J. Field Ornithol., 61(3):339–346

CAN SAW-WHET OWLS BE SEXED BY EXTERNAL MEASUREMENTS?

HELMUT C. MUELLER
Department of Biology
University of North Carolina
Chapel Hill, North Carolina 27599-3280 USA

Abstract.—Measurements of wing chord are being used to determine the sex of individual Northern Saw-whet Owls (*Aegolius acadicus*). The published criteria for sexing the species have a limited basis in reality and many owls are being sexed incorrectly. The criteria are based on biased samples of measurements and improper use of statistics. Furthermore, we do not need sexing criteria based on size; most questions involving differences between the sexes can be examined by using statistics on the data gathered by a bander.

¿PUEDE DETERMINARSE EL SEXO DE *AEGOLIUS ACADICUS* MEDIANTE MEDIDAS EXTERNAS?

Sinopsis.—Medidas del ala están siendo utilizadas para determinar el sexo en búhos (*Aegolius acadicus*). Los criterios utilizados en publicaciones tienen grandes limitaciones al aplicarse a la realidad, y el sexo de las aves está siendo identificado incorrectamente. Los criterios están basados en muestras de medidas con sesgos y en el uso inapropiado de las estadísticas. Además, no se necesitan criterios para identificar el sexo de las aves basados en tamaño; la mayoría de las interrogantes que envuelven la diferencia entre sexos pueden ser analizadas aplicando métodos estadísticos a los datos obtenidos cuando se anillan aves.

Anonymous (1980), Buckholtz et al. (1984), Sheppard and Klimkiewicz (1976), and Weir et al. (1980), have published criteria for sexing live Northern Saw-whet Owls (*Aegolius acadicus*) by wing chord measurement. I show that these criteria are of dubious accuracy and limited utility because: (1) correct use of sampling statistics shows that only a relatively small proportion of the owls can be sexed, and of these, a relatively high proportion will be sexed incorrectly; (2) wing chords of birds decrease when they are made into museum specimens and dried and criteria based on museum specimens are inappropriate for live birds; (3) the only available sample of wing chord measurements from internally sexed museum specimens is significantly biased and thus yields biased sexing criteria; (4) Buckholtz et al. (1984) improperly used a method that can produce estimates of sex ratios to produce sexing criteria that have no validity; (5) the samples of Buckholtz et al. (1984) contain significant rounding errors and are of limited utility in any analysis; (6) age changes in wing chord; and (7) geographic variation in size may further add to the error and uncertainty of sex determination.

CORRECT USE OF SAMPLING STATISTICS

The Bird Banding Laboratory (Anon. 1980) indicates that Saw-whet Owls with wing chords of 131 mm or less are males and those with wing chords of 143 mm or more are females. These criteria, and similar ones of Sheppard and Klimkiewicz (1976) and Weir et al. (1980) appear to be based on Earhart and Johnson (1970) who provided the means and standard errors of wing chord for 57 internally sexed museum specimens. At first glimpse, the criteria appear reasonable: if we assume that the samples of Earhart and Johnson (1970) have normal distributions, then 99.89% of the females would have wing chords greater than 131.4 mm and thus virtually all individuals with wing chords of 131 mm or less would be males. Similarly, 99.84% of males would have wing chords of less than 143.5 mm and thus virtually all larger individuals would be females. Unfortunately, it is improper to use the statistics of a sample to comment on the identity of an individual drawn from without the sample.

The mean and standard deviation of a sample are only estimates of the mean and standard deviation of the population. The mean for males in the sample of Earhart and Johnson (1970) is 132.2 mm, the standard error of the mean is 0.63, the standard deviation is 3.83 and the standard error of the standard deviation is 0.45. At the 99% confidence interval, the mean for males in the population lies between 130.49 and 133.91, the standard deviation lies between 2.61 and 5.06 and the range is 117.5–146.9 mm. Figures 1 and 2 of Buckholtz et al. (1984) present the distributions of wing chord measurements for 1577 HY (birds hatched in that calendar year) and 1011 AHY (birds hatched in any year prior to the year of capture) Saw-whet Owls. Using the criterion of Anon. (1980), 9.5% of the HY birds would be sexed female. If we assume a sex ratio of unity and use my calculations given above, as many as 28.7% of these "females" might actually be males. Similar calculations for AHY show that 15.2% would be sexed female and as many as 27.9% of these might actually be males.

The females in the sample of Earhart and Johnson (1970) have a mean wing chord of 139.0, a standard error of the mean of 0.55, a standard deviation of 2.46 and a standard error of the standard deviation of 0.39. At the 99% confidence interval, the mean of the population of females lies between 137.43 and 140.57, the standard deviation lies between 1.34 and 3.57 and the range is 128.2–149.8. Using the criterion of Anon. (1980), 24.4% of the HY sample of Buckholtz et al. (1984) would be sexed as females and as many as 5.4% of these might actually be males. Similarly, 14.7% of AHY would be sexed as females and up to 14.1% might actually be males.

Sexing only 10% to 24% of a sample with a possible error of 5% to 29% has little scientific merit. Correct use of sampling statistics yields estimates that the longest male wing chord would be 146.9 mm and the shortest female wing chord would be 128.2 mm, both at the 99% confidence interval, yielding criteria of $\geq$148 mm for females and $\leq$127 for males. These criteria would identify 5.3% of HY as male and 0.6% as females, with an error of only 0.5% in both cases. For AHY, 2.6% of the birds would be sexed as male and 1.0% as female, both with an error of 0.5%. If these criteria are used, females are twice (AHY) to eight times (HY) as likely to occur in any sample, leading an investigator using only data from sexed birds to conclude that females are more likely to do whatever is being investigated.

MEASUREMENTS OF MUSEUM SPECIMENS

The calculations in the preceding section assume that the data of Earhart and Johnson (1970) constitute a random and unbiased sample of the statistical population from which banded birds are taken. This is not true: the wing measurement of a bird decreases during the first few months after it is made into a museum specimen. There is no measurement of this shrinkage for any owl, but wing chord measurements taken from freshly dead individuals in samples from seven species were an average of 1.72% longer than those taken from the same museum skins six months or more later (Greenwood 1979, Green 1980, Knox 1980). A difference of this magnitude is about 34% of the difference between the means of the sexes of Saw-whet Owls and obviously would have considerable effect on sexing criteria.

THE SAMPLE OF INTERNALLY SEXED MUSEUM SPECIMENS

Earhart and Johnson (1970) present wing chord measurements for 30 taxa of North American owls. The Saw-whet Owl is the only one to show significant difference between the sexes in the variance in wing chord ($F = 2.43$, $n = 37, 20$, $P < 0.05$, two-tailed). A significant difference at the 5% level in 1 of 30 samples is about what is expected by chance and this appears to be the best explanation for the bias in the sample.

The biased sample results in a biased estimate of the characteristics of the population and in biased sexing criteria. The mean coefficient of variation for all 60 samples of Earhart and Johnson is 2.67. I will use this value to illustrate the effects of biased variances on sexing criteria: the proportion of birds sexed as males would decrease from 8.8% to 2.5% for HY and from 4.4% to 0.1% for AHY. Birds sexed as females would increase from 1.1% to 6.4% for HY and from 2.2% to 6.3% for AHY. These calculations assume that only the variances of Earhart and Johnson are biased; the means are suspect but there is no way of guessing the direction of bias. Note that this slight change reverses the bias in the sex ratio obtained using the variances of Earhart and Johnson (1970).

THE CRITERIA OF BUCKHOLTZ ET AL.

Buckholtz et al. (1984) offer separate criteria of wing chord length for AHY and HY Saw-whet Owls and make the remarkable claim that wing chord permits the "... correct identification of 50% of each sex and misidentification of only 1% of the other sex ..."! The method used by Buckholtz et al. to obtain their criteria relies entirely on the information contained in their samples of wing chords taken from 1577 HY and 1011 AHY live, unsexed owls and the simple (and valid) assumption that females are larger than males. Buckholtz et al. (1984) used the computer program of MacDonald and Pitcher (1979), which alternates between direct search optimization and fast iterative calculations to produce frequency distributions of size classes in a sample. This program was developed to provide estimates of the proportions of year-size-classes in fish.

The program calculates normal distributions for each size class (the two sexes when used on owls) and seeks the best fit between these calculated distributions and the observed distribution of measurements. It is an excellent method for producing estimates of the numbers of male and female Saw-whet Owls in a sample of unsexed birds but it cannot be used to determine the sex of an individual in the sample with any statistical confidence and definitely should not be used to establish criteria for sexing individuals from without the sample. Further, Buckholtz et al. (1984) have misused the method of MacDonald and Pitcher (1979) and have failed to produce the best estimates of the sex ratio in their own sample. The method requires producing estimates of the mean and variance of a measurement for each sex. Buckholtz et al. (1984) believe that the best estimate of the mean and standard deviation of wing chord for males and females can be obtained by ". . . reducing the number of parameters to be estimated. . . ." Their method of "reduction" involves constraining the sex ratio to 1:1, then constraining the variances of the sexes to equality, and ignoring the goodness of fit between their calculated and observed size-frequency distributions. Their final calculations, on which their sexing criteria are based, result in calculated size-frequency distributions that differ significantly from their observed size-frequency distributions for HY birds ($P < 0.001$) and almost significantly for AHY birds $0.06 < P < 0.07$. Thus, the sexing criteria of Buckholtz et al. (1984) for HY birds have no basis in their data.

The MacDonald program, used correctly, reduces the chi-square values with each step and produces a better goodness of fit between calculated and observed size-frequency distributions. Further, when the goodness of fit cannot be improved by further calculations, the initial estimates of some of the parameters should be revised and the program run again (MacDonald and Pitcher 1979). In both of the examples of MacDonald and Pitcher, the use of revised initial values resulted in chi-square values which were less than the number of degrees of freedom, a considerably better fit than any of those obtained by Buckholtz et al. (1984). In contrast, the calculations of Buckholtz et al. increase the chi-square values for the only step for which they provide details. In this step, they constrained the variance so that it was equal for the sexes, increasing the chi-square for AHY birds from 26.0 (22 df) to 34.4 (23 df) and decreasing the probability that the calculated size-frequency distribution of Buckholtz et al. (on which their sexing criteria are based) is the same as the size-frequency distribution from the owls they netted from $0.20 < P < 0.30$ to $0.06 < P < 0.07$. In a previous step, Buckholtz et al. constrained the sex ratio in their sample to unity. Their Figure 1 appears to indicate a skew to the right, suggesting a female-biased sex ratio in their sample and it is likely that the chi-square value would be further reduced by allowing the sex ratio to vary from unity.

At best, the program of MacDonald and Pitcher (1979) provides only estimates of the means and variances for males and females in the sample. It does not identify individuals as males or females and thus does not have the assured accuracy of means and variances derived from individuals of known sex. The only indicator of the accuracy of the estimates of means and variances is the chi-square goodness of fit between observed and calculated distributions, and the poor fit suggests the probability of considerable error.

THE SAMPLE OF BUCKHOLTZ ET AL.

There are problems with the observed size-frequency distributions of Buckholtz et al. (1984), particularly for the HY owls, which may make it impossible to separate the sexes, regardless of the method used. A size difference between the sexes should result in a bimodal frequency distribution, or at worst, unimodal. For example, Mewaldt and King (1986) found bimodal distributions for wing chord in six of seven samples of White-crowned Sparrows (*Zonotrichia leucophrys*). Male White-crowned Sparrows are 4.5% (HY) and 5.4% (AHY) larger than females and the means for the sexes differ by about two standard deviations. Female Saw-whet Owls are 5.1% larger than males and the means for the sexes differ by about two standard deviations. We would thus expect bimodal distributions in the samples of Buckholtz et al. (1984). Instead, their sample of HY birds shows five peaks at 130, 132, 135, 138 and 140 mm, and four of these (all but 138 mm) differ significantly from both adjacent troughs (G values range from 5.75 to 14.97 and P values from <0.02 to <0.0002). In the entire size-frequency distribution for HY owls, measurements ending in 0, 2, 5, or 8 have more birds recorded than both adjacent measurements in 9 of 13 possible cases while no measurement ending in 1, 3, 4, 6, 7 or 9 shows more individuals than both adjacent measurements in 20 possible cases (Fisher exact $P = 0.004$). All five of the peaks in the distribution are at measurements which end in 0, 2, 5 or 8; this is very improbable (Fisher $P = 0.0022$). It is obvious that the measurers of Buckholtz et al. (1984) had a very strong tendency to round measurements at 0, 2, 5, and 8.

The sample of AHY owls of Buckholtz et al. (1984) is not quite as aberrant as their HY sample but it is biased and multimodal. There are five peaks: 133, 135, 138, 140, and 142 mm. The peak at 135 mm differs significantly from the trough at 134 mm ($G = 3.95$, $P < 0.05$) and the peak at 140 mm differs significantly from the trough at 139 mm ($G = 6.52$, $P < 0.02$). Four of the five peaks occur at measurements ending in 0, 2, 5 or 8, indicating measuring bias (Fisher $P = 0.046$). For the entire distribution of AHY birds, measurements ending in 0, 2, 5 or 8 have more birds recorded than both adjacent measurements in 5 of 11 cases while measurements ending in 1, 3, 4, 6, 7 or 9 have more birds recorded than both adjacent measurements in only one of 18 possible cases (Fisher $P = 0.018$).

The measuring error introduced by preference for certain digits should be reduced by grouping measurement intervals before beginning computations to determine the size-class components in a size-frequency distribution (MacDonald and Pitcher 1979). The rounding of Buckholtz et al. (1984) is due to both "lengthening" the wing of birds with shorter than the preferred measurement and "shortening" the wing of birds with longer than the preferred measurement. The only reasonable way to deal with this problem is to lump the data of Buckholtz et al. in 3 mm increments. The MacDonald program could then be run on the data and this might yield reasonable estimates of the numbers of individuals of each sex. The result might be useful to Buckholtz et al. (1984) in discovering interesting phenomena in their sample but would be of no utility in sexing birds in other samples.

AGE CHANGES IN WING CHORD

Mueller and Berger (1967) found that live, unsexed AHY Saw-whet Owls had wing chords that were 2.0 mm longer than those of HY owls. Buckholtz et al. (1984) found a difference of 1.4 mm. Wing chords are longer in AHY than HY birds in many species and these findings are not surprising. However, the accuracy of either estimate of age difference is questionable because the birds were not sexed. If, for example, there are more females than males in the AHY sample than in the HY sample, the apparent age difference could actually be an artifact of sex bias in the samples. There is differential migration of the sexes in many species of birds and sex ratios of unity in a sample of migrants should not be assumed. Korpimäki and Hongell (1986) found that Boreal Owls (*A. funereus*) captured during migration in Finland had mean wing chords that were similar to those of breeding females and much longer than those of breeding males.

DIFFERENCES IN MEASUREMENT TECHNIQUES

A small, but consistent difference in technique between a given bander and the individual(s) that measured the birds for the reference sample on which the criteria are biased can result in considerable error. The mean wing chord for AHY birds is 138.5 mm for the sample of Mueller and Berger (1967) and 137.0 mm for the sample of Buckholtz et al. (1984). It is likely that most of this difference is an artifact of differences in measuring techniques. This difference is 22.1% of the difference between the means for the sexes in the sample of Earhart and Johnson (1970) and thus would have considerable effect on sexing criteria.

GEOGRAPHIC VARIATION IN SIZE

Another possible source of error is geographic variation in size. There is no evidence for or against this possibility in Saw-whet Owls but geographic variation in wing chord occurs within many species and subspecies.

We know very little about geographic variation in size of North American birds. For example, western Cooper's Hawks (*Accipiter cooperii*) are more than 6% smaller than eastern birds, a highly significant difference and an unknown until the publication of Mueller et al. (1981) and Henny et al. (1985).

RECOMMENDATIONS

The currently available criteria for sexing Saw-whet Owls by wing chord should not be used because of the uncertain accuracy and the high probability of error.

Measurements of live or freshly dead owls of known sex are necessary for the development of accurate criteria. Owls can be sexed by behavioral criteria during the breeding season: only the female incubates and broods and the male provides all of the food for the female at this time (Mueller 1986). I have found it very difficult to sex dead HY owls in fall and winter and internal sexing by laparotomy will require great skill. However, the criteria resulting from a sufficient and unbiased sample of owls of known sex and age will probably only enable us to sex a relatively small fraction of a sample.

Multivariate analysis of several external measurements might result in more useful sexing criteria but variation is considerable between banders in the relatively simple measurement of wing chord and I am dubious about the comparability of other external measurements because of the difficulties involved in measuring live birds.

Sexing criteria based on size are of value only because they: (1) encourage banders to measure their birds; and (2) allow an investigator to answer questions about any differences between the sexes in the recoveries of banded birds, using the data from all banders. Most questions about the sexes can be answered by using the measurements without knowing the sex of the individual. Indeed, it is more efficient, and probably more accurate, to use all of the data in a sample rather than discarding all but a few large females and small males. Examining size-frequency distributions can reveal both errors and interesting phenomena that would go unnoticed if one merely assigns sex to individuals.

Testing a hypothesis involving differences between the sexes can be done by comparing samples. For example, Weir et al. (1980) used sexing criteria to suggest that females migrated earlier in the fall than males. This finding probably could be established more scientifically by showing that early owls were significantly larger than later owls and noting that this is best explained by the difference in size between the sexes.

The sex ratio in a sample can be estimated by the method of MacDonald and Pitcher (1979) or one of several other similar methods. The simpler method of Mewaldt and King (1986) probably can be modified to provide an estimate of the sex ratio for many samples. An appreciable bias in the sex ratio can be detected by visual examination of the size-frequency distribution or by testing the distribution for fit to a normal distribution.

Sexing criteria for other species using wing chord or other measures of size should be examined in detail. I suspect that other species are being sexed incorrectly.

ACKNOWLEDGMENTS

I thank M. Green, N. S. Mueller, and C. Verlinden for comments on the manuscript.

LITERATURE CITED

ANONYMOUS. 1980. North American bird banding techniques. Vol. 2, Pt. 6. Can. Wildl. Serv. and U.S. Fish and Wildl. Serv.

BUCKHOLTZ, P. G., M. H. EDWARDS, B. G. ONG, AND R. D. WEIR. 1984. Differences by age and sex in the size of Saw-whet Owls. J. Field Ornithol. 55:204–213.

EARHART, C. M., AND N. K. JOHNSON. 1970. Size dimorphism and food habits of North American owls. Condor 72:251–264.

GREEN, G. H. 1980. Decrease in wing length of skins of Ringed Plover and Dunlin. Ringing & Migr. 3:27–28.

GREENWOOD, J. G. 1979. Post-mortem shrinkage of Dunlin *Calidris alpina* skins. Bull. Br. Ornithol. Club 99:143–145.

HENNY, C. J., R. A. OLSON, AND T. L. FLEMING. 1985. Breeding chronology, molt and measurements of accipiter hawks in northeastern Oregon. J. Field Ornithol. 56:97–212.

KNOX, A. 1980. Post-mortem changes in wing-lengths and wing-formulae. Ringing & Migr. 3:29–31.

KORPIMÄKI, E., AND H. HONGELL. 1986. Partial migration as an adaptation to nest-site scarcity and vole cycles in Tengmalms Owl *Aegelius funereus*. Vår Fågelv. Suppl. 11: 85–92.

MACDONALD, P. D. M., AND T. J. PITCHER. 1979. Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. J. Fish. Res. Board Can. 36:987–1001.

MEWALDT, L. R., AND J. R. KING. 1986. Estimation of sex ratio from wing-length in birds when sexes differ in size but not in coloration. J. Field Ornithol. 57:155–167.

MUELLER, H. C. 1986. The evolution of reversed sexual dimorphism (RSD) in owls: an empirical analysis of possible selective factors. Wilson Bull. 98:387–406.

———, AND D. D. BERGER. 1967. Observations on migrating Saw-whet Owls. Bird-Banding 38:120–125.

———, ———, AND G. ALLEZ. 1981. Age, sex and seasonal differences in size of Cooper's Hawks. J. Field Ornithol. 52:112–126.

SHEPPARD, J. M., AND M. K. KLIMKIEWICZ. 1976. An update to Wood's bird bander's guide. North Am. Bird Bander 1:25–27.

WEIR, R. D., F. COOKE, M. H. EDWARDS, AND R. B. STEWART. 1980. Fall migration of Saw-whet Owls at Prince Edward Point, Ontario. Wilson Bull. 92:475–488.

Received 16 Jun. 1989; accepted 9 Dec. 1989.