Herpetologica, 53(3), 1997, 312–320

VALIDATION TESTS OF AN AMPHIBIAN CALL COUNT SURVEY TECHNIQUE IN ONTARIO, CANADA

LEONARD J. SHIROSE,[1] CHRISTINE A. BISHOP,[1] DAVID M. GREEN,[2] CAMERON J. MACDONALD,[3] RONALD J. BROOKS,[3] AND NATALIE J. HELFERTY[4]

[1]*Environment Canada, Canadian Wildlife Service, Canada Centre for Inland Waters, Box 5050, Burlington, ON, L7R 4A6 Canada*
[2]*Redpath Museum, McGill University, 859 Sherbrooke St. W., Montreal, Quebec, H3A 2K6, Canada*
[3]*Department of Zoology, University of Guelph, Guelph, ON, N1G 2W1, Canada*
[4]*Long Point Bird Observatory, Box 160, Port Rowan, ON, N0E 1M0, Canada*

ABSTRACT: Amphibian call count surveys in Ontario use a common survey protocol. We evaluated the precision and accuracy of data collected by these methods. Multiple observers monitored the same routes to determine inter-observer variation. Inter-observer agreement in evaluation of calling intensity and occurrence was generally high, and the experience of the observer affected estimation of both calling intensity and the number of individuals calling. There was no significant increase in inter-observer agreement in evaluation of calling intensity or occurrence with 5-min versus 3-min surveys. Longer surveys rarely resulted in the identification of species not heard in the first 3 min. Surveys of 30–60 min duration indicated that return for effort declined sharply during the second minute of the survey, although new species were sometimes heard after as long as 15 min. Simultaneous intensive and extensive monitoring was performed on populations of Fowler's toads (*Bufo fowleri*) in southern Ontario and bullfrogs (*Rana catesbeiana*) in central Ontario to determine whether there was a relationship between calling intensity and chorus size or population size.

We found a linear relationship between call counts and chorus size in both species. We conclude that the data generated with the call count surveys were adequately precise, but accuracy remains to be validated. Further investigation of the relationships among population size, chorus size, calling intensity, and environmental factors are required to determine whether extensive monitoring of calling amphibians can be used as an index of actual population size. However, until that time, we recommend that calling amphibian surveys be used as the first part of an integrated program using extensive monitoring to identify areas where populations may be declining, and intensive monitoring to confirm or refute reports of declines.

Key words: Validation; Amphibian call count surveys; Fowler's toad; *Bufo fowleri*; Bullfrog; *Rana catesbeiana*; Extensive monitoring

EXTENSIVE monitoring programs using volunteer observers in conjunction with professional biologists are currently being used in Canada and the United States in an attempt to detect fluctuations and trends in both the abundance and distribution of amphibian populations using standardized techniques. Amphibian call count surveys have been implemented in Quebec, Ontario, Manitoba, Saskatchewan, New York, Pennsylvania, Ohio, Michigan, Indiana, Wisconsin, Illinois, Minnesota, Missouri, and Iowa. The amphibian call count protocol used in Ontario is a point-count method that relies primarily upon the efforts of volunteer observers. Three-minute surveys are conducted along routes consisting of up to 10 permanently marked stations. Because of the difficulty of estimating distance in the dark, no attempt is made to limit the calls counted to a fixed distance around each survey point. Stations are set at intervals of at least 500 m to minimize the risk of counting individuals more than once. Surveys begin 0.5 h after sunset and are completed no later than midnight. Each route is surveyed three times during the spring and early summer, with the dates of surveys depending on the latitude of the route.

Ontario has been divided into north, central, and south regions demarcated at the 43rd and 47th degrees of latitude. The three survey periods are 15 days in duration and are set a minimum of 15 days apart. Participants are asked to survey during specific weather conditions. Minimum air temperatures for surveying are 8 C (first survey), 13 C (second survey), and 21 C (third survey). At each station, participants record observer name, date of survey, survey number, time, and weather conditions including air and water temperatures, precipitation type, cloud cover, and wind speed rated using the Beaufort Wind Scale (World Meteorological Organization, 1970). For more details about the development of this protocol, see Bishop et al. (in press).

Observers estimate the relative abundance of calling species by rating the density of individual calls as 0 = no calls heard, 1 = Individual calls not overlapping, 2 = some overlapping calls, but the number of individuals calling can be reliably estimated, 3 = a continuous chorus of calls in which individual calls could not be discerned. If calling intensity is code 1, the number of individuals calling is counted; if code 2, the number is estimated (Bishop et al., in press).

The value of data gathered through call counts depends upon their precision (repeatability of results) and accuracy (the degree to which the data represent true values; Freedman and Shackell, 1992). In the present study, we deal with an index rather than an estimate of abundance, and use the term "accuracy" to describe the degree to which the index tracks differences in abundance, whereas precision refers to the amount of sampling error measured with multiple, simultaneous samples. Low precision would indicate that larger numbers of survey routes would be required to detect trends in relative abundance. The purpose of this study is to assess the precision and accuracy of several facets of an amphibian call count protocol used in the Marsh Monitoring Program (Long Point Bird Observatory, 1995) and Road Call Count Surveys (Bishop et al., in press) and adopted for use internationally by the North American Amphibian Monitoring Program (NAAMP) (Bishop et al., 1996).

Information on the presence or absence of species can provide useful information on changes in geographic and spatial occurrence only if there is a reasonable expectation that the duration of each point count is sufficient to identify all species that are present and likely to call on that night, and if volunteers have adequate training and experience to identify correctly all species encountered. Previous studies suggest that novice observers provide reliable data on species' presence, but there is much variability in assessment of calling intensity (Bishop et al., in press; Hemesath, in press). Call counts can be used as an index of population size only if there is an acceptably low level of observer bias, a statistical relationship between the estimated number of individuals calling and the actual size of choruses, and if chorus size is statistically related to population size.

This study was done in three phases. (1) Extended surveys of ≥30 min were used to quantify the relationship between survey length and the proportion of species heard. The decline in return for effort over time was used to test whether the minimum survey duration falls within the 3 min prescribed by the monitoring protocol. (2) Groups of volunteers surveyed simultaneously and also monitored stations for an extra 2 min over the prescribed 3-min period required for regular surveys. Inter-observer agreement was measured to test the hypothesis that observer experience is a source of bias. Differences in inter-observer agreement between 3-min and 5-min surveys were used to test whether extending the duration of surveys increases the precision of estimates of calling intensity. (3) Simultaneous intensive and extensive monitoring was performed on populations of Fowler's toads (*Bufo fowleri*) in southern Ontario and bullfrogs (*Rana catesbeiana*) in central Ontario to test the hypothesis that calling intensity can be an index of population size. We used regression analysis to test for linear relationships between the number of calling individuals counted and the number of males captured in intensive sampling of choruses. Testing for a statistical relationship between the number of calling individuals and the actual population size was beyond the scope of this paper, because we had data from only one population of each species.

Materials and Methods

We performed 22 extended surveys to quantify the relationship between survey length and the number of species counted. The sampling protocol was similar to the amphibian call count survey protocol used by volunteer observers, except that we recorded the time when each species was heard for the first time, point counts were extended to 30–60 min, and only expert observers were used. To assess inter-observer variation and the effect of previous experience on results, observers were asked to complete questionnaires to describe their previous experience conducting call counts and familiarity with species calls. Volunteers were then categorized as novice, intermediate, or expert. Persons without previous experience in calling amphibian surveys or a firm knowledge of the calls of all species in Ontario were classed as novices. Persons familiar with the calls of all anurans in Ontario and with previous experience in performing calling amphibian surveys were classed as experts

We required a model in which return for effort is initially high but declines as the proportion of species heard approaches 100% of the species present and calling. Therefore, we used a piecewise linear regression model (Wilkinson, 1990) in which the function relating survey length to the cumulative proportion of species heard has two distinct parts, separated by the breakpoint, MIN. The first part of the function has a non-zero slope and is described by the equation

$$Y = B \times X^2. \quad [1]$$

The second part of the function has a slope of zero and is described by the equation

$$Y = B \times MIN^2 \quad [2]$$

where Y is the cumulative proportion of species heard, X is the length of the survey in min, B is a slope parameter for the curve, and MIN is the breakpoint. We used an iterating, non-linear statistical procedure (Wilkinson, 1990) to fit the curve to the data so that MIN is the time at which return for effort drops off.

To determine if extending the duration of surveys increased the precision of estimates of calling intensity, multiple observers with various levels of experience monitored the same routes on a total of nine routes, three of which were surveyed twice. Groups of 2–11 volunteers recorded their observations separately, but simultaneously, and followed the standard amphibian call count protocol, except that stations were surveyed for an extra 2 min over the usual 3-min period required for regular surveys. Multiple-comparison surveys yielded 403 cases involving expert versus expert comparisons, 1973 cases involving expert versus intermediate comparisons, and 1014 cases involving expert versus novice comparisons. Validation surveys were conducted on three routes in the south region and six routes in the central region. No validation surveys were conducted in the north region. Of the 12 validation surveys, three were conducted during the first survey period, seven during the second survey period, and two during the third survey period.

We used inter-observer agreement measured with Cohen's kappa (Cohen, 1960) as an indicator of repeatability, and therefore of precision. Confidence limits (95%) were constructed to test for differences ($P < 0.05$). Each unique combination of route, station, and species for each pair of observers comprised a single case. Using inter-observer agreement between experts as a standard, agreement between expert and intermediate and between expert and novice was used to determine if there was a biasing effect of observer experience on inter-observer agreement. We used chi-square to test for independence of calling intensity code assigned and experience of the observer for all cases taken together, and separately for cases where the code was not zero. We used paired samples t-tests to test for an effect of observer experience on the number of individuals counted in cases where the code assigned was 1 or 2 (counts were only attempted for these codes).

To investigate the relationship between relative abundance evaluated with auditory surveys and actual numbers present, we coupled amphibian call count surveys with

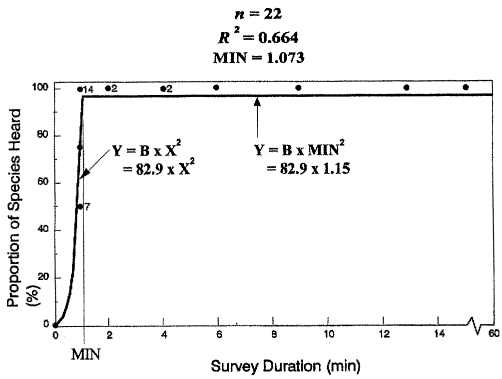

FIG. 1.—The relationship between the proportion of species heard calling and survey length in amphibian call count surveys of 30–60 min. Where data points overlap, the number of points present is indicated to the right.

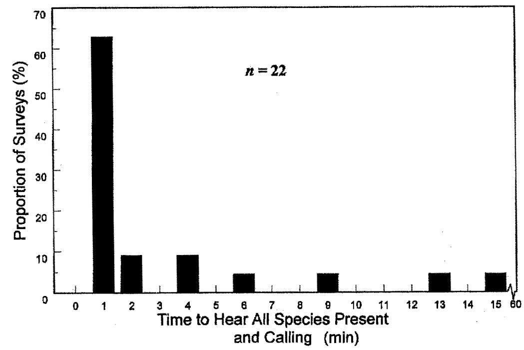

FIG. 2.—Time necessary to hear all species present and calling within 30–60 min amphibian call count surveys.

intensive monitoring (mark-recapture) for populations of Fowler's toads (*Bufo fowleri*) at Long Point in southern Ontario and bullfrogs (*Rana catesbeiana*) at Algonquin Provincial Park in central Ontario. Detailed descriptions of the study sites and mark/recapture sampling protocols for these intensive, long-term studies can be found elsewhere (Fowler's toads: Green, in press; Laurin and Green, 1990; bullfrogs: Shirose et al., 1993). On nights when sites were sampled, auditory surveys were completed first, followed by intensive and exhaustive mark-recapture sampling of choruses.

We used regression analysis to test for linear relationships between the number of calling individuals counted and the number of males captured in intensive sampling of choruses. Counts were summed over entire routes for analysis, because the route rather than the station is the experimental unit in analysis of trends in call count surveys.

RESULTS

Parameters estimated by fitting the piecewise linear regression model to data from extended surveys were $r^2 = 0.664$, B $\pm$ SE = 82.95 $\pm$ 3.094, MIN $\pm$ SE = 1.073 $\pm$ 0.02. The number of new species counted declined early in the second minute of the survey (Fig. 1). In the majority of cases, all species identified at a site were heard within the first minute of the survey. However, in 18.2% of surveys, species were heard for the first time after the first 5 min of a survey (Fig. 2). In 13.6% of surveys, a species was not heard within 60 min, even though it was calling at that site several nights before and after.

The degree of inter-observer agreement was generally high. For 3-min surveys, experts agreed with other experts in 96.30% of cases, with intermediates in 93.36% of cases, and with novices in 93.69% of cases. Values of kappa >0.75 indicate strong agreement, values between 0.40 and 0.75 indicate fair to good, and values <0.4 indicate poor agreement (Wilkinson, 1990). Values of Cohen's kappa $\pm$ SE for 3-min surveys were 0.84 $\pm$ 0.037 (expert versus expert), 0.74 $\pm$ 0.019 (expert versus intermediate), and 0.73 $\pm$ 0.028 (expert versus novice), suggesting that agreement was good to strong. Agreement on presence (intensity rating 1, 2, or 3) or absence (rating of 0) was 98.01% (expert versus expert), 96.91% (expert versus intermediate), and 96.84% (expert versus novice) for 3-min surveys. When cases in which both observers agreed that a species was not present were excluded to remove the confounding effect of high agreement on absence and data were separated by species, inter-observer agreement on calling intensity varied among species and ranged from 46.7–83.3% (Fig. 3a,b,c). Inter-observer agreement on occurrence also varied among species; the degree of agreement on the presence of *Pseudacris crucifer* was

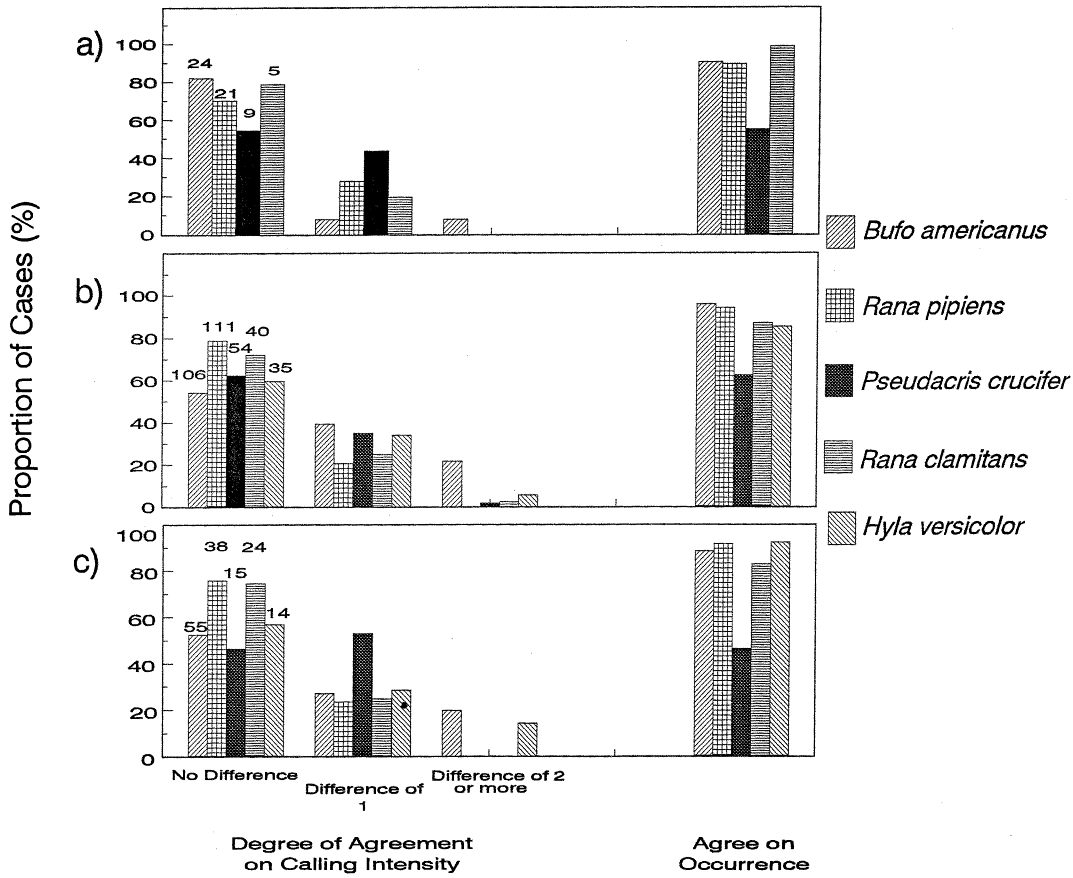

FIG. 3.—Inter-observer agreement (a) among expert observers, (b) between expert and intermediate observers, and (c) between expert and novice observers. From left to right, clusters indicate the proportion of cases in which call intensity levels agreed (0/0, 1/1, 2/2, 3/3), call intensity levels differed by one level (0/1, 1/2, 2/3), call intensity levels differed by two or more levels (0/2, 0/3, 1/3), observers agreed on presence/absence. Numbers above bars indicate the sample size.

lower than it was for the other species tested (Fig. 3a,b,c).

Increasing the survey length from 3 min to 5 min did not produce a significant increase in inter-observer agreement in assessment of calling intensity for any of the three comparisons. Longer surveys rarely resulted in the identification of species not heard in the first 3 min (expert: 0.56% of 897 cases, intermediate: 0.36% of 1105 cases, novice: 0.82% of 728 cases).

We found no significant effect of observer experience on evaluation of calling intensity when all cases were evaluated together. However, the calling intensity assigned by a volunteer was not independent of experience when only non-zero cases were examined ($\chi^2 = 14.2$, df = 4, $P = 0.007$); novice observers tended to underestimate calling intensity and the number of individuals calling, relative to the estimates of experts. Also, the number of calling individuals counted or estimated by novices was lower than that estimated by either intermediate ($t = 1.917$, df = 38, $P < 0.001$) or expert observers ($t = 3.845$, df = 54, $P < 0.001$). There was no significant difference in estimates by expert and intermediate observers.

Intensive and extensive sampling was conducted from 13 May–7 June (Fowler's toads) and from 15 June–12 July (bullfrogs). We found a significant linear relationship between the number of individ-

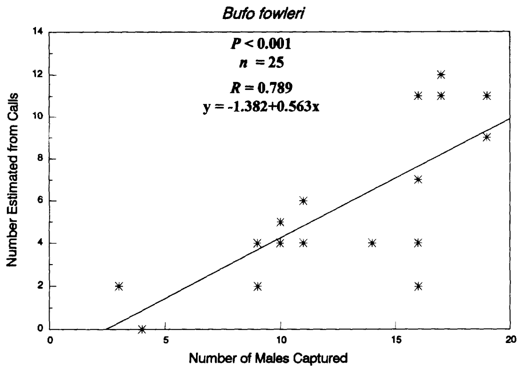

FIG. 4.—The relationship between the number of individuals estimated from amphibian call count surveys and the number captured in intensive sampling of choruses of Fowler's toads (*Bufo fowleri*) at Long Point in southern Ontario.

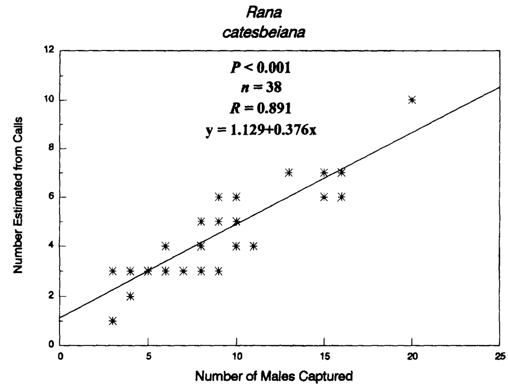

FIG. 5.—The relationship between the number of individuals estimated from amphibian call count surveys and the number captured in intensive sampling of choruses of bullfrogs (*Rana catesbeiana*) in Algonquin Provincial Park in central Ontario.

uals counted or estimated with auditory surveys and the number captured using intensive monitoring techniques for both Fowler's toads ($n = 25$, $P < 0.001$, $r = 0.789$) (Fig. 4) and bullfrogs ($n = 38$, $P < 0.001$, $r = 0.891$) (Fig. 5). The slope of the regression line was <1, indicating that the number captured increased faster than the number heard calling for both species.

DISCUSSION

For maximum efficiency, surveys of calling amphibians that use volunteers should be long enough to allow the identification of all species present and likely to call, but not so long that volunteer observers are unwilling to participate. Our study indicates that return for monitoring effort dropped off early in the second minute of the survey. There was a bias in both the evaluation of calling intensity and the number of individuals counted associated with observer experience. Precision was influenced by which species were calling. For the two species studied intensively, *Bufo fowleri* and *Rana catesbeiana*, the number of individuals estimated to be calling was positively correlated with the number of males captured in choruses using intensive sampling techniques.

The length of point counts in calling amphibian surveys in North America has ranged from 3 min in Quebec, Ontario, Manitoba, and Saskatchewan to 5–10 min in Wisconsin (Huff, 1991) and Iowa (Hemesath, in press). In our area, 3-min point counts appear to be adequate to sample presence/absence and calling intensity of most species. However, a 3-min point count will occasionally fail to detect all species breeding at a site. Increasing the length of point counts will decrease the probability of overlooking species. However, because of the weakness of the relationship between survey length and number of species heard after the initial 2 min of the survey, a relatively large increase in the length of point counts would be required. Even increasing the survey length to 60 min per station does not guarantee that all species present and breeding at a site will be heard, because anuran calling activity is affected by environmental factors, and breeding activity is likely to be interrupted if weather becomes unfavorable. Furthermore, that which constitutes "unfavorable" conditions depends on the species involved. For example, in this study, a hot, dry period in June interrupted the breeding activity of the gray treefrog (*Hyla versicolor*) but did not affect the activity of green frogs (*Rana clamitans*) at the same site. The breeding habits and call characteristics of a species may affect the probability of detection. Bishop et al. (in press) attributed a dearth of records for the northern leopard frog (*Rana pipiens*) to its low calling volume and lack of large, concentrated choruses. Time of night may affect the probability of hearing all species present. All extended (30–60 min) surveys in our study were completed before 0100 h. For some species, such as the mink frog (*Rana septentrionalis*), the nightly peak in calling activity may occur much later in the evening. The length and time of night of surveys should take into account the breeding characteristics of the species that are likely to be present.

The precision of estimates of calling intensity, as measured by inter-observer agreement, was high for 3-min point counts and did not increase significantly when surveys were extended to 5 min. In light of the uncertainty of even a 60-min point count, and the lack of a significant increase in precision with longer surveys, there seems to be little profit in extending point counts beyond 3 min.

Calling amphibian surveys are an extensive monitoring technique and, as such, they necessarily trade certainty for broad geographic coverage. It is not necessary to obtain a 100% probability of detection if a species is present, provided that there is no trend in probability across years. Higher probability of detecting 100% of species increases precision in detecting trends by effectively increasing the sample size (C. Francis, personal communication). However, increasing the length of point counts will not increase sample size if the number of stops per route must be decreased to accommodate the longer counts.

Inter-observer agreement was generally high. When agreement on calling intensity was separated from agreement on presence/absence, we found that cosurveyors were much more uniform in evaluating the latter. The level of inter-observer agreement on calling intensity in this study was similar to that (56–83%) found in study in Iowa (Hemesath, in press).

The inter-observer variation in estimation of calling intensity depends on the species considered. For example, the relatively poor inter-observer agreement on the presence of the spring peeper (*Pseudacris crucifer*) indicates that auditory surveys would be less effective (would require greater sample size to detect trends) for monitoring this species than for the green frog.

The experience of observers should also be taken into account when interpreting data from calling amphibian surveys. In this study, the number of calling individuals counted was lower for novices than for expert or intermediate observers. Novice observers also tended to underestimate calling intensity, relative to the estimates of experts. A novice observer taking over a route from a more experienced observer would be more likely to underestimate relative abundance, which would suggest declines where there were none. In subsequent years, the observer would gain experience and become less likely to underestimate relative abundance. This could mask actual declines. Because novice observers become less likely to underestimate numbers as they gain experience, apparently stable and increasing populations may actually be in decline, whereas apparent declines will likely signify actual declines.

We found a significant linear relationship between call counts and chorus size in both *Bufo fowleri* and *Rana catesbeiana*, which suggests that call counts may be useful as an index of chorus size. However, more research is needed to determine if the relationship is constant across years, which would be necessary if trends in call counts were to be used to track trends in chorus size. It is possible that the relationship between call counts and chorus size is density dependent, with choruses becoming saturated with callers as densities increase and additional males adopting alternative breeding strategies that do not involve calling (Arak, 1983, 1988; Hoglund and Robertson, 1988; Tejedo, 1993).

Testing for a linear relationship between the size of populations, as estimated from mark-recapture data, and either call counts or chorus size in *Bufo fowleri* or *Rana catesbeiana* was beyond the scope of this study. Such a test would require several study sites for each species of interest, or a single site monitored for several years. Extending such a study over several years would have the added advantage of answering outstanding questions of whether or not relationships among variables are constant across years.

Factors such as the level of experience of the observer should be considered along with factors such as ambient and antecedent weather conditions and the species involved in evaluating confidence in reports of declines, and in ranking apparent declines in terms of priority for remedial action. Because chorus size and the intensity of calling activity are likely influenced by ambient and antecedent weather conditions, it is possible that further investigation of the relationships among population size, chorus size, calling intensity, and environmental factors will allow calibration of this method such that call counts may be used as an index of population size. Until that time, the most efficient use of auditory surveys would seem to be using them as the first part of an integrated program using extensive monitoring to identify putative declines in populations, and intensive monitoring to confirm or refute reports of declines.

CONCLUSIONS

Three-minute point counts provide an adequate balance between stop duration and number of stops in areas with species diversity and climate similar to those in Ontario. Point counts as brief as 1 min appear to be adequate, but 3-min counts provide some margin for inexperienced observers and suboptimal monitoring conditions.

Inexperienced observers tend to underestimate the number of calling individuals relative to more experienced observers. It is therefore important to know the level of experience of observers. Because of the nature of the bias, apparently stable or increasing populations may actually be declining, whereas apparently declining populations probably indicate actual declines.

The results of monitoring must be interpreted with consideration of the specific habitat requirements and breeding behavior of the species involved.

There is a positive, linear correlation between the number of individuals estimated from call counts and the actual number of individuals in a chorus. Thus call counts may be used as an index of chorus size.

Trends in the number of individuals estimated from call counts will reflect trends in actual population size only if there is no trend in the relationship between call counts and chorus size from year to year and there is a significant relationship between chorus size and population size. Further research is necessary to determine if these conditions are met. This will likely require a research program including several study sites for each species of interest and extending over several years.

Acknowledgments.—We are grateful to the 117 volunteers who participated in this study. We thank M. Krawchuk, R. McDougall, G. Nancekivell, J. Rabinowitz, N. Schlecht, A. Stagni, and M. Stone for their efforts and expertise in the intensive monitoring portion of this study. This research was supported by Environment Canada and by grants from the Ontario Ministry of Natural Resources and Natural Sciences and Engineering Research Council of Canada (to R. J. Brooks and D. M. Green) and the World Wildlife Fund of Canada (to D. M. Green).

LITERATURE CITED

ARAK, A. 1983. Male-male competition and mate choice in anuran amphibians. Pp. 181–210. *In* P. P. G. Bateson (Ed.), Mate Choice. Cambridge University Press, Cambridge, U.K.

———. 1988. Callers and satellites in the natterjack toad: Evolutionary stable decision rules. Anim. Behav. 36:416–432.

BISHOP, C. A., R. T. BROOKS, S. FACCIO, M. LANNOO, L. MERTZ, L. MONTI, P. RODGERS, AND M. SREDL. 1996. Protocols and Strategies for Monitoring North American Amphibians. Proceedings from the 1995 North American Amphibian Monitoring Program Conference. [http://www.im.nbs.gov//amphib/naampcall.html]

BISHOP, C. A., K. E. PETTIT, M. E. GARTSHORE, AND D. A. MCLEOD. Monitoring anuran populations using call counts and road transects. In press. *In* D. M. Green (Ed.), Amphibians in Decline: Canadian Studies of a Global Problem. Herpetological Conservation Vol. 1, Society for the Study of Amphibians and Reptiles, Saint Louis, Missouri

COHEN, J. 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Measmt. 20:37–46.

FREEDMAN, B., AND N. L. SHACKELL. 1992. Amphibians in the context of a national environmental monitoring program. Pp. 101–104. *In* C. A. Bishop and K. E. Pettit (Eds.), Declines in Canadian Amphibian Populations: Designing a National Monitoring Strategy. Occasional Paper, Canadian Wildlife Service, Ottawa, Canada.

GREEN, D. M. Temporal variation in abundance and age structure in Fowler's toads, *Bufo fowleri*, at Long Point, Ontario. In press. *In* D. M. Green

(Ed.), Amphibians in Decline: Canadian Studies of a Global Problem. Herpetological Conservation Vol. 1, Society for the Study of Amphibians and Reptiles, Saint Louis, Missouri.

HEMESATH, L. M. Iowa's frog and toad survey, 1991–1994. In press. *In* M. J. Lannoo (Ed.), The Status and Conservation of Midwestern Amphibians. University of Indiana Press, Bloomington, Indiana.

HOGLUND, J., AND J. G. M. ROBERTSON. 1988. Chorusing behaviour, a density-dependent alternative mating strategy in male common toads (*Bufo bufo*). Ethology 79:324–332.

HUFF, J. 1991. Frog and toad survey. Pp. 114–123. *In* Wisconsin Department of Natural Resources Technical Report, Madison, Wisconsin.

LAURIN, G., AND D. M. GREEN. 1990. Spring emergence and male breeding behaviour of Fowler's toads (*Bufo woodhousei fowleri*) at Long Point, Ontario. Can. Field-Nat. 104:429–434.

LONG POINT BIRD OBSERVATORY. 1995. Training Kit and Instructions for Surveying Marsh Birds, Amphibians and Their Habitats. Long Point Bird Observatory, Port Rowan, Ontario, Canada.

SHIROSE, L. J., R. J. BROOKS, J. R. BARTA, AND S. S. DESSER. 1993. Intersexual differences in growth, mortality and size at maturity in bullfrogs in central Ontario. Can. J. Zool. 71:2363–2369.

TEJEDO, M. 1993. Do male natterjack toads join larger breeding choruses to increase mating success. Copeia 1993:75–80.

WILKINSON, L. 1990. SYSTAT: The System for Statistics. SYSTAT Inc., Evanston, Illinois.

WORLD METEOROLOGICAL ORGANIZATION. 1970. The Beaufort Scale of wind force. WMO Commission for Maritime Meteorology, Marine Sciences Affairs Report 3:1–22.

Accepted: 11 November 1996
Associate Editor: Richard Howard