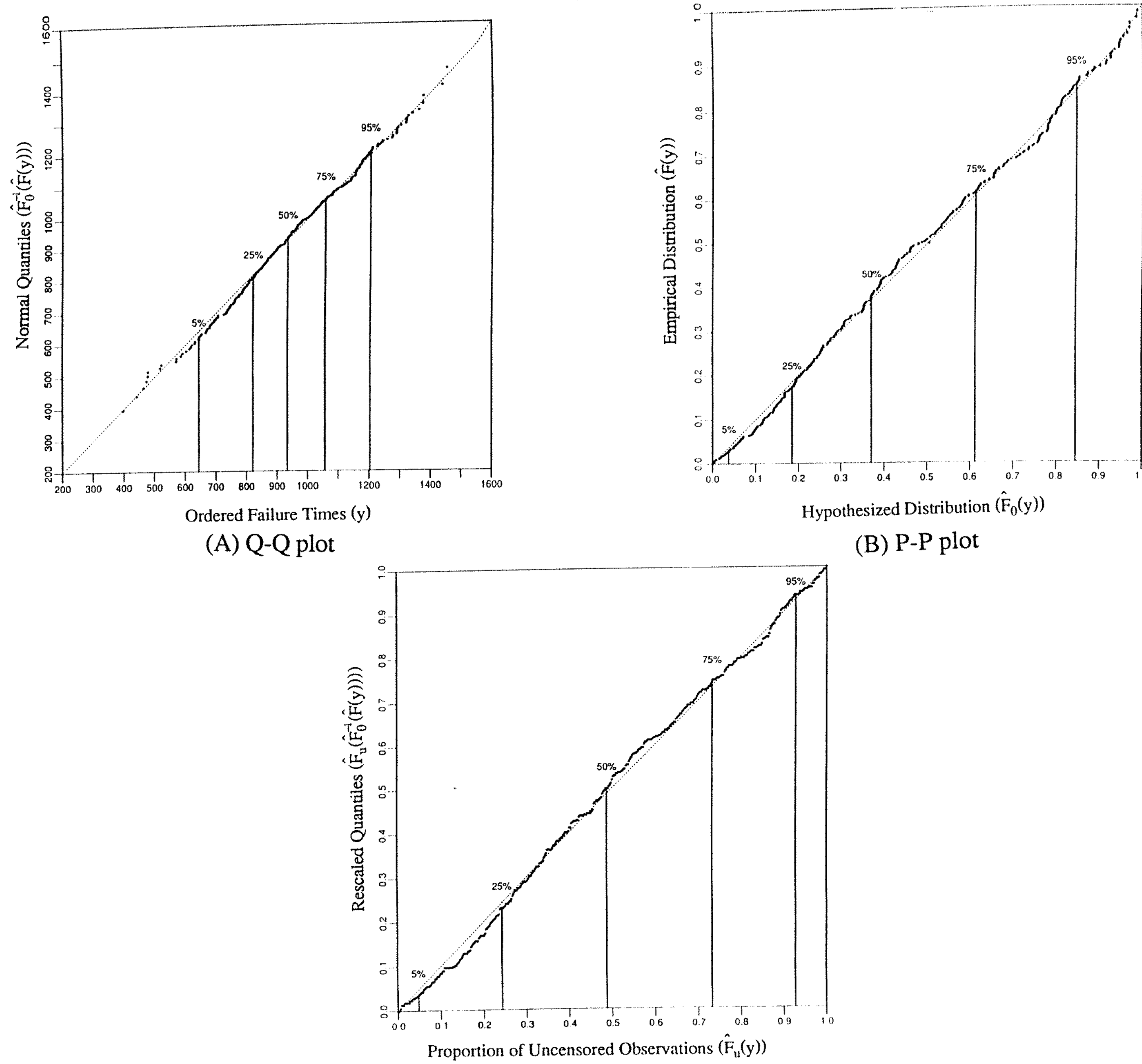


(A) Q-Q plot

(B) P-P plot

(C) Empirically Rescaled plot

Figure 4. Probability plots for simulated censored normal data.

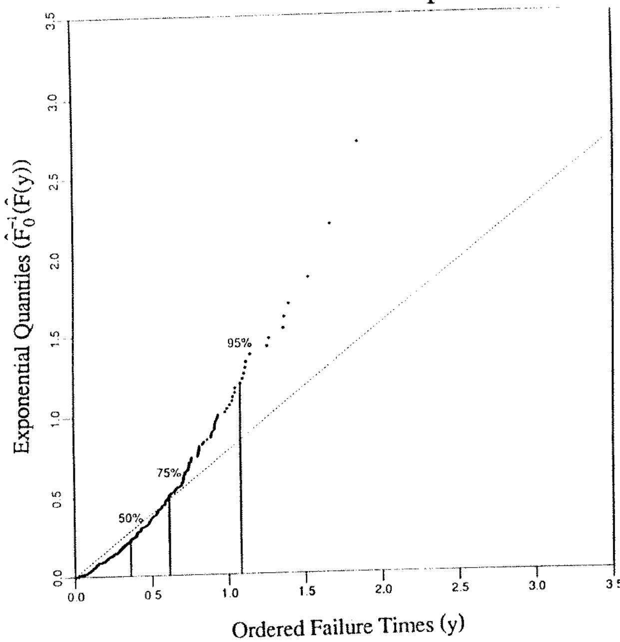


(A) Q-Q plot

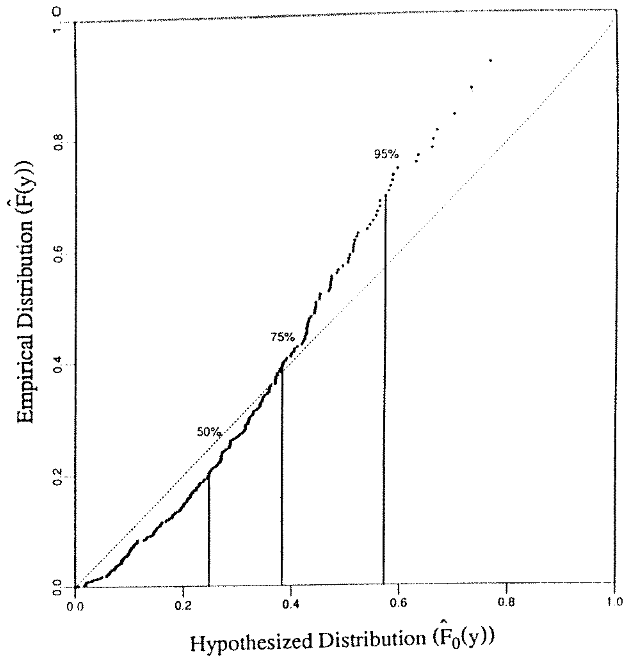
(B) P-P plot

(C) Empirically Rescaled plot

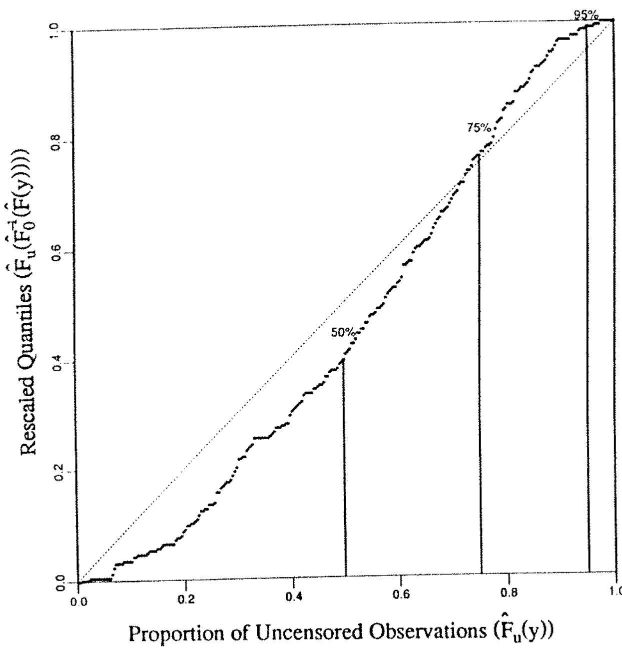
Figure 6. Probability plots for simulated censored Weibull data with hypothesized exponential null distribution.



(A) Q-Q plot



(B) P-P plot



(C) Empirically Rescaled plot