J. Amer. Soc. Hort. Sci. 126(4):503–510. 2001.

Estimating Mean Fruit Weight and Mean Fruit Value for Apple Trees: Comparison of Two Sampling Methods with the True Mean

Richard P. Marini

Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-3083

Additional index words. *Malus sylvestris* var. *domestica*, statistics, fruit size

Abstract. **Data obtained over two years from chemical thinning experiments with 'Redchief Delicious' apple [*Malus sylvestris* (L.) Mill. var. *domestica* (Borkh.) Manst.] on Malling 26 (M.26) rootstock were used to estimate mean fruit weight (MFW) and mean fruit value (MFV) using two sampling methods. The estimated values were compared with the true MFW and the true MFV calculated from the entire crop from a tree. Statistical techniques were used to assess agreement between the values obtained with estimation methods and the true values. Estimates of MFW obtained from a 20-fruit sample per tree may differ from the true value by ≈13% and estimates obtained from weighing all fruit on three limbs per tree may range from 11% to 19% of the true mean. Estimates of MFV obtained from packouts of a 20-fruit sample may differ from the true value by about $0.04 (U.S. dollars)/fruit and estimates from packing out all fruit on three limbs per tree may differ from the true mean by about $0.07/fruit. Analysis of variance was performed on each data set. The resulting *P* values differed for the three methods of calculating MFW and MFV. Therefore, erroneous conclusions may result from experiments where MFW and MFV are estimated from subsamples. Error associated with estimating fruit weight and fruit value from the sampling methods employed in this study may be larger than many pomologists can accept. Until protocols for sampling apple trees, which account for the important sources of within-tree variation, are developed, researchers should consider harvesting the entire crop to calculate MFW and MFV.**

Fruit size and fruit value (the economic value of a fruit) are important factors influencing apple (*Malus sylvestris* var. *domestica*) orchard profitability. Pomologists want to know which practices and treatments influence mean fruit weight (MFW) and mean fruit value (MFV). Obtaining the true MFW for a tree is expensive and involves dividing the weight of the crop by the number of fruit. Researchers have reported true MFW (Barritt et al., 1995; Dozier, et al., 1982; Hampson et al., 1997), but due to financial and labor limitations, MFW is often estimated with various subsampling schemes. Sources of variation for FW within a tree, such as limb position, canopy position, and fruit position within a cluster, have not been characterized for apple. Therefore, sampling protocols for estimating true MFW have not been published. Data for MFV are rarely published. Sampling to estimate MFW usually involves harvesting five to 30 fruit per tree (Dozier et al., 1980; Greene, 1986; Miller, 1982), and while such sampling schemes are assumed to provide reasonable estimates of MFW, they have not been evaluated statistically.

Sometimes a quantity cannot be easily measured and techniques are developed to estimate the true value. To objectively judge their acceptability in research or commercial settings, the precision, accuracy, and repeatability of such techniques must be studied. One approach is to compare a new or established technique with an accepted reference method. If the test method compares favorably, it is considered acceptable. Detailed schemes, based on least-squares analysis, *F* test, *t* test, and correlation coefficients, for performance evaluation have been published (Barnett, 1965; Barnett and Youden, 1970; Broughton et al., 1969). Westigard and Hunt (1973) reported that the improper use or faulty interpretation of the statistical parameters might result in invalid judgments on the acceptability of new methods. Biometricians have published new statistical methods for assessing the agreement between two approximate measurement techniques (Bland and Altman, 1986) and for comparing a new technique to the gold standard or a method that provides the true value (Laurent, 1998). The statistical calculations cannot determine if a method is acceptable, but they can provide specific estimates of the types and magnitudes of errors. The absence of errors substantiates acceptable performance. Decisions on acceptability are judgments made by individual researchers as to what amount of error is tolerable.

The purpose of this study was to compare two methods of estimating MFW and MFV for apple trees. MFW and MFV were estimated with each method and were compared to the true values obtained by harvesting and grading all fruit on a tree. Statistical techniques were used to provide estimates of errors, which can be used to make decisions on acceptability of these methods.

Materials and Methods

Treatments. Data were obtained from chemical thinning experiments in 1997 and 1998. These treatments were used to obtain a range in crop load and MFW that are typical of experiments involving comparison of cultural practices. Ten-year-old 'Redchief Delicious'/M.26 trees were used each year. In 1997 a completely randomized design (CRD) was used with a factorial treatment structure involving four levels of methyl N',N'-dimethyl-N-[(methylcarbamoyl)oxy]-1-thiooxamimidate (oxamyl) at 0, 300, 600, and 900 mg·L^{-1} and four levels of 1-naphthaleneacetic acid (NAA) at 0, 2, 4, and 6 mg·L^{-1}. In 1998 a CRD was used for a factorial experiment with two levels of oxamyl (0 and 600 mg·L^{-1}) and six levels of NAA (0, 1, 2, 3, 4, and 5 mg·L^{-1}). Both years there were three single-tree replications per treatment.

Data collection. Data were collected in the same manner each year. At harvest, 20 representative fruit were selected from around the periphery of each tree at ≈1.5 to 2.2 m above ground. The sample mean fruit weight (SMFW) was calculated by dividing the total weight for the sample by 20. During bloom, three

Received for publication 31 Aug. 2000. Accepted for publication 2 Mar. 2001. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

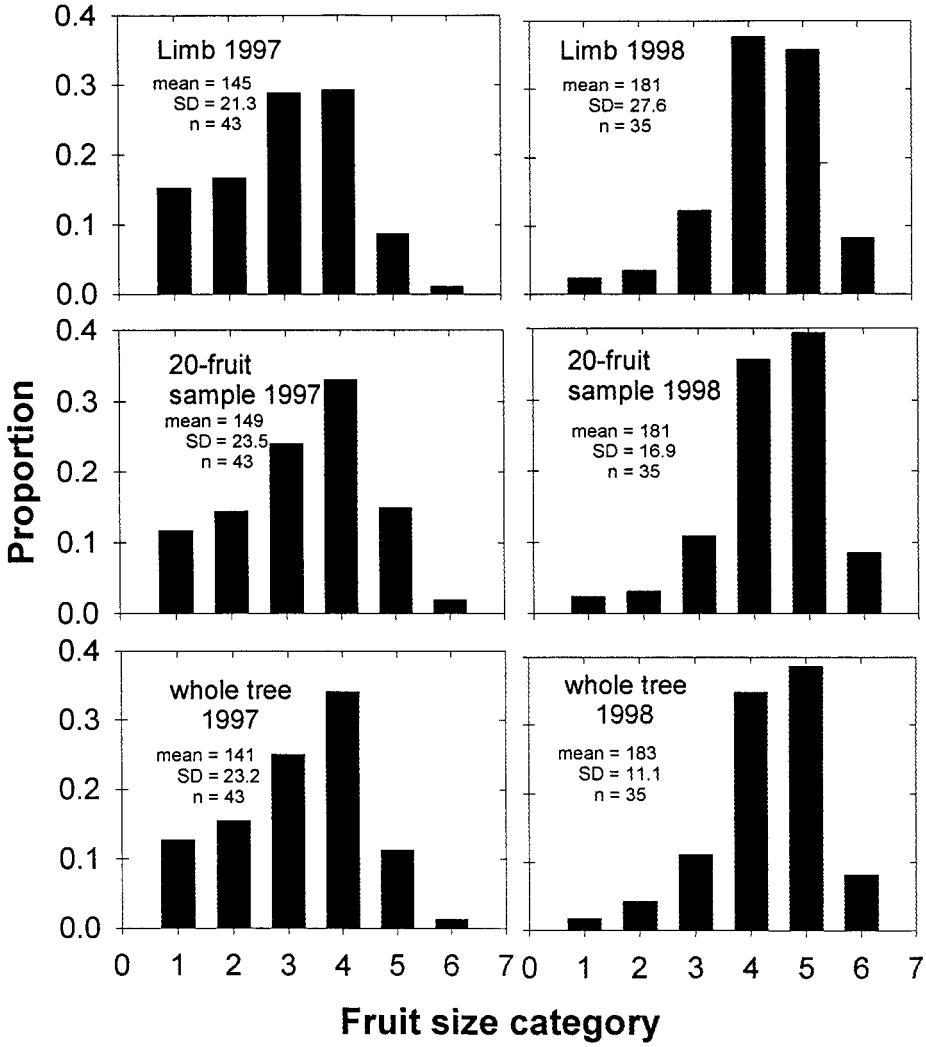

Fig. 1. Proportion of fruit in six size categories resulting from packouts of fruit on three limbs per tree, a random sample of 20 fruit per tree, and all the fruit on a tree in 1997 and 1998. Mean fruit weight for the size categories was: 1 = 91 g, 2 = 119 g, 3 = 142 g, 4 = 174 g, 5 = 200 g, and 6 = 281 g. Included are means, standard deviations (SDs) and number of observations for each sampling method.

limbs per tree were tagged for blossom and fruit set counts. At harvest, all fruit on each limb were counted and weighed to calculate MFW for that limb. The mean value for the three limbs per tree will be referred to as the limb mean fruit weight (LMFW). Sometimes a fruit from one of the tagged limbs was selected for the 20-fruit sample. In such cases, after calculating the SMFW, the same fruit was used to calculate the LMFW. All fruit remaining on the tree were then harvested. The fruit from each 20-fruit sample, each limb, and the fruit remaining on each tree were then segregated into six size categories with a weight sizer. The proportion of fruit falling into each category per tree was calculated for each sampling method. To obtain the true mean fruit weight (TMFW) for each tree, all fruit from a tree were combined and the total weight of the crop was divided by the number of fruit. The crop value (value of all the fruit on a tree) for each sampling method per tree was then calculated using prices for Appalachian 'Delicious' during Oct. 1997, where the value per kilogram for different fruit grades was: juice (<64 mm diam) = $0.088 (U.S. dollars), bags = $0.49, 113 count = $0.50, 100 count = $0.55, 88 count & larger = $0.61. The value of a fruit was calculated by dividing the crop value for the tree by the number of fruit in the sample.

Statistical analyses. Graphical procedures were used to compare the sampling methods to the TMFW (Bland and Altman, 1986). Scatter plots of values for each method against those of the true values are presented for each year. REG Procedure of SAS (Freund and Littell, 1991) was used to evaluate the linear relationships for LMFW and SMFW against TMFW for each year. Values for LMFW and SMFW were subtracted from TMFW for each tree per year. For each method of estimating MFW, differences were plotted against the TMFW for a tree. Lack of agreement was evaluated by calculating the relative bias, estimated by the mean of the differences ($\bar{d}$), and the standard deviation (SD) of the differences. Normality was visually evaluated with a frequency distribution of individual differences. Univariate Procedure of SAS (SAS Inst., Inc., 1990) was used to obtain a Wilkes-Shapiro W statistic as a formal test for normality.

To determine if sampling methods for estimating MFW may influence conclusions drawn from an experiment, GLM Procedure of SAS was used to perform an analysis of variance (ANOVA) for each data set. Orthogonal polynomials were used to evaluate the effects of chemical concentrations on the response variables. The same procedures were used to compare the sampling methods for estimating MFV.

Results and Discussion

Fruit size distribution and descriptive statistics for fruit weight are presented in Fig. 1. The number of fruit sampled per tree was constant for the 20-fruit sample, but for the three-limb sample in 1998, the total number of fruit sampled per tree varied from 0 to 90 (data not presented). The percentage of the total crop sampled averaged ≈17% and 5%, respectively, for the three-limb method and the 20-fruit method (data not presented).

Distributions of fruit in the six size classes resulting from the two sampling methods roughly approximated the distribution for the whole trees (Fig. 1). For both years the three-limb method slightly overestimated the proportion of fruit in the fourth size and underestimated the proportion of fruit in the fifth size. In 1998, the 20-fruit method slightly overestimated the proportion of fruit in smallest size and largest sizes. Not only did the distributions appear similar in 1997, estimates for means and SDs from both sampling methods were similar to the population parameters. To evaluate bias due to sampling method, a paired *t* test was performed on the absolute differences (SMFW–TMFW and LMFW–TMFW). The null hypothesis was that the mean of the differences was equal to zero. The probability of a greater *t* was >0.30 for both methods, indicating that for both sampling methods the bias (absolute differences) was similar over the entire range of true fruit weights. The SDs estimated from each sample data set were compared to the population SD obtained from the whole tree (σ)

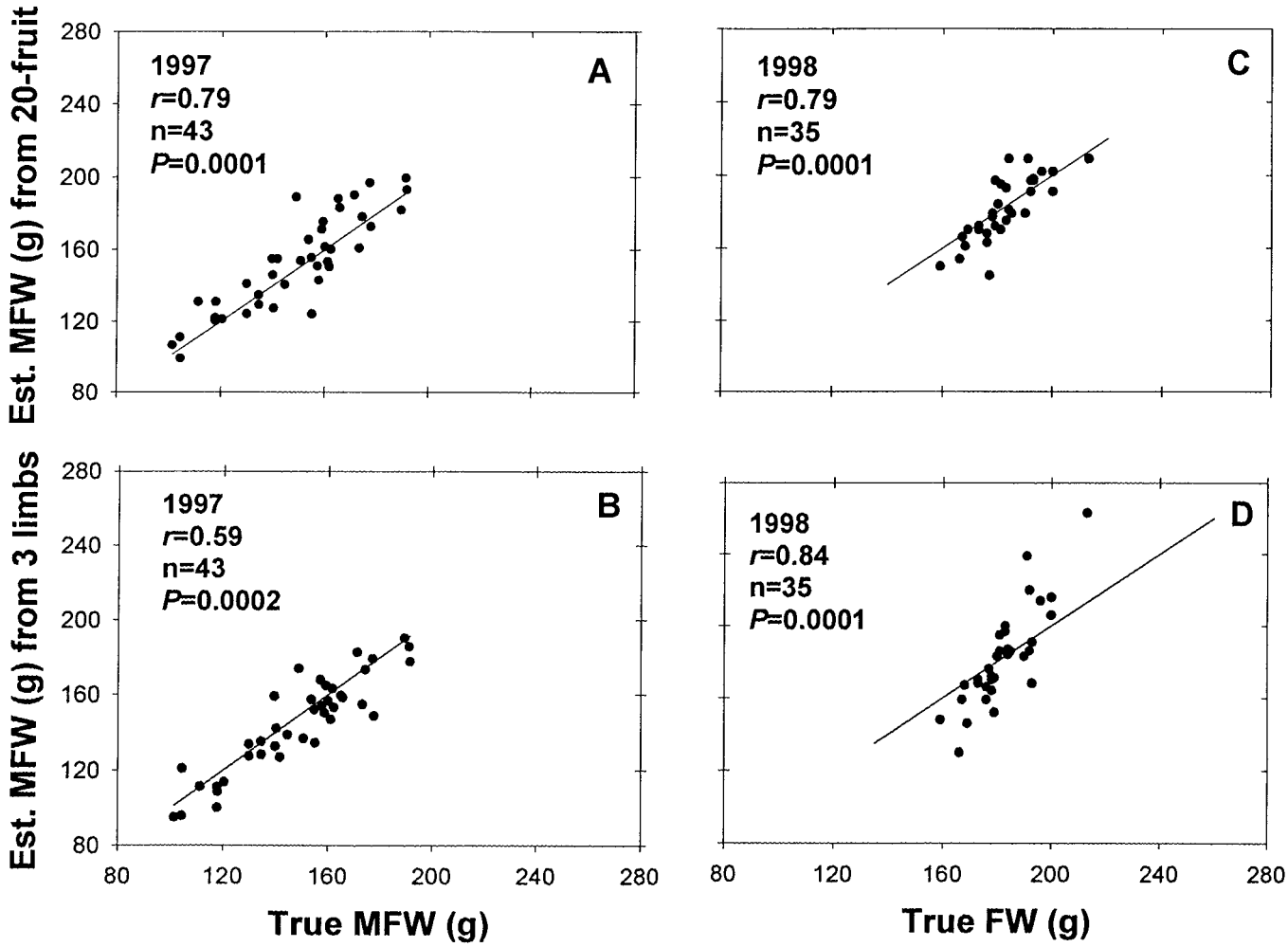

Fig. 2. (**A** and **C**) Plots of mean fruit weight (MFW) estimated from a 20-fruit sample (SMFW) vs. the true mean fruit weight (TMFW) calculated from all the fruit on a tree and (**B** and **D**) MFW estimated from all fruit on three limbs per tree (LMFW) vs. TMFW. Along with estimated MFW for each tree is the line of equality on which all points would lie if the estimation method gave the true value for every tree.

using the formula $(n-1/SD^2) \times \sigma^2$. This statistic follows a chi-square distribution. The probability of a greater χ^2 for both sampling methods was >0.25, indicating the SDs estimated from the sampling methods were similar to the population SD. Paired *t* tests for 1998 data also indicated the bias due to both sampling methods was similar over the range of true MFW values. The hypothesis that the mean of the absolute differences was equal to zero was not rejected for both sampling methods $P > 0.333$. However, in 1998 both sampling methods provided unacceptable estimates of the SDs because the hypothesis that SD = σ was rejected ($P < 0.005$).

Estimates for LMFW and SMFW were plotted against the true values and plots contain a line of equality on which all points would lie if the estimation methods gave exactly the same value as the true value (Fig. 2). Data for 1997 tend to be clustered fairly close to the line, but in 1998 there is more variation for both estimation methods. Correlation coefficients were calculated for each plot to test the null hypothesis that the values for the estimation methods are not linearly related to the true values. Values obtained from the estimation methods were linearly related to the true values ($r > 0.59$, $P < 0.0001$).

Method comparison studies are frequently misanalyzed (Altman, 1991; Altman and Bland, 1983). Correlation between the values of two methods is often calculated, and a high correlation coefficient is interpreted as an indication of good agreement. According to Altman (1991), there are several reasons why correlation is an inappropriate analysis. 1) The correlation coefficient is a measure of strength of linear association between two variables, which is not the same as a measure of agreement. 2) There may be a high degree of correlation when the agreement is poor. There is perfect agreement only if the points lie along the line of equality, but there is perfect correlation if points lie along any straight line. Data in Fig. 2D clearly illustrate this point. LMFW values were highly correlated with TMFW values ($r = 0.84$, $n = 35$, $P = 0.0001$), but LMFW clearly underestimates true MFW at low values of TMFW and overestimates true MFW at high TMFW values. 3) The test of significance may show that two methods are related, but the significance is irrelevant to the question of agreement. In most situations, a correlation significantly different from zero indicates good agreement. For two methods that are assumed to measure the same thing, the agreement between them may not be close, because what may be a high correlation in other contexts is not high when comparing things that should be highly related. It can be wrong to infer from a high correlation that the methods may be used interchangeably (Hallman and Teramo, 1981). In this case $r = 0.84$, indicating that nearly 30% of the variation in TMFW is not explained by the variation in LMFW. Therefore, LMFW would not be a very good predictor of TMFW. 4) The least-squares parameters (slope, intercept, standard error of estimate, and *t* test) may be inaccurate when

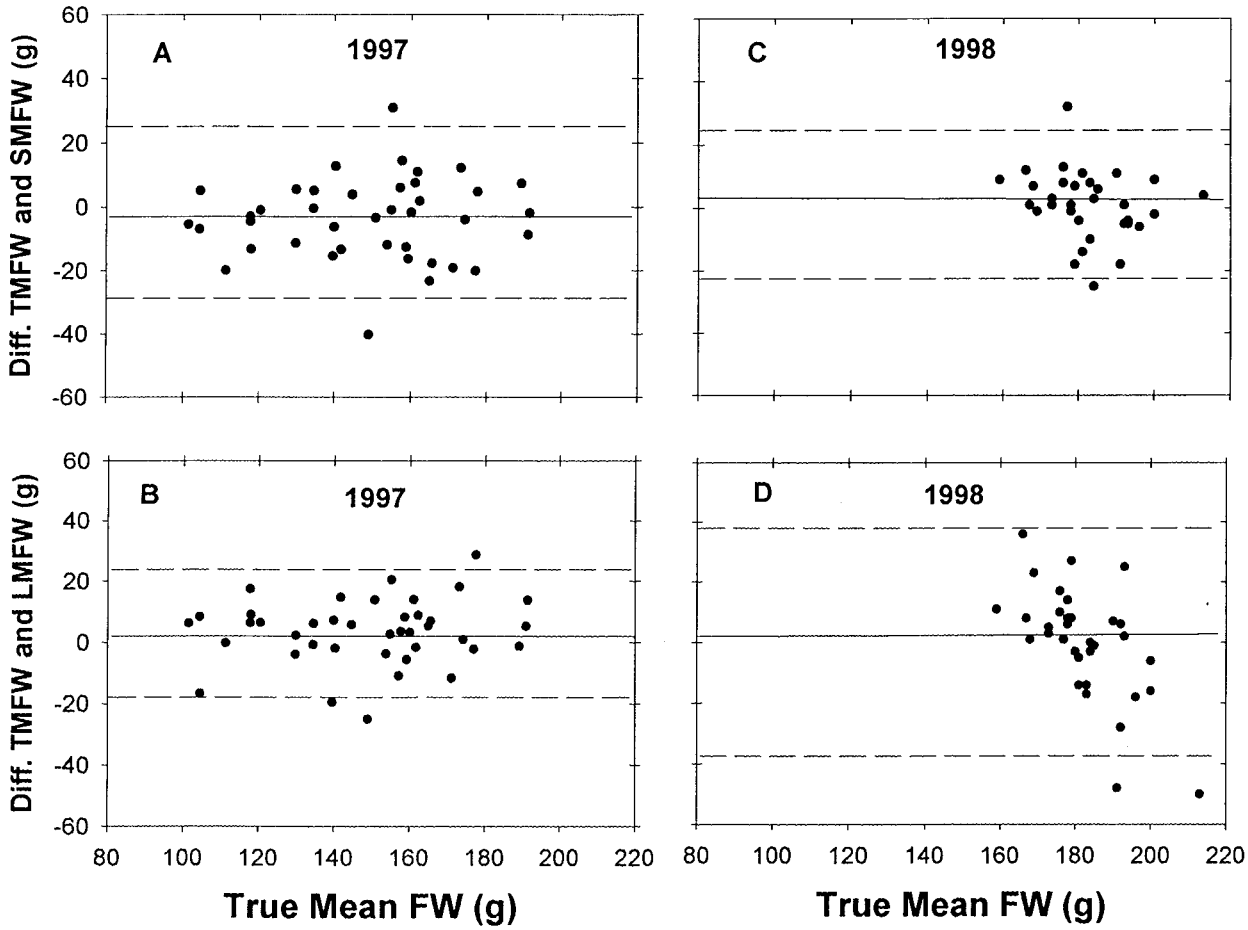

Fig. 3. (**A** and **C**) The difference between mean fruit weight (MFW) estimated from a 20-fruit sample (SMFW) vs. the true mean (TMFW) calculated from all fruit on a tree and (**B** and **D**) the difference between MFW estimated from all fruit on three limbs per tree (LMFW) vs. TMFW. The solid line is the mean of the differences. The broken lines are the limits of agreement, calculated as $d \pm 2SD$; where d = the mean of the differences and SD = the standard deviation of the differences. If the differences are normally distributed, 95% of the differences in a population will lie between the limits of agreement.

random error is large and the range of data is small. Correlation depends on the range of the true values in the sample. If the range is large, the correlation will be greater than if it is small. Estimated values in Fig. 1C and D are highly correlated with true values, but the estimated values clearly do not agree with the true values.

Examining the data with plots as in Fig. 2 is useful, but it is difficult to assess between-method differences. A plot of the difference between the methods against the true values may be more informative (Bland and Altman, 1986). Considerable lack of agreement between the values of MFW obtained with the two estimating methods and the true values is apparent in Fig. 3. Plotting differences against the true value also allows investigation of possible relationships between the measurement error and the true values. Three plots (Fig. 3A–C) indicate no obvious relation between the differences and the true values. The difference between the TMFW and LMFW in 1998 (Fig. 3D) shows that the differences were positive for lower values of TMFW and differences were negative for larger values of TMFW and the relationship is somewhat linear.

Lack of agreement between an estimation method and the true value can be evaluated by calculating the bias, estimated by the mean of the differences (d) and the SD of the differences. If there is consistent bias, an adjustment can be made by subtracting $\bar{d}$ from the new estimation method. In Fig. 3A–D a solid line represents the mean of the differences. If the differences are normally distributed, 95% of the differences will lie between $d \pm 2SD$. These values are referred to as the limits of agreement. Such differences are likely to follow a normal distribution because much variation between subjects was removed and the measurement error remains (Bland and Altman, 1986). The frequency distributions and the W statistics for the four plots of the differences indicate that the data adequately approximate Normality for parametric tests to be valid (Fig. 4) (Anderson and McClean, 1974).

Provided that differences within the limits of agreement are not important, the estimation methods can be used instead of calculating the true MFW. Estimates from the 20-fruit sample may be 24 g below or 28 g above the true value in 1997 (Fig. 3A), and 22 g below and 24 g above the true value in 1998 (Fig. 3C). Estimates from sampling limbs may be 18 g below and 25 g above the true value in 1997 (Fig. 3B), and 36 g below and 38 g above the true value in 1998 (Fig. 3D). Most values for TMFW were <180 g in 1997 and < 200 g in 1998. Thus the 20-fruit sample method provided estimates within ≈13% of the true mean both years. The limb-sampling method provided estimates within ≈11% of the true mean in 1997 and within ≈19% of the true mean in 1998. Pomologists must decide if this level of error is acceptable. Estimates of MFW are often obtained from samples of 10 to 30 apples per tree and treatment means usually differ by 8% to 20% (Dozier, et al., 1980; Greene, 1986; Miller, 1982; Stover et al., 1999). Apple prices are usually related to fruit sizes (box count). Difference in MFW between 125-count and 100-count

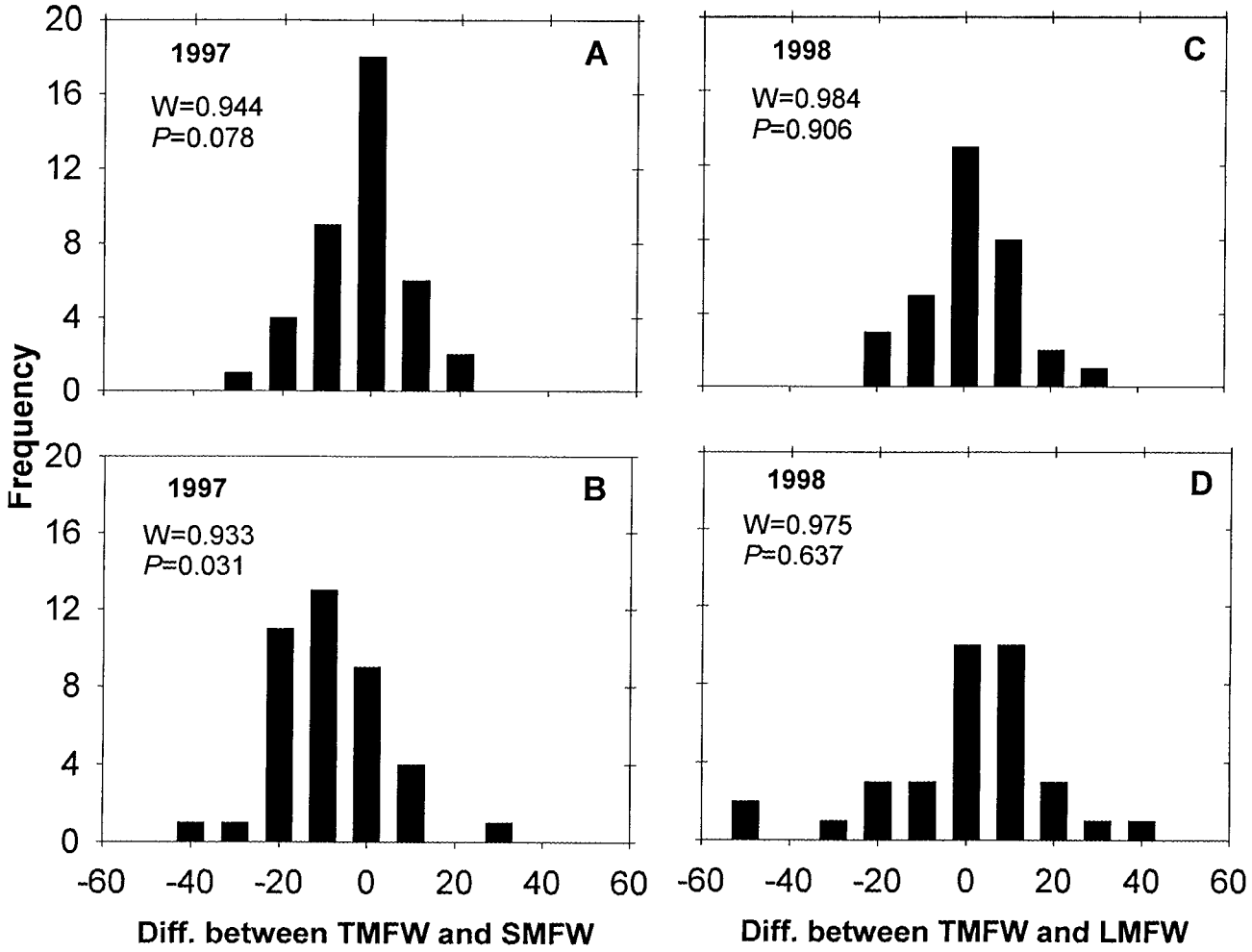

Fig. 4. Frequency plots of the differences plotted in Fig. 3. Also presented are coefficients and *P* values for Wilkes Shapiro W test where the null hypothesis is that the differences come from a normal distribution.

fruit is 20% and the difference between 100-count and 88-count fruit is 12%. If the TMFW for an experiment is ≈191 g (100-count), the measurement error associated with the estimation methods used in this study is on the order of one box size. Thus, when using these methods, treatment differences of one box size may be due to measurement error. Pomologists would probably consider the level of error associated with these two methods of estimating MFW too high to detect treatment differences that would be considered biologically and economically important.

Dorsey and McMunn (1938) evaluated five different methods for estimating apple fruit size. Methods that did not provide accurate estimates included sampling 20 fruit per tree, sampling 100 fruit per tree, harvesting all the fruit on one limb per tree, and harvesting all the fruit on a vertical section of the tree. Harvesting 91 kg/tree, which was ≈25% of the crop, provided estimated MFWs within ≈10% of the true mean. They concluded that conscious attempts to select small representative samples are unreliable and that those methods that most nearly approximate a random sample should be used.

The bias due to sampling method for estimating fruit values was tested with paired *t* tests. Probability of a greater *t* was >0.25 for both sampling methods in 1997 and 1998, indicating the sampling methods had consistent bias over the entire range of data and that the mean of the differences did not significantly differ from zero.

Plotting estimated mean fruit values against the true mean fruit values shows that the estimates from the 20-fruit samples are clustered around the line of unity for 1997 and 1998 (Fig. 5A and C). Estimates from limb samples are clustered around the line of unity in 1997, but in 1998 deviation from the line of unity was high at true mean fruit values >0.09.

Estimated fruit values from the 20-fruit samples were within $0.04 below to $0.035 above the true value in 1997 and $0.02 above to $0.02 below the true value in 1998 (Fig. 6A and C). In 1998, at high true mean fruit values, some of the estimates from the limb samples substantially underestimated the true values. Estimates from the limb-sampling method were within $0.037 below to $0.043 above the true value in 1997 and $0.02 below to $0.019 above the true value in 1998. These seemingly small differences are magnified when multiplied by the number of fruit on a tree. The mean number of fruit per tree was ≈500 and 300 in 1997 and 1998, respectively. The crop value estimated from the 20-fruit sample would have been underestimated by $20.00 and $6.10 per tree in 1997 and 1998, respectively. The limb method sample overestimated true crop value by $21.50 per tree in 1997 and underestimated the crop value by $18.70 per tree in 1998. Depending on the year and the sampling method, the estimated crop value would have deviated from the true crop value by $3,148 to $11,094/ha (516 trees/ha).

To determine if interpretation of experimental results would be affected by the method used for estimating MFW, each of the six data sets was analyzed with ANOVA and *P* values for the main

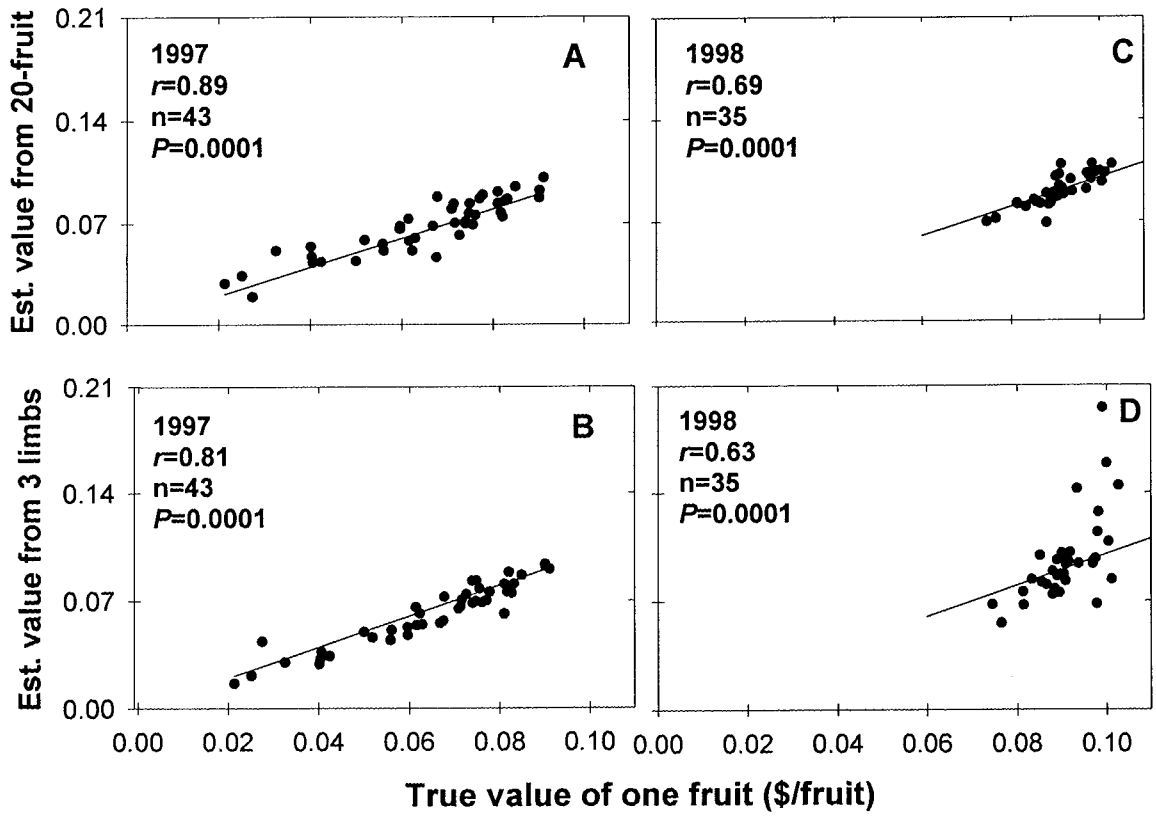

Fig. 5. (**A** and **C**) Plots of mean fruit value ($/fruit) estimated from a 20-fruit sample (SMFV) vs. the true mean fruit weight (TMFV) calculated from all the fruit on a tree and (**B** and **D**) mean fruit weight (MFW) estimated from all fruit on three limbs per tree (LMFV) vs. TMFV. Along with estimated MFW for each tree is the line of equality on which all points would lie if the estimation method gave the true value for every tree.

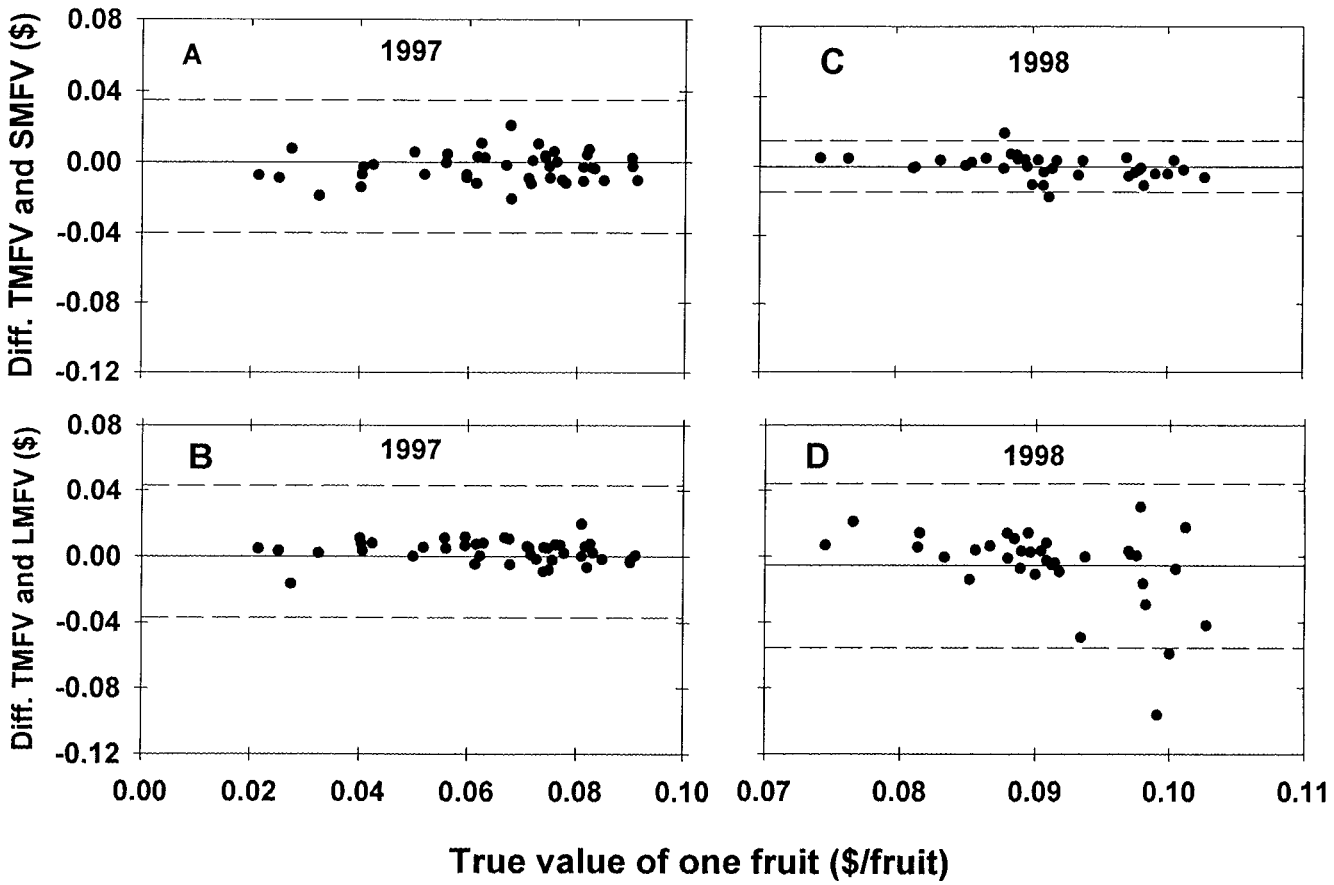

Fig. 6. (**A** and **C**) The difference between mean fruit value (MFV) estimated from a 20-fruit sample (SMFV) vs. the true mean (TMFV) calculated from all the fruit on a tree and (**B** and **D**) the difference between MFV estimated from all fruit on three limbs per tree (LMFV vs. TMFV. The solid line is the mean of the differences. The broken lines are the limits of agreement, calculated as $d \pm 2sd$ where d = the mean of the differences and sd = the standard deviation of the differences. If the differences are normally distributed, 95% of the differences in a population will lie between the limits of agreement.

Table 1. *P* values for main effects and interactions from apple thinning experiments in 1997 and 1998 where mean fruit weight (MFW) was calculated by weighing and counting all the fruit on each tree (TMFW), the MFW was estimated from a 20-fruit sample per tree (SMFW), or the MFW was estimated by harvesting all fruit on three limbs per tree (LMFW).

1997				1998			
Source of variation	TMFW	SMFW	LMFW	Source of variation	TMFW	SMFW	LMFW
NAA (N)	0.442	0.412	0.445	N	0.128	0.435	0.985
Lin	0.740	0.353	0.362	Lin	0.408	0.867	0.851
Quad	0.120	0.157	0.522	Quad	0.859	0.759	0.679
Oxamyl (O)	0.001	0.001	0.001	O	0.001	0.015	0.005
Lin	0.001	0.001	0.001	N × O	0.201	0.208	0.224
Quad	0.035	0.109	0.012	$N_{Lin} \times O_{Lin}$	0.091	0.480	0.070
N × O	0.270	0.774	0.049	$N_{Quad} \times O_{Lin}$	0.991	0.708	0.792
$N_{Lin} \times O_{Lin}$	0.799	0.884	0.093				
$N_{Quad} \times O_{Lin}$	0.052	0.183	0.325				
$N_{Lin} \times O_{Quad}$	0.874	0.757	0.038				
$N_{Quad} \times O_{Quad}$	0.492	0.585	0.221				

Table 2. *P* values for main effects and interactions from apple thinning experiments in 1997 and 1998 where mean fruit value (MFV) was estimated from all the fruit on each tree (TMFV), MFV was estimated from a 20-fruit sample per tree (SMFV), or MFV was estimated by harvesting all fruit on three limbs per tree (LMFV).

1997				1998			
Source of variation	TMFW	SMFW	LMFW	Source of variation	TMFW	SMFW	LMFW
NAA (N)	0.233	0.389	0.808	N	0.126	0.602	0.849
Lin	0.801	0.594	0.783	Lin	0.782	0.582	0.959
Quad	0.048	0.103	0.535	Quad	0.330	0.951	0.863
Oxamyl (O)	0.001	0.001	0.001	O	0.002	0.024	0.022
Lin	0.001	0.001	0.001	N × O	0.222	0.263	0.473
Quad	0.004	0.027	0.022	$N_{Lin} \times O$	0.038	0.643	0.289
N × O	0.199	0.797	0.528	$N_{Quad} \times O$	0.266	0.531	0.628
$N_{Lin} \times O_{Lin}$	0.965	0.553	0.513				
$N_{Quad} \times O_{Lin}$	0.024	0.079	0.276				
$N_{Lin} \times O_{Quad}$	0.614	0.902	0.694				
$N_{Quad} \times O_{Quad}$	0.569	0.711	0.479				

effects and interactions are presented in Table 1. In 1997 the ANOVA for TMFW indicated that NAA did not significantly influence MFW, there was a quadratic effect of oxamyl concentration, and the $NAA_q \times oxamyl_l$ interaction was significant (P = 0.052). Analysis of estimates from the 20-fruit sample indicated that NAA did not influence MFW, there was a linear effect of oxamyl, and the NAA × oxamyl interaction was not significant (P = 0.18). Analysis of estimates from limb samples indicated that NAA did not influence MFW, there was a quadratic response to oxamyl, and the $NAA_l \times oxamyl_q$ interaction was significant (P = 0.038). In 1998 all three analyses indicated that the main effect of NAA was not significant ($P > 0.128$) and the main effect of oxamyl was significant ($P < 0.015$). However, the P values for the × NAA_l × oxamyl interactions were 0.09, 0.48, and 0.07 for TMFW, SMFW, and LMFW, respectively. Some pomologists might consider P values <0.1 for interaction terms to be meaningful for thinning experiments (Marini, 1999).

Results for fruit value also indicate nonagreement between the three sampling methods (Table 2). In 1997, all three methods indicated that NAA did not significantly affect fruit value, but only the analysis of the whole-tree data indicated there was a quadratic effect of NAA on fruit value. All methods detected significant linear and quadratic effects of oxamyl, but the NAA_q × $oxamyl_l$ interaction was significant for only the whole-tree estimates. In 1998, a significant effect of oxamyl and a nonsignificant effect of NAA was detected for all sampling methods. However, the NAA_l × oxamyl interaction was detected only when estimates were based on whole trees. These analyses show that conclusions concerning treatment effects on MFW and MFV estimated from a 20-fruit sample may be incorrect. Although better than the 20-fruit sample, conclusions based on estimates from harvesting all fruit on three limbs/tree also did not agree very well with the true values. Conclusions from thinning experiments, performed by different researchers, with similar treatments often differ. One reason for differing conclusions may be that researchers use different sampling schemes to estimate MFW. Results from ANOVA show that different sampling schemes may lead to different conclusions.

In this study, test statistics generated with least squares analyses indicated MFW and MFV could be estimated fairly accurately from a 20-fruit sample or by weighing all fruit on three limbs per tree because sample means and SDs did not significantly differ (P = 0.05) from the population parameters. Additionally, paired t tests indicated the mean differences were symmetrical around the true mean and correlation analyses indicated sample values were correlated with true values. However, these results are misleading

because graphical techniques showed both sampling schemes provided estimated values that did not consistently agree with the true values.

Sampling procedures for estimating MFW and MFV of apple trees have not been developed and sources of variation for FW within apple trees have not been identified and quantified. There is a need for experiments designed to obtain estimates of the important sources of variability so trees can be properly sampled to accurately estimate MFW and MFV. Until results of such studies are published, it seems that researchers should count and weigh all fruit on a tree to obtain the true MFW for each tree in an experiment.

Literature Cited

Altman, D.G. 1991. Practical statistics for medical research. Chapman and Hall. New York.

Altman, D.G. and J.M. Bland. 1983. Measurement in medicine: The analysis of method comparison studies. The Statistician 32:307–317.

Anderson, V.L. and R.A. McClean. 1974. Design of experiments: A realistic approach. Marcel Dekker, Inc. New York.

Barnett, R.N. 1965. A scheme for the comparison of quantitative methods. Amer. J. Clinical Pathol. 43:562–567.

Barnett, R.N. and W.J. Youden. 1970. A revised scheme for the comparison of quantitative methods. Amer. J. Clinical Pathol. 54:454–460.

Barritt, B.H., B.S. Honishi, and M.A. Dilley. 1995. Performance of three apple cultivars with 23 dwarfing rootstocks during 8 seasons in Washington. Fruit Var. J. 49:158–170.

Bland, J.M and D.G. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancets I:307–310.

Broughton, P.M.G., M.A. Buttoph, A.H. Growenlock, D.W. Neill, and R.G. Skentelbery. 1969. Recommended scheme for the evaluation of instruments for automatic analysis in the clinical biochemistry laboratory. J. Clinical Pathol. 22:278–283.

Dorsey, M.J. and R.L. McMunn. 1938. A comparison of different methods of taking samples of apples in experimental plots. Proc. Amer. Soc. Hort. Sci. 36:619–626.

Dozier, Jr., W.A.,C.C. Carlton, and K.C. Short. 1982. Effects of mechanical and hand pruning, tree spacing, and limb bending on tree development and yield on hedgerow 'Delicious' apples on Malling Merton 106 rootstock. J. Amer. Soc. Hort. Sci. 107:295–298.

Dozier, Jr., W.A., W.A. Griffey, and H.F. Burgess. 1980. Effect of growth regulators on the development of 'Delicious' apples. HortScience 15:743–744.

Freund, R.J. and R.C. Littell. 1991. SAS system for regression. SAS Inst., Cary, N.C.

Greene, D.W. 1986. Effect of paclobutrazol and analogs on growth, yield, fruit quality, and storage potential of 'Delicious' apples. J. Amer. Soc. Hort. Sci. 111:328–332.

Hallman, M. and K. Teramo. 1981. Measurement of the lecithin/sphingomyelin ration and phosphatidylglycerol in amniotic fluid: An accurate method for the assessment of fetal lung maturity. British J. Obstetrics Gynaecol. 88:806–813.

Hampson, C.R., H.A. Quamme, and R.T. Brownlee. 1997. Performance of dwarfing rootstocks in five trials in British Columbia, Canada. Fruit Var. J. 51:183–191.

Laurent, R.T.S. 1998. Evaluating agreement with a gold standard in method comparison studies. Biometrics 54:537–545.

Marini, R.P. 1999. Are nonsignificant differences really not significant? HortScience 34:761–762.

Miller, S.S. 1982. Regrowth, flowering, and fruit quality of 'Delicious' apple trees as influenced by summer pruning. J. Amer. Soc. Hort. Sci. 107:975–978.

SAS Inst., Inc. 1990. SAS/STAT user's guide. Version 6, 4th ed. vol. 2. SAS Inst. Cary, N.C.

Stover, E., M. Fargione, R. Risio, W. Stiles, and K. Iungerman. 1999. Prebloom foliar boron, zinc, and urea applications enhance cropping of some 'Empire' and 'McIntosh' apple orchards in New York. HortScience 34:210–214.

Westgard, J.O. and M.R. Hunt. 1973. Use and interpretation of common statistical tests in method-comparison studies. Clinical Chem. 19:49–57.