PREDICTING EFFECTIVENESS OF WOOD PRESERVATIVES FROM SMALL SAMPLE FIELD TRIALS

Carol L. Link

Mathematical Statistician

and

Rodney C. DeGroot

Research Plant Pathologist
U.S. Department of Agriculture, Forest Service
Forest Products Laboratory[1]
One Gifford Pinchot Drive
Madison, WI 53705-2398

(Received January 1989)

ABSTRACT

Field tests of wood preservatives use groups of stakes treated at various retentions. Although only an average value is reported for a given group of stakes, the lifetime of individual stakes is quite variable. This paper explores presentations of data that reflect such variability. We also consider the feasibility of predicting the effectiveness of a preservative before all stakes fail. For sample sizes of ten replicate stakes, we suggest that reports include box plots of the actual failure times, and that studies use the sample median for the reported lifetime value rather than the sample mean and report the first quartile as a lower bound for the population average.

Keywords: Box plots, censored data, nonparametric statistics.

INTRODUCTION

Wood has considerable natural variability in its properties. Engineering design values account for this feature in different ways. For example, grading rules for identified wood species quote a strength value that is near a low percentile of the strength distribution, whereas the stiffness value is closer to the population average. However, current procedures for field testing preservative-treated wood do not include variability as a major component of the data interpretation.

The effectiveness of wood preservatives is usually summarized by sample mean values. For example, Gjovik and Gutzmer (1989) report sample lifetime mean values, and Colley (1970) and Hartford (1972) report mean stake ratings over time. Ideally, statistical statements should be made about the population rather than the sample mean. The analysis should also include a measure of variability. Consumers might be more interested in the minimal lifetime of a treated stake rather than the average lifetime. However, as a retention of a wood preservative is often tested with a series of ten replicate stakes, little information exists for the lower part of the distribution of stake lifetimes. Data from tests of different preservatives and retentions cannot be readily combined to give distributional

[1] The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

information that would be valid for an individual preservative and retention. With ten replicate stakes, only average trends are usually discussed. The questions are: Does the current method of reporting best describe the materials on test and Is this method an adequate predictor of the performance of treated wood products that might subsequently go into service?

The purpose of this paper is to explore alternative methods of presenting information about preservative-treated stakes in field trials to reflect data variability. Using parametric and nonparametric procedures, we will also explore the feasibility of using early failures in a group of replicate stakes to predict the sample median of that group. We define failure as a stake destroyed by decay and/or termites [rated 5 or E by the Forest Products Laboratory (FPL) rating scheme; Gjovik and Gutzmer 1989]. However, other definitions of failure could be used, such as time to significant decay and/or termite attack—a log score of 7 (as used by Colley 1970) or a decay index of 50 (as used by Hartford 1972).

DATA UNDER CONSIDERATION

This report considers data from stakes placed on test by FPL. During the past 50 years, FPL has tested the effectiveness of wood preservatives using field tests of 19,000 treated and untreated stakes in Florida, Louisiana, Maine, Mississippi, and Wisconsin. The vast majority of stakes are tested in groups of ten replicates at a given preservative and retention level. The progress of these field trials has been periodically reported in a series of research notes, FPL-02 (most recent report, Gjovik and Gutzmer 1989). Each group of replicate stakes is described in terms of the percentage of stakes that are good, serviceable but showing signs of decay and/or termites, or removed from test because of decay and/or termites. After all stakes have failed, the mean lifetime of a group of replicate stakes is reported.

Too many stakes exist to consider all of them at once. Therefore, in this report we initially confine our attention to groups of ten replicate 2- by 4- by 18-in. Southern Pine stakes placed on test in Mississippi. We will first analyze the variability of 510 control stakes. Then, we will consider methods for estimating the sample median lifetime before all stakes fail. Finally, we will consider confidence intervals for the population median.

Not all available data were used for modeling purposes. Some data were used to create or choose a model, and other data were used to validate the model. The data used for modeling purposes were from 79 groups of Southern Pine 2- by 4- by 18-in. stakes on test in Mississippi, which were treated with chromated zinc chloride, fluor chrome arsenate phenol (type A), copper naphthenate, coal-tar creosote, and pentachlorophenol, from plots 2, 5, 6, 20, 24, 38, 41, 48, 55, 59, and 67. These treatments were chosen because they represent a variety of inorganic and organic preservatives in water- or oil-based formulations. Historical data were also chosen so that adequate failure information was available for the statistical analysis. The data sets used to validate the models were from other preservatives and types of wood products on test in Mississippi and Wisconsin.

Variability of data

The failure times of any group of replicate stakes vary. The difference between the first and last failure times of a particular group of ten stakes ranges from years to decades. Stake plots are inspected at discrete intervals, usually ranging from every 6 months after the plot is first installed to every 2 years after that plot has been on test for a decade or more. Failure time is currently defined as the length of time between the installation date and the inspection date when the stake is observed to have failed. In reality, the stake failed some time between the last inspection date before failure and the inspection date of failure. Because of the discrete nature of the inspection times, there are often multiple failures at a given inspection date. To better reflect the fact that the deterioration process is continuous, we adjusted the recorded failure times so that only one failure occurs at any specific time. For example, if four failures were discovered at the year 2 inspection, and the prior inspection was at year 1, then the failures are assumed to have occurred at 1.25, 1.5, 1.75, and 2 years. This conservative approach did not account for seasonal influence on rate of deterioration and had the most impact on data for untreated control stakes that tended to deteriorate over a relatively short time span. Compared to failure times for untreated stakes, failure times for treated stakes were spread out over a greater number of years and were less clustered; hence, the data for treated stakes were less influenced by this procedure. Unique failure times were necessary for some statistical procedures used in this paper. The slight adjustment of failure times did not significantly affect any results, given the large variability in failure times.

The variability of individual lifetimes is due to the inherent variability in wood stakes, in the actual preservative retention, and in the presence of decay fungi and termites over time in a particular location of the test plot. However, there are no quantitative measurements of any of these factors other than individual stake preservative retention values.

Variability of controls

To explore the variability of stake data, let us first consider the controls. When each plot is established, a set of untreated stakes is included to serve as controls, or a reference baseline. If the underlying nature of the wood or the experimental environment has not changed over time, similar failure patterns should occur for control stakes of similar species. Therefore, the variability observed within or between control stakes in the plots is likely to serve as a lower bound for the variability expected in field tests of wood preservatives.

We will consider 51 groups (1 per plot) of Southern Pine sapwood 2- by 4- by 18-in. control stakes installed between December 1938 and May 1982 at the Harrison Experimental Forest in Saucier, Mississippi. The variability of the stake lifetimes within a given plot can be presented in different ways. In Table 1, failure time data are represented as a histogram and a box plot. In the modified histogram or grouped frequency distribution, time is broken into a ¼-year grid with the number of failures listed in each cell. In the box plot (Velleman and Hoaglin 1981), a box surrounds the center 50% of the data, the median is represented by a perpendicular bar, and "whiskers" extend to the smallest and largest observations, except for outliers, which are denoted by an asterisk.

Actual failure times range from a few months to 6 years. Neither representation indicates sample means. However, the box plots show the median lifetime as well as some indication of the variability in each group. The sample mean and median lifetimes are similar within a plot and range from 1.4 to 3.6 years. Because the

TABLE 1. *Presentation of failure times of Southern Pine 2- by 4- by 18-in. controls.*

Installation Date	Plot	Modified histogram[a] Failure time (yr) 0 1 2 3 4 5 6	Box plot[b] Failure time (yr) 0 1 2 3 4 5 6
12/38	2	1 1 11121 1 1	
1/40	3	1 1 2222	
6/40	4	1212 1 1 1 1	
2/41	5	1111112 1 1	
2/42	7	231 111 1	
2/42	9	12321 1	
2/42	10	1 21211 1 1	
12/43	12	1 212111 1	
12/43	13	21211111	
12/44	14	11 111 11111	
12/45	15	11 2111 1 1 1	
4/48	20	1111111 1 1 1	
12/48	24	32221	
12/48	25	32221	
12/49	26	2121111 1	
12/49	27	1 1111111 1 1	
12/49	28	1 1 21211 1	*
3/52	32	111111111 1	
3/52	33	1 2111111 1	

TABLE 1. *Presentation of failure times of Southern Pine 2- by 4- by 18-in. controls.*—con.

Installation Date	Plot	Modified histogram[a] Failure time (yr) 0 1 2 3 4 5 6	Box plot[b] Failure time (yr) 0 1 2 3 4 5 6
3/52	34	211111111	
3/52	35	111 111 111 1	
3/52	36	11 1111111 1	
12/52	37	1221111 1	
4/53	38	111111111 1	*
12/54	40	1 22221	
12/54	41	22211 1 1	
12/54	42	21211 1 1 1	
12/56	47	2111111 1 1	
12/57	48	111121111	
12/58	53	11111111 1 1	
12/59	54	1 11111121	
12/59	55	1 21111111	
12/60	56	121211 1 1	
12/60	57	11121211	
7/61	59	1111121 1 1	
12/63	61	121111111	
12/63	62	211111111	
12/63	63	123111 1	
1/67	66	1 1 1 12111 1	*

TABLE 1. *Presentation of failure times of Southern Pine 2- by 4- by 18-in. controls.*—con.

Installation Date	Plot	Modified histogram[a] Failure time (yr) 0 1 2 3 4 5 6	Box plot[b] Failure time (yr) 0 1 2 3 4 5 6
11/67	67	2232 1	
12/71	68	11211121	
5/76	74	21121111	
12/76	75	1 11 1 2121	
11/78	78	111112111	
4/80	80	11211111 1	
6/80	81	1111111 1 1 1	
6/80	82	111121111	
5/81	83	1 1 11211 1 1	
12/81	85	22221 1	
5/82	86	1 1212111	
5/82	87	1331 1 1	

[a]Time broken into a 1/4-year grid with number of failures.

[b]A box runs from the first to the third quartile, the median is represented by a perpendicular bar within the box, and "whiskers" extend to the smallest and largest observations, except for outliers denoted by *. Outliers are those data points that are far away from the center of the data, compared to the interquartile range (Velleman and Hoaglin 1981).

median lifetimes do not show a pattern over time, no evidence exists that the lifetime of the control stakes has increased or decreased over time. It is important to note the variability of failure times within each plot, which is not summarized in an average lifetime.

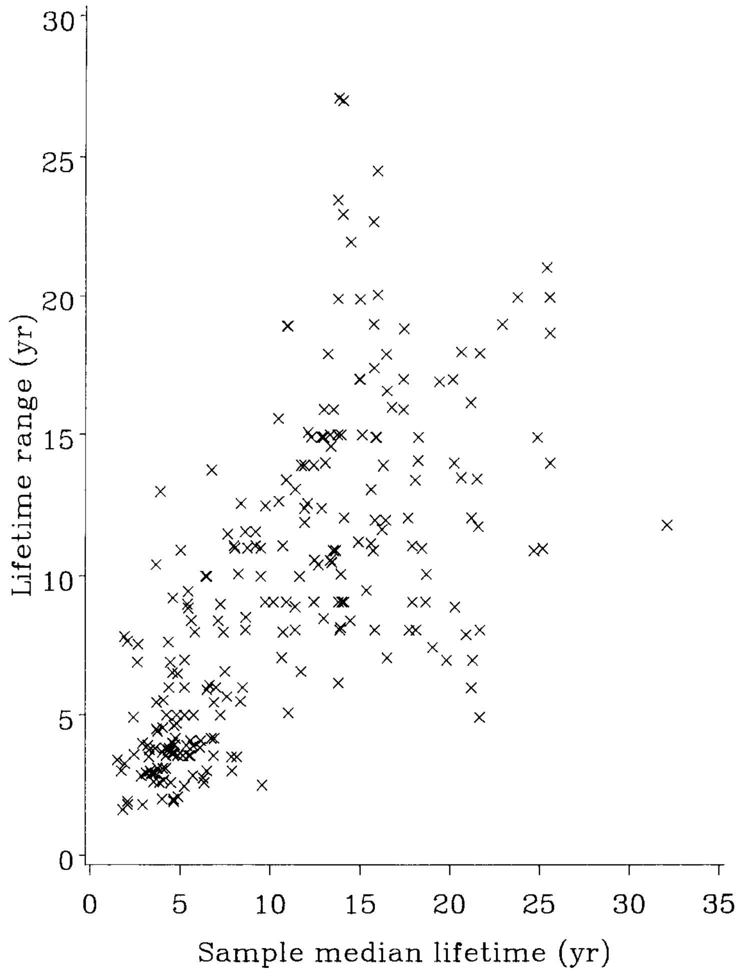


FIG. 1. Median lifetime compared to lifetime range for a group of ten failed replicate stakes treated with different retentions of chromated zinc chloride, fluor chrome arsenate phenol, copper naphthenate, coal-tar creosote, or pentachlorophenol.

Variability of treated stakes

Like the controls, the individual failure times vary widely within a group of ten replicate stakes. For groups of ten replicate stakes treated with either chromated zinc chloride, fluor chrome arsenate phenol (type A), copper naphthenate, coal-tar creosote, or pentachlorophenol in which all stakes have failed, variability in failure time is demonstrated by plotting the sample range (maximum–minimum) against the sample median lifetime (Fig. 1). We detected no difference in distribution patterns of individual stake failure times about the median values for each treatment group. An average lifetime, whether sample mean or median, by itself conceals the large underlying variability of the data.

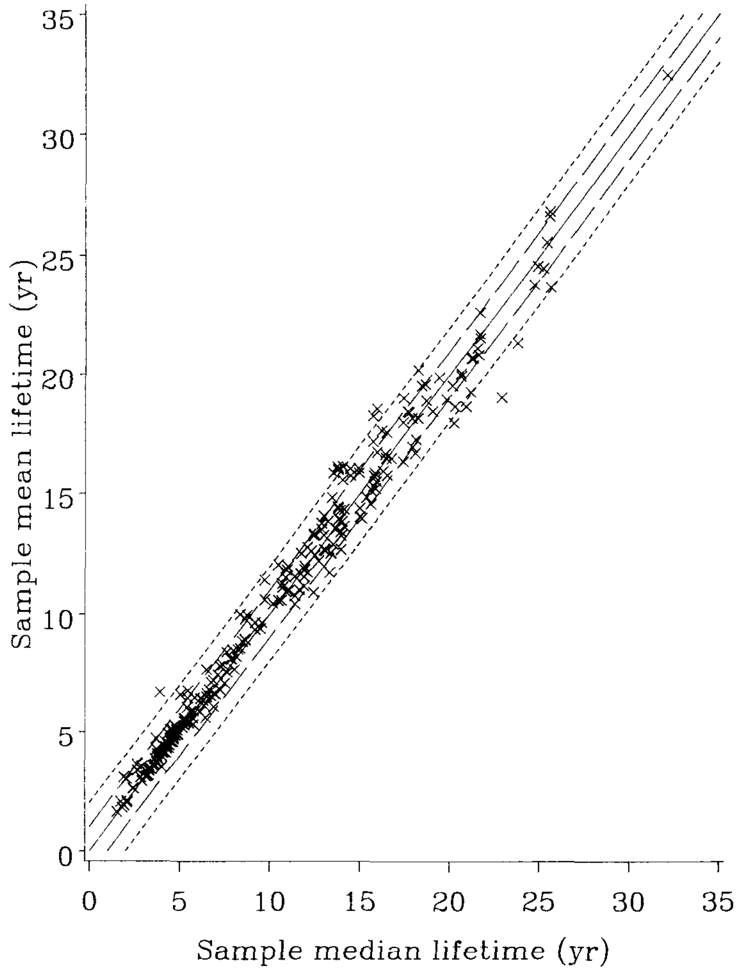


FIG. 2. Mean compared to median stake lifetime. Solid line indicates identical mean and median values; long-dashed lines include points where mean and median differ by at most 1 year; short-dashed lines include points where mean and median differ by at most 2 years.

PREDICTING PRESERVATIVE EFFECTIVENESS

This section deals with predicting the sample average lifetime. Currently, the effectiveness of a wood preservative at a given retention level as reported in the FPL research notes is measured by the sample mean lifetime of a group of stakes treated with a certain preservative to a specified retention level. A lower percentile estimate might be a more appropriate descriptor if one is concerned with the lower portion of the lifetime distribution. However, the sample sizes currently used in preservative field trials allow meaningful comparisons of only an average value. Moreover, a confidence interval for the population average lifetime would indicate the variability associated with the estimate. Given the sample sizes currently under test, recommended confidence intervals (or lower bounds of confidence intervals) for the population average lifetime are based on nonparametric

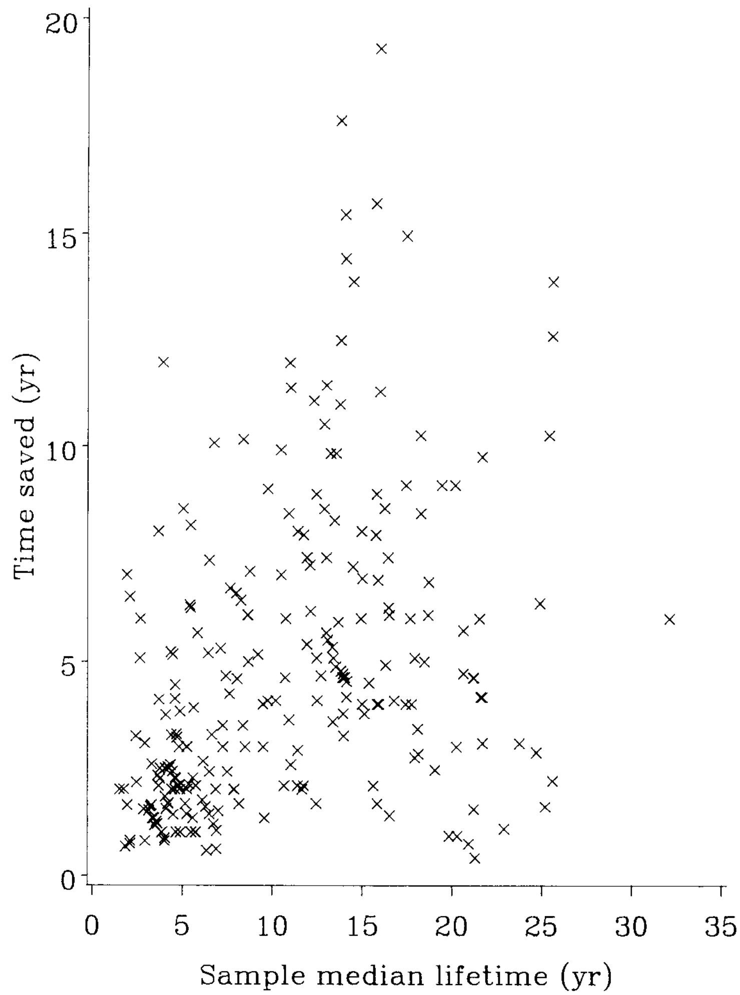


FIG. 3. Time saved by using median instead of mean lifetime to determine average lifetime of groups of ten replicate stakes. Time saved represents difference between inspection date when all stakes had failed and date when sixth stake had failed.

statistics. A good reference for nonparametric statistics is Conover (1980). The nonparametric statistics are presented in the following sections.

Median and mean values

Mean and median values are both estimates of the sample average lifetime. However, the mean can be computed only when all stakes have failed. The median can be calculated when the sixth stake fails (out of ten replicate stakes). Because the sample mean and median are both measures of average lifetimes, it is not surprising that the values are similar in any one group of stakes. In Fig. 2, the difference between the sample mean and median values is less than 1 year, 80% of the time. Similarly, the difference between the sample mean and median value is less than 1.5 (2) years, 90 (95)% of the time.

The use of the median instead of the mean allows earlier determination of the sample measure of average lifetime. The time that is saved is the difference between the inspection date when the last stake has failed and the inspection date of the sixth (out of 10) failure. This ranges from 0, when the last five stakes failed at the final inspection date, to 20 years, with a median of 4 years saved in the time until the sample average can be reported. As the sample median increases, the potential for larger time savings exists. However, a larger sample median does not always imply a larger time savings (Fig. 3). In some preservative-retention combinations that have median lifetimes in excess of 20 years, the difference between the sixth and tenth failures is less than 2 years.

First quartile

Thorne (1918) and MacLean (1957) report that for treated and untreated railroad ties in service, the time when 22% of the ties have failed is 75% of the average lifetime. Does a relationship of this sort exist for preservative-treated stakes? For ten stakes, there is very little difference between the 22d and 25th percentiles. Therefore, we will consider the use of the sample 25th percentile (first quartile) to estimate the 50th percentile (median) of the sample.

Figure 4a contains the data sets from which one can estimate the sample median lifetime; that is, the sixth stake out of ten has failed. The solid line is the relationship proposed by Thorne and MacLean. Some data may be censored; we do not know the exact lifetime, but only that the lifetime must exceed some value. Censored data arise if stakes are lost because of factors other than decay or termite damage, such as fire or mechanical damage, or if the stakes are still on test and not yet failed. If censoring is present, the Kaplan-Meier (1958) product-limit estimator is used to obtain the sample median and first quartile. [Additional discussion of the Kaplan-Meier estimate can be found in Lee (1980).] Note that in the majority of cases, the median lifetime is within 2 years of the first quartile, with a scatter of points in which the median value is considerably larger than the first quartile, up to 10 to 15 years. When predicting the effectiveness of a preservative, a conservative approach would be to underestimate the sample median lifetime.

The subset of the Mississippi Southern Pine 2 by 4 data used for modeling is presented in Figure 4b. The regression equation of the sample median as a function of the first quartile is expressed by the following formula: median = 1.5 + 1.16 (first quartile) (Fig. 4, long-dashed line). This equation gives a 16% increase over the first quartile plus 1.5 years. The increased variability of the sample median lifetime as the sample first quartile increases will show up in any confidence intervals. To be conservative and to avoid obtaining an estimated sample median that is larger than the actual sample median, one would have to use the sample first quartile as the estimate of the sample median (Fig. 4, short-dashed line). The plots in Fig. 4 provide a visual impression of these three models.

The difference between actual and estimated sample medians as a function of the first quartile (Fig. 5) shows that this difference increases as the first quartile increases. This increased variability shows up in confidence intervals. For example, approximate 80% confidence intervals for the predicted sample median from the first quartiles are calculated as $1.5 + 1.16(\text{first quartile}) \pm [1 + 0.13(\text{first quartile})]$. Similarly, 90 and 95% confidence intervals are calculated by the same formula, substituting 1.5 and 2 years, respectively, for the constant 1. In Fig. 5,

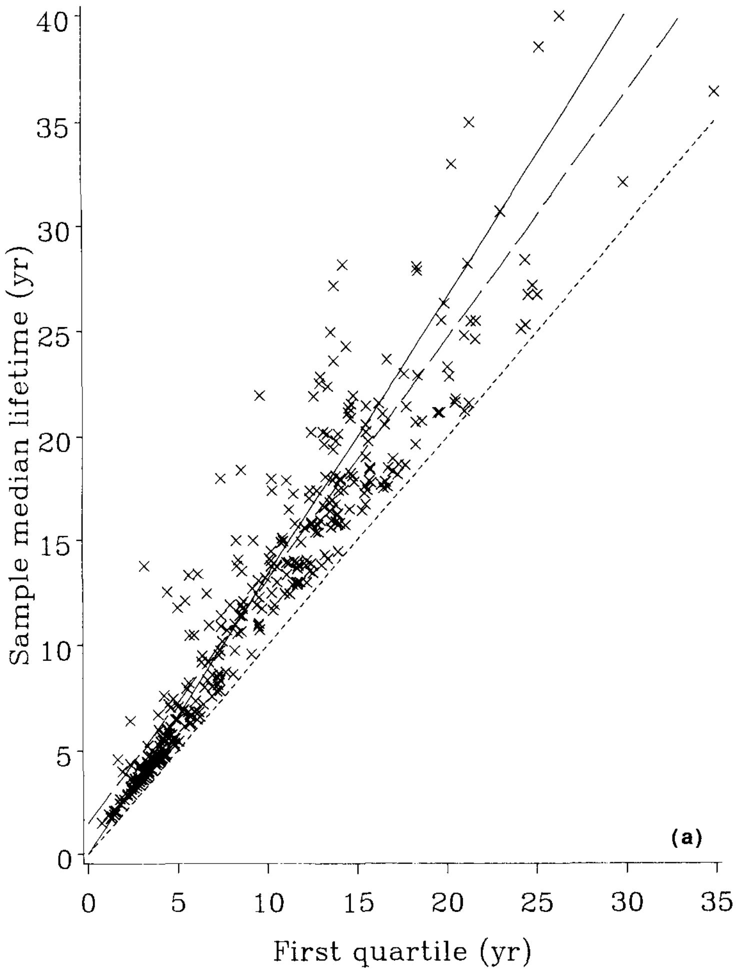
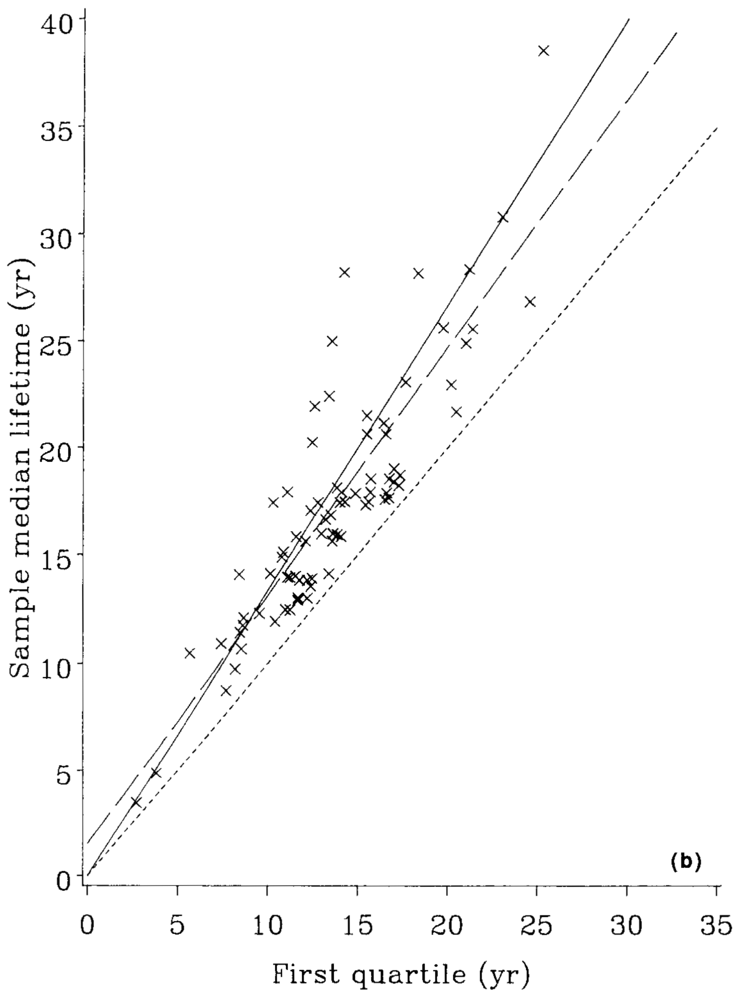


FIG. 4. Median lifetime compared to first quartile (25th percentile of stake sample). (a) All Southern Pine 2 by 4 stakes where the sample median exists. Solid line, relationship between median lifetime and 22d percentile (Thorne and MacLean); long-dashed line, linear regression of sample median on

80 and 95% confidence intervals are shown by long- and short-dashed lines, respectively. Because of the skewed data, the lower bounds on these confidence intervals are conservative. The use of the first quartile to estimate sample median lifetimes saves an additional 2 years, but the penalty is larger variability.

Because one goal of this section is to predict the sample median lifetime, we would like to compare different estimators. So far, we have considered three estimators of the sample median that use the first quartile: (1) Thorne and MacLean's equation, (2) the regression equation, and (3) the first quartile. These estimators are compared in Fig. 4. Alternatively, the estimators can be compared with box plots that show the difference between actual and estimated sample medians (Fig. 6). These estimators are defined in Table 2. The difference between the sample mean and median for complete data groups is given as estimator 0. Positive differences between median and estimated values indicate that the actual median value is underestimated.



results of first quartile; short-dashed line, median and first quartile values identical. (b) Subset used for modeling.

It is probably preferable to underestimate rather than overestimate the median. The Thorne and MacLean estimator overestimates the sample median much more frequently than the regression model. The variability of the difference between sample median and estimated value is larger for the Thorne and MacLean estimator than for the regression estimator. Therefore, the regression estimator is preferable. To avoid overestimating, the first quartile would be used to estimate the sample median.

To validate the regression model, we will consider three other data sets (all groups of ten replicate stakes): (1) Southern Pine 2 by 4 stakes tested in Mississippi that were not used for modeling purposes, (2) stakes of other sizes and species tested in Mississippi, and (3) stakes tested in Wisconsin. Again, we will look at the differences between estimated and actual sample median lifetimes shown in Fig. 6. These differences follow the same pattern as discussed earlier, and they are not systematically larger, which indicates the validity of the model.

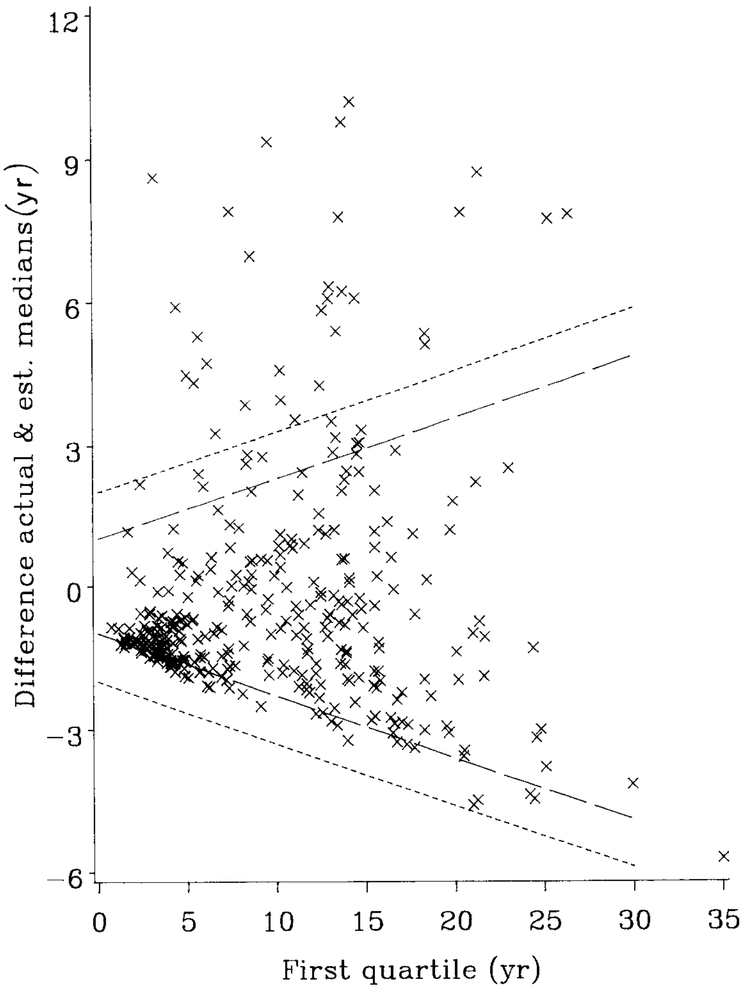


FIG. 5. Difference between actual and estimated sample median lifetimes as a function of first quartile. Vertical axis shows difference between actual and estimated sample median lifetimes in years. Short-dashed lines, 95% confidence interval; long-dashed lines, 80% confidence interval.

Censored parametric distribution functions

If the time to failure data are assumed to come from some parametric distribution, then better average estimates can often be obtained using maximum likelihood estimators of the distributional parameters. However, one must be comfortable with the distributional assumption. For the sample size of ten stakes,

---

→

FIG. 6. Estimators for median lifetime compared by box plots (Velleman and Hoaglin 1981). A box runs from the first to the third quartile, the median is represented by a perpendicular bar within the box, and "whiskers" extend to the smallest and largest observations, except for outliers, which are denoted by *. Outliers are those data points that are far away from the center of the data, compared to the interquartile range (Velleman and Hoaglin 1981). Estimators defined in Table 2.

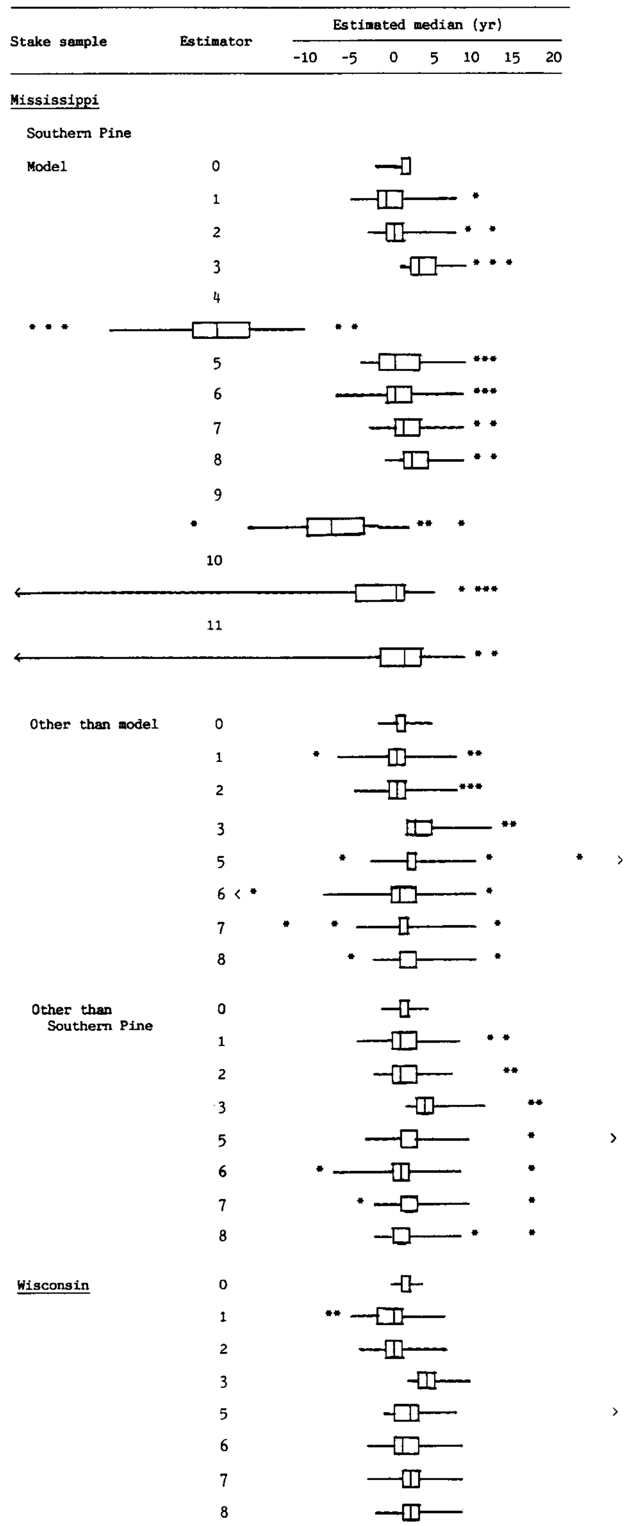

Stake sample
Estimator
Estimated median (yr)
-10 -5 0 5 10 15 20
Mississippi
Southern Pine
Model
0
1
2
3
4
5
6
7
8
9
10
11
Other than model
0
1
2
3
5
6
7
8
Other than Southern Pine
0
1
2
3
5
6
7
8
Wisconsin
0
1
2
3
5
6
7
8

TABLE 2. *Estimators for predicting sample median lifetime.*

Estimator	Definition
0	Estimated median = mean
1	Estimated median = 1.33(first quartile)[a]
2	Estimated median = 1.5 + 1.16(first quartile)[b]
3	Estimated median = first quartile[c]
4	One-parameter exponential
5	Normal
6	Two-parameter lognormal
7	Two-parameter Weibull
8	Extreme value
9	Two-parameter exponential[d]
10	Three-parameter lognormal[d]
11	Three-parameter Weibull[d]

[a] Thorne (1918) and MacLean (1957) relationship.
[b] Regression equation.
[c] Conservative.
[d] Includes location parameter.

it is impossible to determine if the data come from any given distribution. Therefore, we looked at many distributions: one- and two-parameter exponential distributions, normal distribution, two- and three-parameter lognormal distributions, extreme value distribution, and two- and three-parameter Weibull distributions. Nelson (1982) is a good reference for parametric estimation of censored data.

We compared the effectiveness of parametric estimation to the use of the first quartile to estimate the sample median using the same data sets as before. We censored the data so that each group contained only three failures (the minimum number of failures to estimate three distributional parameters). The other stakes in the group were coded as having been censored at the time of the third failure. This method provided equivalent information as when the first quartile was used as an estimator because the first quartile is a function of the second and third failure times. Using the parameter estimates found using maximum likelihood estimation, we then calculated the estimated sample median. The difference between estimated and actual sample median values is shown in box plots in Fig. 6. Parametric distributional estimators are numbers 4 through 11.

Estimators 4 (one-parameter exponential), 9 (two-parameter exponential), 10 (three-parameter lognormal), and 11 (three-parameter Weibull) cannot compete with any of the first quartile estimators for the group of Southern Pine stakes used for modeling. Therefore, these estimators are not included in the other parts of Fig. 6. Apparently, a location parameter (estimators 9, 10, and 11) cannot be estimated nor can the exponential distribution be used as the underlying distribution. One of these parametric estimators, No. 9, merits further discussion. Gillespie et al. (1969) suggest that a two-parameter exponential distribution function fits the data from stake tests. The second parameter of the two-parameter exponential distribution is a location parameter—a value that indicates that no stake fails before this time. The maximum likelihood estimator of the location parameter is the first failure time. However, the first failure is often a poor indicator of the sample median (Fig. 7). Although the two-parameter exponential may fit complete sets of data, it is not useful as a predictor with censored data.

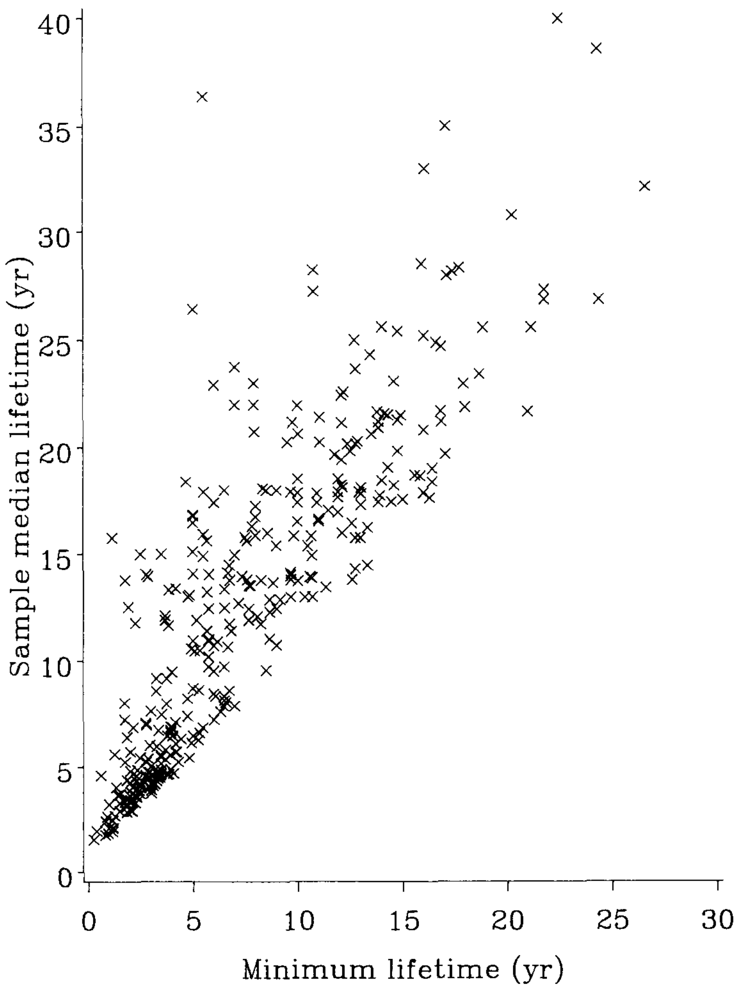


FIG. 7. Comparison of median and minimum lifetimes.

Distribution estimators 5 through 8 do equally well as the estimators that use the first quartile. However, more outliers appear, primarily because occasionally parameters cannot be properly estimated from three failures. Not surprisingly, there is no clear choice for any of these estimators because the small sample size prohibits the choice of a particular failure distribution. Therefore, parametric estimation of the sample median may be appropriate if one has strong reasons to use the selected distribution and the resulting estimates appear reasonable. Without strong reasons to use a distributional assumption, nonparametric estimates are preferable.

CONFIDENCE INTERVALS FOR POPULATION MEDIAN

The previous three subsections dealt with estimating the sample median lifetime. However, statistical comparison of two preservatives or retentions requires confidence intervals for the population as opposed to the sample average. A sample size of ten allows two choices for a confidence interval for the population average.

If the group of data is complete and the data are assumed to come from a normal distribution, the confidence interval for the mean is mean $\pm\ ts/n^{1/2}$, where t is the appropriate value from a table of Student's t-distribution, s is the sample standard deviation, and n is the sample size (in this case ten). The other alternative is to give nonparametric confidence intervals for the population median. Given a sample size of ten, a 99.8% confidence interval would span from the 1st to the 10th failure times; similarly, 98 and 89% confidence intervals would span from the 2d to 9th and 3d to 8th failure times, respectively. Whereas confidence intervals for the population mean (using the normal distribution assumption) are symmetric about the sample mean, nonparametric confidence intervals are not symmetric about the sample median.

A comparison of the widths of 89% parametric (normal) and nonparametric confidence intervals for the Mississippi Southern Pine stakes where both types of intervals can be calculated shows that the nonparametric confidence intervals are not systematically larger than the corresponding parametric intervals (Fig. 8). The solid line shows where the parametric and nonparametric confidence intervals have the same width. Therefore, the nonparametric confidence intervals are preferable because they do not require a distributional assumption. Even if the data are censored, the third failure time gives a lower bound for the 89% confidence interval for the population median. Similarly, the second failure time gives a lower bound for the 98% confidence interval.

CONCLUDING REMARKS

We have emphasized that individual lifetimes of stakes in field tests of preservatives are quite variable. This variability needs to be incorporated into published results of field tests. We suggest that reports include box plots of the actual failure times, using the sample median rather than the mean, and reference the first quartile to provide some idea of the underlying variability of the data.

Estimating the effectiveness of a preservative by using the sample mean lifetime is not totally appropriate. One needs to consider what type of value best measures effectiveness. Is a lower percentile or average value more appropriate? The sample sizes on test today allow comparisons of average values only.

For sample sizes of ten replicate stakes, we have shown that sample mean and median values report the same average value. However, the sample median can be calculated after the sixth of ten replicate stakes fails, whereas the sample mean cannot be calculated until the last stake fails. The use of the median rather than the mean allows one to report an average value approximately 4 years sooner. The difference between sample mean and median values is generally less than 2 years.

The use of the Kaplan-Meier product-limit estimator for the sample empirical distribution function permits us to use data from stakes that have disappeared because of factors other than decay and/or termites. Currently, if the lifetimes of these stakes are censored, the sample size of stakes under test is decreased. The information from censored stakes should not be disregarded.

The sample first quartile can be used to estimate the sample median about 2 years earlier, but the penalty is a larger difference between the actual and estimated sample median values.

We have determined that parametric estimation using the exponential distri-

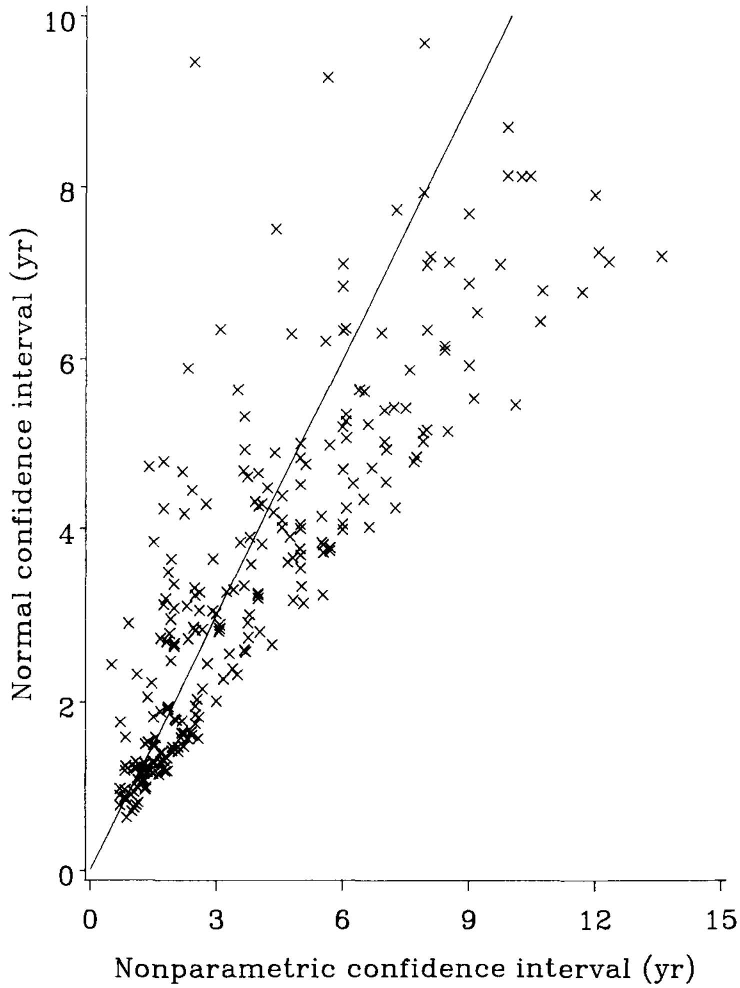


FIG. 8. Widths of normal compared to nonparametric 89% confidence intervals for Mississippi Southern Pine stakes. Solid line indicates where normal and nonparametric intervals have the same width.

bution or a distribution containing an estimated location parameter is useless to estimate the sample median lifetime when considering three failures in a group of ten stakes. However, normal, two-parameter lognormal, two-parameter Weibull, and extreme value distributions may be useful to estimate the sample median if a particular distributional assumption is appropriate and unreasonable values are discarded.

Statistical comparisons of different preservatives or retentions require nonparametric confidence intervals for the population median, not the sample median. The size of these confidence intervals (where we can estimate them) indicates that a difference in sample median lifetimes of less than 4 years is unlikely to be statistically significant. The lower boundary of an 89% confidence interval for the population median is the third failure time (second failure time for 98% confidence interval). Changes in field test procedures will be needed for estimating a lower percentile of the lifetime distribution.

Our recommendation that variability be incorporated in reports of field tests applies to all field trials with preservatives, independent of stake size, wood species, plot location, soil type, climate, and other biological and environmental variables. Improved evaluation of performance data will allow more definitive determination of the potential durability of wood product/preservative combinations in various ecosystems.

ACKNOWLEDGMENT

This research was supported in part by grant 87-FSTY-9-0254.

REFERENCES

COLLEY, R. H. 1970. The practical meaning of preservative evaluation tests. Proceedings of the American Wood-Preservers' Association, pp. 16–37.

CONOVER, W. J. 1980. Practical nonparametric statistics, 2nd ed. John Wiley and Sons, New York.

GILLESPIE, T., H. P. SEDZIAK, AND J. KRZYZEWSKI. 1969. The analysis of service life experiments on wood treated with preservatives. Wood Sci. 2(2):73–76.

GJOVIK, L. R., AND D. I. GUTZMER. 1989. Comparison of wood preservatives in stake tests. 1987 Progress report, USDA Forest Service, Forest Products Laboratory Research Note FPL-RN-02, Madison, WI.

HARTFORD, W. H. 1972. The interpretation of test plot data. Proceedings of the American Wood-Preservers' Association, pp. 67–82.

KAPLAN, E. L., AND P. MEIER. 1958. Nonparametric estimation from incomplete observations. J. Statist. Assoc. 53:457–481.

LEE, E. T. 1980. Statistical methods for survival data analysis. Wadsworth Press, Belmont, CA.

MACLEAN, J. D. 1957. Percentage removals and average life of railway ties. USDA Forest Service, Forest Products Laboratory, Report No. 886, Madison, WI.

NELSON, W. 1982. Applied life data analysis. John Wiley and Sons, New York.

THORNE, M. E. 1918. Relation between average life of ties and percentage of renewals. Proceedings of the American Wood-Preservers' Association, pp. 150–156.

VELLEMAN, P. F., AND D. C. HOAGLIN. 1981. Applications, basics, and computing of exploratory data analysis. Duxbury Press, Belmont, CA.