Forest Sci., Vol. 31, No. 1, 1985, pp. 260–268

An Evaluation of Percentile and Maximum Likelihood Estimators of Weibull Parameters

STANLEY J. ZARNOCH
TOMMY R. DELL

ABSTRACT. Two methods of estimating the three-parameter Weibull distribution were evaluated by computer simulation and field data comparison. Maximum likelihood estimators (MLB) with bias correction were calculated with the computer routine FITTER (Bailey 1974); percentile estimators (PCT) were those proposed by Zanakis (1979). The MLB estimators had superior smaller bias and mean square error but larger variance than the PCT estimators. The MLB bias correction in FITTER increased the bias of parameter *c*, suggesting that for the three-parameter Weibull, the MLB estimators should be used without the correction. Comparisons of predicted percentages indicate that either MLB or PCT estimators, which are simpler to use, can model pine plantations equally well. FOREST SCI. 31:260–268.

ADDITIONAL KEY WORDS. FITTER, diameter distribution, Weibull distribution, growth and yield, modeling, estimation.

THE WEIBULL DISTRIBUTION, introduced by Bailey and Dell (1973) as a model for diameter distributions, has been used extensively in forestry. The probability density function as used in forestry applications is

$$f(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left\{-\left(\frac{x-a}{b}\right)^{c}\right\} \quad \text{for } x \geq a,\ a \geq 0,\ b > 0,\ c > 0$$
$$= 0, \text{ elsewhere}$$

where
x = dbh,
a = location parameter,
b = scale parameter, and
c = shape parameter.

Although it has been shown that the Weibull distribution adequately fits data from many different types of forest stands, the estimation of its three parameters can be difficult. Currently the most prevalent method is that of maximum likelihood, which requires a costly, iterative computer algorithm when all three parameters must be estimated.

As an alternative to maximum likelihood, Zanakis (1979) examined several simple proposed estimators of the Weibull parameters. He proposed a set of percentile estimators which were simple to calculate and even more accurate than the maximum likelihood estimators when $c < 2$ and the sample size is small. In

The authors are Mathematical Statisticians, USDA Forest Service, Southern Forest Experiment Station, Room T-10210, Postal Services Building, 701 Loyola Avenue, New Orleans, LA 70113.
Manuscript received 8 November 1983.

most forestry applications, however, and especially in the estimation of plantation growth and yield where the Weibull has been heavily relied upon, the parameter c is greater than 2 and may range to 6.

The objective of this research was to assess the usefulness of the percentile estimators by comparing them to those of the maximum likelihood estimators over the range of parameters usually encountered in modeling plantation growth and yield. Computer simulation techniques were used to generate known Weibull populations and the estimators were compared using appropriate statistical properties. In addition, the estimators were analyzed using loblolly pine plantation data.

METHODS

Estimators.—The percentile estimator (PCT) for the location parameter a as defined by Zanakis (1979) is

$$\hat{a} = \frac{x_1 x_n - x_2^2}{x_1 + x_n - 2x_2} \text{ if } x_2 \text{ is closer to } x_1 \text{ than to } x_n, \text{ and } x_1 \text{ otherwise}$$

where x_i = the i^{th} ordered value (from smallest to largest) in the sample and n = sample size. If $\hat{a}$ is less than zero, which may occur occasionally, it is set equal to x_1. The scale parameter is estimated as

$$\hat{b} = -\hat{a} + x_{[0.63n]}$$

where [ ] indicates rounding up. The shape parameter is estimated as

$$\hat{c} = \frac{\ln\left[\dfrac{\ln(1 - p_k)}{\ln(1 - p_i)}\right]}{\ln\left[\dfrac{x_{[np_k]} - \hat{a}}{x_{[np_i]} - \hat{a}}\right]}$$

where $p_i = 0.16731$ and $p_k = 0.97366$.

The solution of the three-parameter maximum likelihood equations could be achieved by any of a number of nonlinear algorithms. Zanakis (1977) evaluated the performance of several nonlinear optimization algorithms and discussed their merits. Zutter and others (1982) developed computer algorithms based on the method of constrained modified quasilinearization (Wingo 1973) for computing maximum likelihood estimates of various censored samples and truncated Weibull populations.

In this study the maximum likelihood estimators were calculated with the computer routine FITTER (Bailey 1974) with the bias correction for parameter c. This was an extension of the correction given by Thoman and others (1969) for the two-parameter Weibull distribution where the parameter a is known. In this paper, unless specified otherwise, the maximum likelihood estimator (MLB) will imply use of the bias correction for the three-parameter case. (Estimating criteria needed for the execution of the program consisted of parameters controlling the iteration procedure. These were set at EBC = 0.001, EA = 0.1, BOUND = YY(1), and LM = 20.)

Simulation Study.—For the comparison of methods, computer simulation techniques were used to generate Weibull samples of 25-tree and 50-tree plots separately from 26 populations defined by all possible combinations of

a = 0.00, 1.00, 2.00,
b = 2.00, 5.00, 8.00,
c = 2.00, 4.00, 6.00,

except the case of $a = 2.00$, $b = 2.00$, and $c = 6.00$ which gave extremely variable results and hindered drawing meaningful conclusions from the data. These values represent the plot sizes and range of parameter values commonly associated with southern pine plantation research plots. A Weibull variate from a specified population was easily generated as

$$x = a + b[-\ln(1 - R)]^{1/c}$$

where R is a random variate on the interval 0 to 1. For each of the 26 populations and two plot sizes, 100 replications of the simulation experiment were made, requiring a total of 195,000 Weibull variates.

Field Data Comparison.—It is of interest to determine if significant differences exist between parameter estimates when field data are fitted to the Weibull. Thus, to further evaluate the two sets of estimators, dbh field data from 20 loblolly pine plantation research plots were fitted to the Weibull with the PCT and MLB estimators. These data were supplied by the Southern Forest Experiment Station, Research Work Unit SO-1102, Study 3.22.

Statistical Criteria.—For quantitative comparison of different estimators of a parameter, several sample statistics may be used. The simplest is the sample bias, defined as

$$\hat{B}(\hat{\alpha}) = \sum_{i=1}^{n} \hat{\alpha}_i/n - \alpha$$

where
α = the value of the parameter to be estimated,
$\hat{\alpha}$ = the estimator under consideration for the parameter, and
$\hat{\alpha}_i$ = the estimate for simulation i ($i = 1, 2, \ldots, n$).

The sample bias is an estimate of how far, on the average, the estimator will vary from the true value of the parameter. A negative bias would imply underestimation, a positive bias overestimation.

An equally important statistic for comparison is the sample variance,

$$\widehat{\text{Var}}(\hat{\alpha}) = \sum_{i=1}^{n} \left(\hat{\alpha}_i - \sum_{i=1}^{n} \hat{\alpha}_i/n\right)^2/(n - 1).$$

This is a measure of how far the estimator is expected to vary from its mean value, not from the true value of the parameter. It is analogous to precision and does not incorporate the bias of the estimator.

A third sample statistic which expresses both the bias and variance is the mean square error (M.S.E.), a measure of the accuracy of the estimator. It is concerned with the difference between the estimate and the true value of the parameter, and is defined as

$$\widehat{\text{M.S.E.}}\ (\hat{\alpha}) = \sum_{i=1}^{n} (\hat{\alpha}_i - \alpha)^2/(n - 1) = \widehat{\text{Var}}(\hat{\alpha}) + [\hat{B}(\hat{\alpha})]^2.$$

Results

Simulation Study.—In the simulation study, the values obtained for the parameters a, b, and c using the percentile and maximum likelihood estimators with

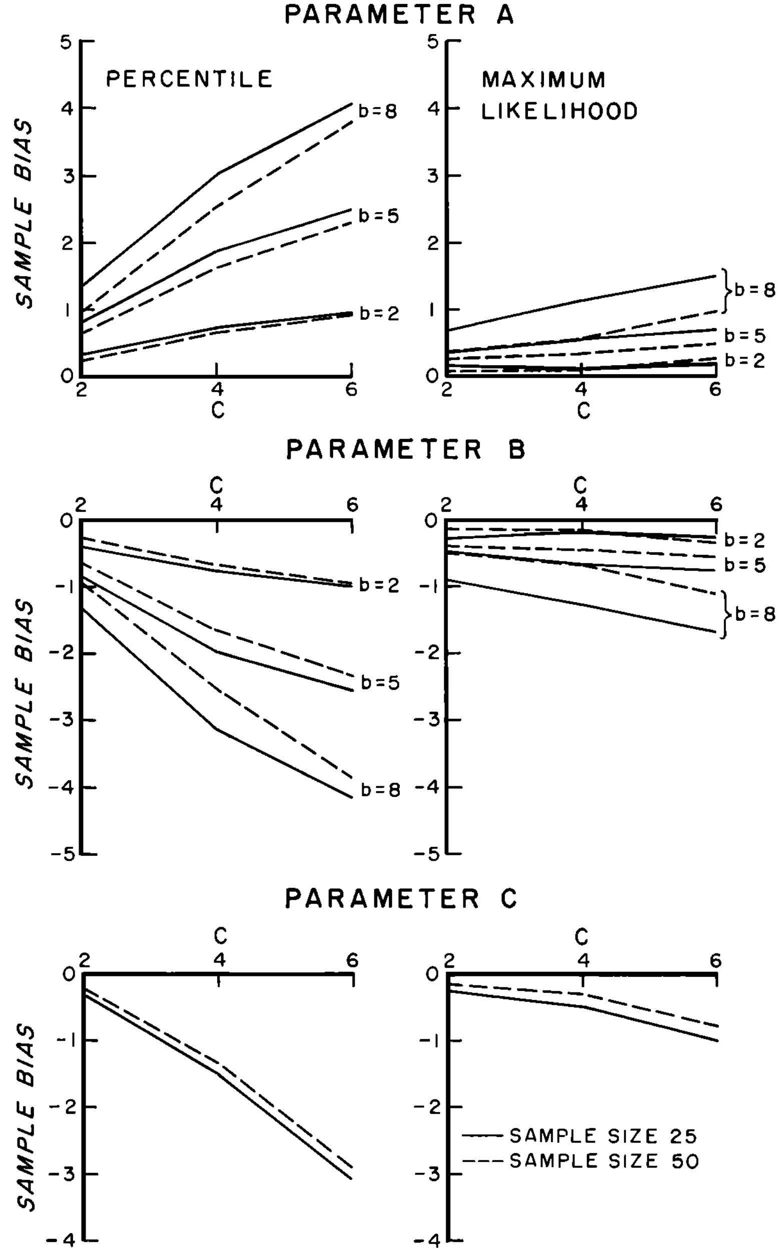

FIGURE 1. Sample bias of percentile and maximum likelihood estimators for parameters *a*, *b*, and *c*. Note that for parameters *a* and *b*, sample biases were combined over the parameter *a* classification and hence each point is the average of 300 replications except in the case of $b = 2$ and $c = 6$, where there were 200 replications; for parameter *c*, sample biases were combined over the parameter *a* and *b* classification, and hence each point is an average of 900 replications, except in the case of $c = 6$, where there were 800 replications.

bias correction were compared graphically. Only pertinent results and trends are given here.

The magnitude and direction of bias varied according to the parameter (Fig. 1). Because the sample bias of parameter *a* for both PCT and MLB estimators

TABLE 1. Comparison of sample variance of estimates calculated by percentile and maximum likelihood methods for 26 Weibull populations (100 replications, plot sizes 25 and 50 trees).

Parameter and plot size	Sample variance			
	Percentile		Maximum likelihood	
	Minimum	Maximum	Minimum	Maximum
a (location)				
25	0.035	1.580	0.058	7.954
50	.018	1.316	.021	5.692
b (scale)				
25	.073	2.106	.116	8.361
50	.044	1.339	.048	6.047
c (shape)				
25	.125	2.011	.176	9.053
50	.074	1.039	.072	5.349

was independent of the population value of *a*, the data were pooled over the three values of *a*. Independence was judged by visual inspection of graphs of the sample bias for the values of the *a* parameter. The results revealed an increasing positive bias with increasing values of the parent population parameters *b* and *c*. The bias of the PCT estimator was approximately four times that of the MLB.

Parameter *b* reflected bias properties similar to those of parameter *a*. The bias was independent of parameter *a* but increased negatively with increasing *b* and *c*. For any given population, the bias in *b* was approximately of the same magnitude as that of *a* but of opposite sign.

Because the sample bias of parameter *c* was apparently independent of *a* and *b*, the data were pooled over the appropriate populations. The bias increased negatively with increasing *c* and was about three times greater for the PCT estimator as compared to the MLB. It is interesting to note that despite the bias correction, the MLB estimators were still biased. This point will be discussed later.

The increase in sample size from 25-tree plots to 50-tree plots reduced the absolute magnitude of the bias for all parameters for nearly all estimates; exceptions were probably due to variability in the experiment.

The sample variance for all three parameters was markedly greater for the MLB estimators in nearly all situations. Generally, the variability of the MLB estimators was four times that of the PCT (Table 1). The variance of parameter *a* increased with increasing *b* and *c* for both estimators. It was independent of the value of *a* for the PCT but increased with increasing *a* for the MLB. The variance of parameter *b* for the PCT was independent of *a* and *c* but increased with increasing *b*; that of the MLB increased with increasing *a, b,* and *c*. The variance of parameter *c* for the PCT estimator revealed no trends with *a* and *b* but increased with increasing *c*. The MLB estimator increased with increasing *a* and *c* but decreased with increasing *b* value.

Generally, the variance decreased with larger plot size. This trend was more easily detectable in the MLB, because the variance was large.

The mean square error incorporates the bias and variance and is thus a measure of the total error. The M.S.E. allows an overall interpretation of the quality of an estimator. Generally, the PCT estimators of all three parameters gave a slightly higher M.S.E. than the MLB estimators at values of *c* close to 2. As *c* increased

TABLE 2. Estimates calculated by percentile and maximum likelihood methods for Weibull parameters for 20 loblolly pine plots.

Estimates					
Percentile			Maximum likelihood		
a	*b*	*c*	*a*	*b*	*c*
0.4	6.37	3.94	0.0	6.74	4.55
3.2	3.95	3.00	.0	7.00	5.65
1.7	5.30	3.08	.0	6.80	4.06
2.6	3.92	3.31	.0	6.57	5.10
1.9	4.59	3.44	.0	6.48	4.62
3.3	3.90	2.48	1.6	5.68	3.36
1.3	6.18	5.07	.5	6.90	5.20
1.6	6.50	4.06	.0	8.11	4.44
3.5	4.30	2.50	.0	7.67	4.93
2.9	5.01	3.07	.0	7.72	4.56
2.5	5.04	3.81	.0	7.39	5.47
4.4	3.41	1.82	.4	7.18	5.74
3.0	4.25	3.67	.0	7.20	5.50
.7	7.47	4.65	.0	7.77	4.75
3.3	4.44	2.26	.0	7.67	4.84
3.1	4.93	3.04	.0	7.74	5.78
2.2	4.51	4.03	.0	6.80	5.47
.8	6.38	3.72	.0	7.06	3.76
3.7	3.83	2.55	2.0	5.68	3.76
1.6	5.20	2.86	.0	6.52	3.60

toward 6, however, the M.S.E. for the PCT estimators increased much more rapidly as compared to the MLB, often being two to three times as great. Similar trends were noted by Zanakis (1977, 1979).

Field Data Comparison.—The results of the field study (Table 2) indicate that the PCT estimates were significantly different from the MLB estimates which, based on the simulation study, were considered the best. The nonparametric procedures associated with Wilcoxon signed ranks for paired replicates were performed (Hollander and Wolfe 1973). The MLB estimate for parameter *a* was found to be significantly different from its PCT counterpart. A 95 percent confidence interval for the difference (MLB-PCT) was (−2.70, −1.65). The MLB estimates for *b* and *c* were also significantly different from the PCT estimates, with confidence intervals for the difference found to be (1.55, 2.62) and (0.90, 1.93), respectively.

Discussion

Bias Correction for c.—Surprisingly, the bias correction for parameter *c*, used with the computer routine FITTER, did not prevent negative bias of the MLB estimator, for the populations studied. For instance, given a population with $a = 1.00$, $b = 8.00$, and $c = 6.00$, the mean MLB estimate based on 100 replications of a 25-tree plot was $c = 4.449$. This indicates a negative bias of approximately 26 percent. A similar bias, though considerably smaller, is undoubtedly present in the estimators of *a* and *b*,

The bias correction for *c* for the two-parameter Weibull is a multiplicative reduction correction based on a monotonically increasing function of sample size, asymptotically converging toward 1 as the sample size increases. This correction implies that the maximum likelihood estimator for *c* for the two-parameter Wei-

TABLE 3. Sample bias for the maximum likelihood estimator of parameter c, *with and without the bias correction.*

Value of *c*	Sample bias*				Number of replications
	25-tree plots		50-tree plots		
	With correction	Without correction	With correction	Without correction	
2	−0.219	−0.114	−0.158	−0.107	400
4	−0.484	−0.274	−0.225	−0.118	100
6	−1.145	−0.856	−0.849	−0.704	300

* Difference between estimated and true values.

bull is positively biased. To investigate the bias in the three-parameter Weibull, simulation experiments at eight selected points in the experimental design were performed using FITTER without the bias correction. Since the bias in c was independent of a and b, the results were combined for various populations and presented with respect to c (Table 3). The results show a somewhat smaller negative bias than that of results with the bias correction. Hence, a more suitable bias correction would be one of a multiplicative increasing type.

Distribution Aspects.—Although the results reported here indicate that the MLB estimators are superior to the PCT for estimating the three parameters, they do not allow comparison of estimation of the overall distribution. It is conceivable that the Weibull distribution may be insensitive to the magnitudes of the bias in the PCT estimators. To investigate this property, selected Weibull populations were generated, the parameters were estimated by both techniques, and various percentiles were calculated and compared (Table 4). The results indicate that although the estimators may be considerably inaccurate, the resulting percentiles are remarkably well estimated. In one example, although both estimating procedures appear poor, the selected percentiles are all within 0.1 of the true population value. Generally, the MLB estimators appear to produce slightly better estimates but the PCT are probably well within the margin of error anticipated by most researchers.

Conclusion

For populations similar to forest stands, the MLB estimators were found to be superior in accuracy. Analysis of simulation data indicated that the MLB estimators showed a smaller M.S.E. In field data tests, the PCT estimates were significantly different from corresponding MLB estimates. Thus, although the PCT estimates may be comparable to MLB for a restricted set of Weibull populations, they are inferior under conditions commonly encountered in pine plantation growth and yield research.

The PCT estimators, however, should not be dismissed as totally unsuitable. Their simplicity is a valuable attribute and their behavior when c is near or below 2 is comparable to or better than that of the MLB. In addition, this research has demonstrated that although the bias exceeds that of the MLB, the PCT estimators' variances are much smaller. This property encourages the development of a bias correction for the PCT estimators.

In calculating maximum likelihood estimates for the three-parameter Weibull with FITTER, the bias correction should not be used. It was formulated for the two-parameter Weibull distribution and is not appropriate for the three-param-

TABLE 4. Relation between estimated parameters and percentiles for selected Weibull populations. Each set of values is based on a single trial of 25 observations.

Item	Parameters			Percentiles				
	a	*b*	*c*	10	25	63	75	90
True	0	2	2	0.6	1.1	2.0	2.4	3.0
PCT	0.1	1.9	1.8	0.6	1.0	2.0	2.4	3.1
MLB	0	2.2	1.6	0.5	1.0	2.2	2.8	3.8
True	0	2	6	1.4	1.6	2.0	2.1	2.3
PCT	1.2	0.7	2.1	1.4	1.6	1.9	2.0	2.2
MLB	0.6	1.3	4.2	1.4	1.6	1.9	2.0	2.2
True	0	8	2	2.6	4.3	8.0	9.4	12.1
PCT	0.5	7.4	2.1	3.1	4.6	7.9	9.2	11.5
MLB	0	8.1	2.2	2.8	4.5	8.0	9.4	11.9
True	0	8	6	5.5	6.5	8.0	8.4	9.2
PCT	4.8	3.0	1.8	5.7	6.3	7.8	8.4	9.6
MLB	4.1	3.8	2.3	5.6	6.4	7.9	8.5	9.6
True	1	5	4	3.8	4.7	6.0	6.4	7.2
PCT	3.4	3.0	1.8	4.2	4.9	6.3	6.9	8.1
MLB	2.9	3.3	2.1	4.1	4.8	6.2	6.8	7.8
True	2	2	2	2.6	3.1	4.0	4.4	5.0
PCT	2.3	1.7	2.4	2.9	3.3	4.0	4.2	4.7
MLB	1.8	2.1	2.8	2.8	3.2	3.9	4.2	4.7
True	2	2	6	3.4	3.6	4.0	4.1	4.3
PCT	2.8	1.2	6.3	3.7	3.8	4.0	4.1	4.2
MLB	0	4.0	14.3	3.4	3.7	4.0	4.1	4.3
True	2	8	2	4.6	6.3	10.0	11.4	14.1
PCT	3.8	7.6	1.3	5.2	6.8	11.4	13.5	18.0
MLB	3.5	6.5	1.4	4.8	6.2	10.0	11.8	15.4
True	2	8	6	7.5	8.5	10.0	10.4	11.2
PCT	7.0	3.0	1.4	7.7	8.3	10.0	10.8	12.3
MLB	6.0	3.8	2.4	7.5	8.3	9.8	10.3	11.4

eter; there the bias correction for c actually corrects in the wrong direction, increasing an already negative bias.

The insensitivity of the Weibull distribution to variations in the parameter estimates has been demonstrated. Because the parameters are correlated, various combinations of parameters can lead to very similar distributions. The ability of the simple PCT estimators to mimic the distribution well makes them useful to those concerned with the distribution and not with interpreting individual parameters.

Literature Cited

Bailey, R. L. 1974. Announcement: Computer programs for quantifying diameter distributions with the Weibull function. Forest Sci 20(3):229.

Bailey, R. L., and T. R. Dell. 1973. Quantifying diameter distributions with the Weibull function. Forest Sci 19(2):97–104.

Hollander, M., and D. A. Wolfe. 1973. Nonparametric statistical methods. Wiley and Sons, New York. 503 p.

Thoman, D. R., L. J. Bain, and C. E. Antle. 1969. Inferences on the parameters of the Weibull distribution. Technometrics 11(3):445–460.

Wingo, D. R. 1973. Solution of the three-parameter Weibull equations by constrained modified quasilinearization (progressively censored samples). IEEE Trans on Reliability R-22(2):96–102.

Zanakis, S. H. 1977. Computational experience with some nonlinear optimization algorithms in deriving maximum likelihood estimates for the three-parameter Weibull distribution. *In* Algorithmic methods in probability (M. F. Neuts, ed), North-Holland/TIMS Studies in Manage Sci 7:63–77.

Zanakis, S. H. 1979. A simulation study of some simple estimators for the three-parameter Weibull distribution. J Stat Comput Simulation 9:101–116.

Zutter, B. R., R. G. Oderwald, R. M. Farrar, Jr., and P. A. Murphy. 1982. WEIBUL a program to estimate parameters of forms of the Weibull distribution using complete, censored, and truncated data. Va Polytech Inst and State Univ, Sch For and Wildl Resour Publ FWS-3-82, 17 p.

Errata

Inoculation of Three Quercus *Species with Eleven Isolates of Ectomycorrhizal Fungi. II. Foliar Nutrient Content and Isolate Effectiveness,* by R. J. Mitchell, G. S. Cox, R. K. Dixon, H. E. Garrett, and I. L. Sanders, Forest Science 30(3):563–572.

On pages 566, 567, and 568, in Tables 2, 3, and 4, the unit of measurement for columns 6–11 (B, Cu, Fe, Mn, Mo, Zn) should be μg.

(Cont'd. from outside back cover)

versity Microfilms or Dissertation Abstracts reference where applicable; credit unpublished theses, communications, etc., by parenthetical references or footnotes. Abbreviations should follow *American Standards for Periodical Title Abbreviations* (American Standards Association).

Nomenclature and Terminology: At first mention of the species in the text, give its common name, if any, immediately followed in parentheses by its italicized scientific name and authority. Some deviation from this rule may be desirable to avoid awkwardness when names are numerous. The authority for nomenclature of North American tree species is *Checklist of United States Trees (Native and Naturalized)*, U S Dep. of Agric. Handbook 541. Authorities of nomenclature of other plants and of animals are listed in the *Style Manual for Biological Journals.* Where applicable, indicate how or by whom taxa were identified, and location of voucher specimens. Technical usage in forestry and allied fields follows *Forestry Terminology,* SAF.

Number of Copies: Submit the ribbon copy (or sharp photocopy), plus three copies with illustrations to facilitate review.

Suggestions to Secretaries: Double space all copy, including tables, captions, and literature citations, on 8 × 10½ or 8½ × 11 inch white bond or, preferably, manuscript paper with numbered lines. Type each table on a separate sheet with title at top. Type captions for individual figures together on a separate sheet placed at the end of the manuscript. Attach individual captions to figures in review copies. Indicate footnotes in text by consecutive, superior numerals, never asterisks; and treat similarly to figure captions. Large tables typed for photographic reproduction should be *single spaced,* neat and compact; they may be as much as 10 inches wide.