Evaluation of methods and data quality from a volunteer-based amphibian call survey

Kristen S. Genet and Lori G. Sargent

Abstract The Michigan Frog and Toad Survey (MFTS) is an annual volunteer-based anuran census. One major concern with data collected by volunteers is the information's quality and consistency. The goal of this study was to evaluate the effect of observer experience on data quality. Questionnaires and an audio CD with a simulated anuran survey route were mailed to all active volunteers. We were able to determine volunteer demographics and commitment to program; species characteristically missed, misidentified, over- or underestimated; and influence of volunteer background on data quality. Volunteers were reasonably reliable in their abilities to determine species presence, but there was extensive variability in abundance estimation. Some species were characteristically confused by volunteers, and additional species frequently were recorded even when absent from a site. Prior experience and background had little influence on the ability to identify or estimate abundance of calling anurans. Our results indicate that such survey approaches are easy for volunteers to learn and provide reasonable estimates for species' presence, but do not necessarily estimate abundance well. These results will be used to improve data-collection protocols for the MFTS and better analyze and interpret data collected, and also could be beneficial for other regional amphibian monitoring programs.

Key words amphibian monitoring, anuran, call survey, data assessment, Michigan, observer evaluation, volunteer-based census

In the last decade, there has been increased interest, debate, and research concerning the apparent global declines of amphibian populations, although the causes have yet to be unambiguously determined (Blaustein et al. 1994, Sarkar 1996, Green 1997, Alford and Richards 1999). The primary obstacle that must be overcome in evaluating potential amphibian declines is separating effects due to anthropogenic influences from natural population fluctuations (Pechmann and Wilbur 1994, Travis 1994). Long-term data from extensive areas are needed (Blaustein et al. 1994), and regional amphibian monitoring programs can contribute such information.

Surveys of calling anurans (hereafter call surveys) are being used for amphibian monitoring in many states and Canadian provinces (e.g., Huff 1991, Bishop et al. 1997, Lepage et al. 1997, Mossman et al. 1998) and have the potential to provide valuable information about population size and status (Zimmerman 1994, Shirose et al. 1997, Driscoll 1998). The North American Amphibian Monitoring Program (NAAMP) has developed a unified protocol for volunteer-based call surveys that

Address for Kristen S. Genet: Department of Zoology, Michigan State University, East Lansing, MI 48824, USA; e-mail: ksgenet@msu.edu. Address for Lori G. Sargent: Department of Natural Resources, Wildlife Division, P.O. Box 30180, Lansing MI 48909, USA.

has been implemented in 24 states (Weir and Mossman, in press). This technique provides a fairly efficient and effective method of surveying anurans over large geographical regions (Scott and Woodward 1994), and has successfully allowed collection of data on presence and abundance of anurans in many areas. Call-survey protocols are easy for volunteers to learn and provide an excellent means for promoting education about wetlands and amphibian conservation.

Annual monitoring data for Michigan's 13 species of frogs and toads can be used to evaluate short-term trends in population dynamics and guide research and conservation efforts for these species. The Michigan Frog and Toad Survey (MFTS) protocol originally was modified from that developed for the Wisconsin Department of Natural Resources (WDNR) frog survey (Mossman et al. 1998). The first surveys in Michigan were carried out in 1988 on a limited basis but were discontinued a few years later due to lack of personnel to coordinate the survey. In 1996 the Michigan Department of Natural Resources (MDNR) initiated a statewide annual frog and toad survey using a network of volunteer observers throughout the state to monitor breeding anuran populations (Sargent 2000). NAAMP developed survey protocols in 1995, but these were not approved for national use until 2000 (Weir and Mossman, in press). Since routes already had been established in Michigan and data had been submitted for 5 years by the time the NAAMP protocols were officially available, the MFTS continued to use their original protocols. However, some MFTS routes adhere to NAAMP protocols, and these data are submitted for use at both the state and national levels. In the MFTS, volunteers establish routes by first submitting a map and descriptions of a series of 12 potential survey sites identified without knowledge of wetlands or presence of amphibians (i.e., sites identified in winter or otherwise outside the breeding season when anurans are inactive); upon approval of the state coordinator (L.G.S.), the route is driven and sites are established (Sargent 2000). NAAMP routes use randomly generated driving maps (Weir and Mossman, in press). The volunteers then survey their routes 3 times each spring, corresponding to the peak breeding times of anurans, and record the identity of each species and an index of their calling intensity (0=absent, 1=few individuals with nonoverlapping calls, 2=many individuals with overlapping but distinguishable calls, 3=full chorus with individual calls indistinguishable). MFTS protocols instruct volunteers to listen for 5 minutes at each site (Sargent 2000).

Blanchard's cricket frog (*Acris crepitans blanchardi*) was once common in the Great Lakes basin, but many populations in the Midwestern United States have suffered dramatic declines over the last few decades. Blanchard's cricket frogs reach the edge of their range in southern Michigan, and are classified a species of special concern.

Volunteer training is an important component that contributes to the success of a regional monitoring program. Prior to onset of data collection in 1996 and again in 2001, training workshops were held throughout the state. These workshops provided information on general anuran biology and habitat requirements, how to distinguish species by sight and sound, and instructions for establishing a route and conducting the survey (Sargent 2000). In addition to these initial workshops, instructional packets including information on protocols and a training cassette with examples of all species' calls were distributed to interested participants.

Several factors (e.g., prior experience or hearing ability) may influence observers' abilities to correctly identify anurans and estimate their abundance. Previous studies suggested that novice observers provide reliable data on species' presence, but their assignment of call index values was more variable (Bishop et al. 1997, Shirose et al. 1997, Hemesath 1998). Differences among observers could influence survey data quality and should be incorporated into analyses of population changes (Sauer et al. 1994).

Because data for the MFTS are collected each year by hundreds of volunteers with varying expertise (range of 198-293 routes submitting data during 7 years of program; most routes had at least 2 volunteers), we initiated a study to identify factors that influenced data collection so we could incorporate into data analyses modifications to deal with these factors. Although problems of observer bias and subsequent data analyses have been investigated for monitoring programs such as the North American Breeding Bird Survey (BBS, Sauer et al. 1994, Kendall et al. 1996, Link and Sauer 1996), differences among volunteer observers and their implications for data analyses need to be more thoroughly addressed for amphibian monitoring programs (but see Bishop et al. 1997, Shirose et al. 1997, Hemesath 1998). Methods for assessing observer experience have been proposed (Bishop et al. 1997, Shirose et al. 1997, Hemesath 1998) but have not been widely implemented.

The overall goal of this study was to evaluate observer experience and data accuracy and precision for the MFTS. Based upon this goal, the primary objectives of this study were to evaluate how volunteer background and experience influenced anuran identification and assignment of call index values, determine implications of observer differences on analyses of population trends, and establish an observer evaluation process that could be implemented by other states with similar monitoring programs.

Methods

We administered a mail survey between July and December of 2001. The survey was thoroughly evaluated for ethical appropriateness by the Michigan State University Committee on Research Involving Human Subjects (UCRIHS IRB# 01-324). In mid-July we sent a questionnaire and audio CD to all MFTS volunteers who had submitted data in 1999, 2000, or both years (n=355). We sent a second mailing of reminder or thank-you postcards in early September, and we also provided additional questionnaires or CDs to volunteers who requested them at that time. Finally, we mailed the annual MFTS update in early December, which included a preliminary report of the survey data and encouragement to send in responses if volunteers had not already done so. We did not accept any responses after 31 December 2001. We removed from the sample surveys returned as undeliverable by the United States Postal Service or cases where the recipient was deceased. We included in the sample size surveys returned blank or indicating no interest in participating in the survey. We considered these "nonresponses" (i.e., counted in final corrected sample size, but data omitted from analyses). Respondents provided on their questionnaires information about basic demographics, participation, and experience with the MFTS and wildlife in general, and commitment to this and other wildlife monitoring programs (specific questions from the questionnaire are found in the Appendix).

We also enclosed with the survey an audio CD with 12 tracks. Each track represented a wetland breeding site typical of those visited by volunteers as they collect data for the MFTS. We instructed volunteers to listen and, using data forms identical to those used for annual surveys, record each species and assign an abundance index for each CD track in the same manner as for sites they monitor annually. Each track was 5 min in length, the recommended listening time for MFTS protocol (Sargent 2000).

If respondents failed to follow survey instructions or submitted only a partial questionnaire or datasheet, we edited the data as follows. If the respondent gave a range when asked for a numeric response, we assigned the arithmetic mean. If respondents indicated they would submit data as long as possible in response to question 3 (Appendix), we assigned a value of 10 years. We omitted from analyses nonresponses on individual questions. We categorized observers as novice (n=18), intermediate (n=46), or expert (n=90) based on number of years they had submitted data to the MFTS (1-6 yr) and their perceived level of expertise (5-point rating scale, range 2-5). We summed these 2 values (summed score range: 3-11) and divided the resulting scores equally into 3 experience categories: novice=3-5, intermediate=6-8, expert=9-11. The call index values assigned by

The northern leopard frog (*Rana pipiens*) was once a common species throughout its range. Leopard frogs have experienced population declines in the Midwestern United States and have been the impetus for the establishment of many regional amphibian monitoring programs using volunteers.

respondents were compared to values independently predetermined by experts (both authors and 2 additional experts in the field) for each species on each of the 12 tracks. We performed statistical analyses using SAS statistical software (SAS Institute, Cary, NC) and methods described in the following paragraphs (Zar 1998).

We used chi-square analyses to test whether species identifications deviated from expected (i.e., all volunteers correctly identified all species on each track) and to determine whether respondents assigned the expected call index values (predetermined by authors and 2 other experts) for each species on each track. Volunteers in other regional call surveys have had difficulty discriminating between the 2 highest call index values (L. B. Johnson, University of Minnesota-Duluth, personal communication). Thus, we tested 2 separate hypotheses for each track of the CD: 1) respondents correctly assigned 4-category call index values (0, 1, 2, 3), and 2) respondents correctly assigned 3-category call index values (0, 1, 2+3 combined). For the 3 species with restricted distributions within Michigan (Fowler's toads, *Bufo fowleri;* mink frogs, *Rana septentrionalis;* and Blanchard's cricket frogs, *Acris crepitans blanchardi*), we used contingency table analyses to determine whether there was a significant association between living within a species' range and correct identification. We also used contingency-table analyses to determine whether observer experience level (i.e., novice, intermediate, or expert) influenced abundance estimation of each species (i.e., correct, over-, or underestimation of abundance, based on authors' predetermined values). We used parametric and nonparametric (Kruskal-Wallis) analysis of variance (ANOVA) to determine whether observer experience influenced correct identification of species. Nonparametric tests were implemented in cases where data did not meet assumptions of parametric tests (i.e., small sample size). Data are reported as means±SE unless otherwise noted, and a *P*-value <0.05 was considered statistically significant.

Results

Volunteer demographics

Of the 355 questionnaires sent to volunteers, 4 were undeliverable, 1 recipient was deceased, and 179 were returned, yielding a 51.1% response rate (corrected sample size, $n=350$). Volunteers indicated extensive experience with wildlife and commitment to this program (Table 1). The typical respondent was just over 50 years old, and 26% of respondents indicated they were retired. Slightly more females (53.7%) than males (46.3%) participated, although the difference was not significant ($\chi^2_1=0.5746, P = 0.464$). Over 70% of respondents also were avid birders, and 44% were involved in at least one additional wildlife monitoring program.

Table 1. Demographic characteristics of survey respondents representing active MFTS volunteers in Michigan, 2001. Means and SE were calculated from all volunteers in the MFTS returning completed questionnaires (sample size for each characteristic is presented in the last column.

Demographic character	Mean	SE	Range	n
Involvement with program to date (yrs.)	4.67	0.13	1–6	175
Total anticipated involvement (yrs.)	10.04	0.32	3–30	142
Wildlife experience (yrs.)	24.01	1.28	1–65	168
Age (yrs.)	50.52	0.86	17–81	177

The typical respondent felt he or she had improved approximately 2 points on a rating scale from 1 (low) to 5 (high) from beginning in the MFTS to their current level of expertise (beginning: 1.66±0.07, current: 3.56±0.05). With respect to training, volunteers predominantly listened to the training tape (97.7%) and attended training workshops (66.5%). The majority of volunteers participated in more than one type of training activity (82.4%), but fewer practiced in the field (11.4%) or participated in some other type of training activity (33.0%). On average, 2 people participated in each survey route, and in >85% of cases, the number and identity of those people were consistent from year to year. Most respondents (80%) indicated that a primary observer had been designated for their route (if >1 observer on that route). The primary observer was responsible for data forms and establishing consensus when there were discrepancies among observers. Thirty-five percent of respondents indicated there had been discrepancies among observers in terms of species heard or call index values assigned. The most common methods for resolving these discrepancies were to discuss among observers until consensus was reached (58.3%), listen longer at the site (51.7%), check calls against the training tape (28.3%), or allow primary observer to make decision (8.3%).

Species identification

Volunteers were able to identify species by their calls (Tables 2 and 3), but correct identifications ranged from 60% for Fowler's toads to >98% for northern green frogs (*Rana clamitans*) and bullfrogs (*R. catesbeiana*). The majority of species present on a track were correctly identified by ≥80% of volunteers (Tables 2 and 3). Many respondents confused northern leopard frogs (*R. pipiens*) and pickerel frogs (*R. palustris*) (>10%, Table 2). Fowler's toads were correctly identified by only 60% of respondents (Table 2), and a large proportion of respondents indicated presence of one (or both) of the gray treefrog species (*Hyla versicolor, H. chrysoscelis*) either instead of or in addition to Fowler's toads. Several respondents also confused calls of mink frogs and wood frogs (*R. sylvatica*) (Table 2). No other species appeared to be characteristically misidentified or confused with a similar-sounding species.

On 6 of the 12 tracks, volunteers' responses significantly differed from correct identifications ($P<0.005$), indicating they had missed or misidentified a significant number of species on those tracks. For the remaining tracks, volunteers correctly identified all species present; however, additional species not calling on the CD were also recorded (Tables 2 and 3). For the 3 species with restricted ranges, living within the species' range did not affect observers' abilities to correctly identify these species (Table 4, $P>0.05$). Novice, intermediate, and expert observers did not differ significantly in their abilities to correctly identify species' presence based on their calls ($H_2 = 1.468$, $P = 0.480$). Respondents in all 3 experience categories recorded correct identifications of all species combined (i.e., all species calling in concert at a breeding site) in >85% of cases. Observer experience level also had no significant influence on respondents' abilities to correctly identify individual species ($P>0.05$).

Abundance estimation

There were discrepancies in assignment of call indices for some species, even among those respondents who had correctly identified the species on each track of the CD recording. In many cases there was consensus among respondents, but in others respondents were relatively equally split among different call index values (Figure 1). For example, spring peepers (*Pseudacris crucifer*) were almost unanimously assigned call index 3 on

Table 2. Proportion of respondents in the MFTS identifying each species on each track of the CD recording of Michigan anurans, 2001. Species present on each track are indicated with an asterisk.

Track	1	2	3	4	5	6	7	8	9	10	11	12
Fowler's toad (*B. fowleri*)	<0.01	0.01	<0.01	0.02	0	0.02	0.60*	0	<0.01	0.03	0.06	0.01
wood frog (*R. sylvatica*)	0.16	0.03	0	<0.01	0.87*	0.04	0.03	0.02	0.07	0.01	0	0.24
west. Chorus frog (*P. triseriata*)	0.76*	0.03	0	0.04	0.64*	0.04	0.13	0.03	0.03	0.10	0.01	0
spring peeper (*P. crucifer*)	1.0*	0.04	<0.01	0.90*	0.99*	0.91*	0.60*	0.03	0.80*	0.93*	0.08	<0.01
north. leopard frog (*R. pipiens*)	0.90*	0	0	<0.01	0.56*	0.85*	0	0	0.87*	0.22	0	<0.01
pickerel frog (*R. palustris*)	0.10	0	0	<0.01	0.13	0.17	<0.01	<0.01	0.11	0.69*	<0.01	0
east. American toad (*B. americanus*)	0.02	0	0.99*	<0.01	<0.01	0.99*	0.04	0.01	0.90*	0	<0.01	0.02
gray treefrog (*H. versicolor*)	0.05	0.15[a]*	0.03	0.93*	0.02	0.89*	0.56	0.13[a]*	0.02	0.02	0.94*	<0.01
Cope's gray treefrog (*H. chrysoscelis*)	<0.01	<0.01	<0.01	0.06	0	0.05	0.32	0.03	0	0	0.87*	<0.01
Blanchard's cricket frog (*A. c. blanchardi*)	0.01	0.85*	0	0	0.05	0.02	0	0.86*	<0.01	0.01	0.02	0.03
mink frog (*R. septentrionalis*)	0	0.04	0	0	0.09	0.04	0	0.06	0.10	0.04	0.01	0.71*
north. green frog (*R. clamitans*)	0.01	0.99*	<0.01	0.99*	0.01	0.03	0	0.98*	0	0.01	0	0.97*
bullfrog (*R. catesbeiana*)	0	0.99*	0	0	0	0.02	<0.01	0.98*	0	0.01	0	0.04

[a] Gray treefrogs may be calling from location outside survey site that cannot be determined from CD recording. This species was omitted from statistical analyses for these tracks.

Table 3. Summary of correct identifications and identification errors for Michigan anurans averaged over all 12 CD tracks from volunteers in the MFTS, 2001.

Species	Correct[a] mean[d]	SE[e]	Missed[b] mean[d]	SE[e]	Incorrect[c] mean[d]	SE[e]
Fowler's toad (*Bufo fowleri*)	0.602		0.398		0.015	0.005
wood frog (*Rana sylvatica*)	0.864		0.136		0.055	0.023
west. chorus frog (*Pseudacris triseriata*)	0.701	0.059	0.299	0.059	0.041	0.013
spring peeper (*Pseudacris crucifer*)	0.874	0.052	0.126	0.052	0.033	0.013
north. leopard frog (*Rana pipiens*)	0.795	0.080	0.205	0.080	0.029	0.027
pickerel frog (*Rana palustris*)	0.688		0.312		0.051	0.019
east. American toad (*Bufo americanus*)	0.957	0.031	0.043	0.031	0.010	0.004
gray treefrog (*Hyla versicolor*)	0.917	0.014	0.083	0.014	0.099	0.077
Cope's gray treefrog (*Hyla chrysoscelis*)	0.873		0.127		0.043	0.028
Blanchard's cricket frog (*Acris crepitans blanchardi*)	0.858	0.006	0.142	0.006	0.014	0.005
mink frog (*Rana septentrionalis*)	0.705		0.295		0.034	0.011
north. green frog (*Rana clamitans melanota*)	0.984	0.004	0.016	0.004	0.009	0.003
bullfrog (*Rana catesbeiana*)	0.981	0.006	0.019	0.006	0.008	0.004
all species combined	0.831	0.034	0.169	0.034	0.034	0.007

[a] Correct identifications indicate volunteers who correctly recorded a species as present when it was calling.

[b] Missed identifications indicate volunteers who recorded a species as absent when it was actually present.

[c] Incorrect identifications indicate volunteers who recorded a species present when it was absent from the recording.

[d] All means represent proportion of total respondents.

[e] Species with no reported SE values were present on only one track, precluding estimation of variability.

track 1 (Figure 1*a*), while Blanchard's cricket frogs and green frogs were almost equally split between call index values 2 and 3 on tracks 2 and 8 (Figures 1*b* and 1*h*). For 11 of the 12 tracks, call index values assigned by respondents differed from those predetermined by the authors ($P < 0.05$). Combining call index values 2 and 3 only improved the respondents' abundance estimation on one track. In most cases the deviation from the expected call index values was due to either not recording a species as present (assigning a call index of 0) or discrepancies among volunteers in abundance estimation when a species was recorded as present (Figures 1*a*–*l*). Track 8 was an exact repeat of track 2, and 75.2% of respondents identified the same species in both tracks. However, of those respondents who identified the same species in tracks 2 and 8, only 43.8% assigned the same call indices to those species in both tracks. In 3 × 3 contingency table analyses of volunteer experience level (novice, intermediate, or expert) and abundance estimation (correct, over-, or underestimation) for each species, only 2 tests produced significant associations (6.7% of comparisons). Novice observers tended to overestimate northern leopard frogs on track 9 while intermediate and expert observers underestimated calling intensity ($\chi^2_4 = 51.0047, P < 0.001$), and correct index assignment was positively associated with observer experience for mink frogs on track 12 ($\chi^2_4 = 10.8867$, $P = 0.0279$).

Discussion

The MFTS was modified from protocols originally developed by the WDNR and NAAMP. NAAMP supplies randomly generated route maps to interested volunteers, and the volunteer then chooses sites based on an equidistant or stratified-by-habitat method (Weir and Mossman, in press). The MFTS instructs volunteers to select the area where they first establish a route, then choose sites without regard to anuran presence (Sargent 2000). While there likely are some volunteers who are knowledgeable about wetland locations and who have paid attention to the frogs calling at these wetlands before involvement in the MFTS, most volunteers have no prior experience with the wetlands or the anurans in an area before establishing a survey route (L.G.S., personal observation). As such, sites are chosen haphazardly, only omitting areas that are "too dangerous or too noisy to hear" calling frogs. The ability to survey a route near an observer's

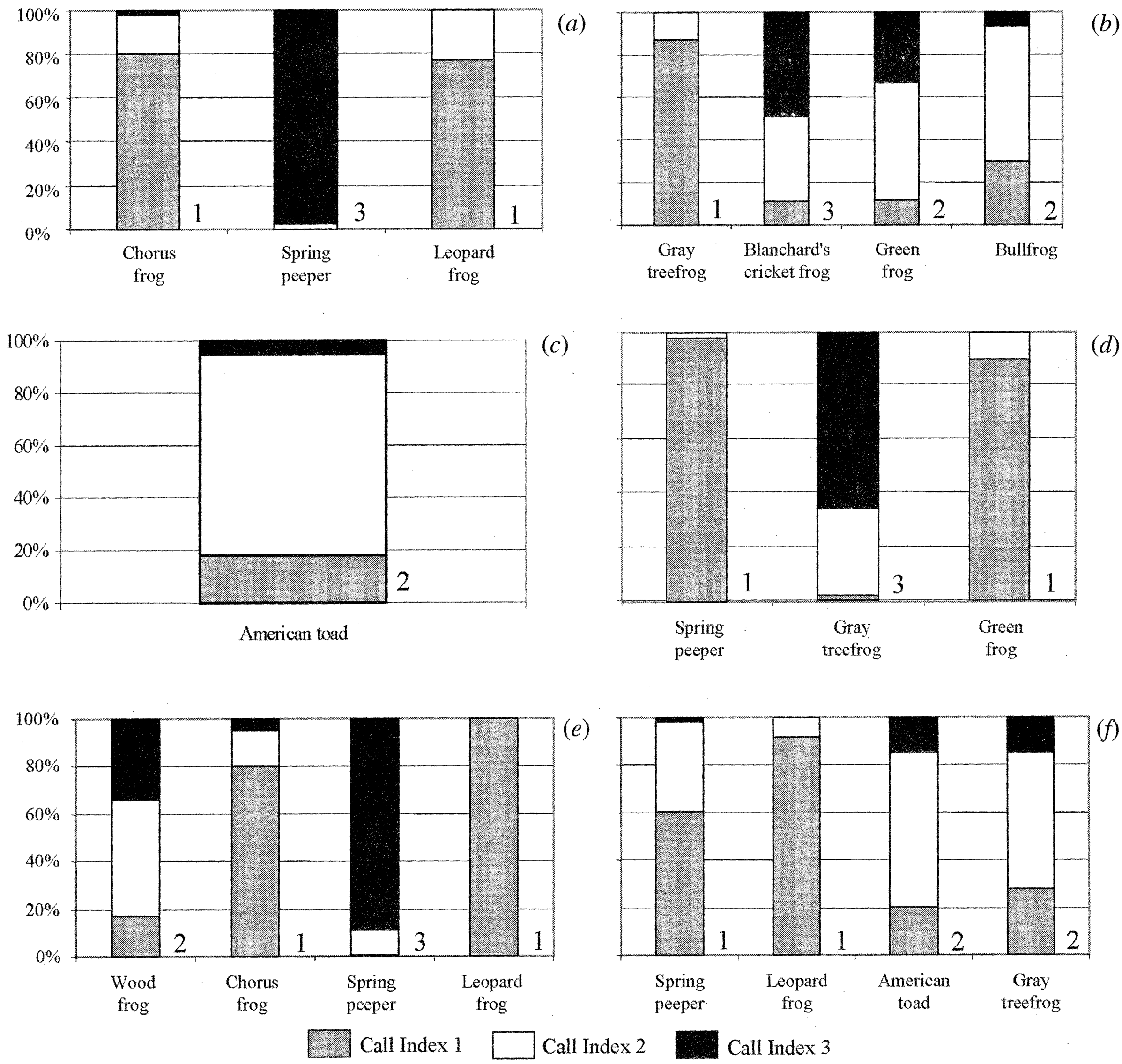

Figure 1 (*a–f*). Call indices assigned by MFTS respondents for each CD track in Michigan, 2001. *Y*-axis (percentage of respondents) is the same for all panels. Correct call index values are given at the base of each bar. (*a*) Track 1, (*b*) Track 2, (*c*) Track 3, (*d*) Track 4, (*e*) Track 5, (*f*) Track 6. (Figure 1 (*g–l*) continued next page.)

home significantly increases the likelihood of data submission and volunteer retention. Furthermore, the longevity and large number of routes established for the MFTS dilute any bias that site selection may inherently introduce. The MFTS, NAAMP, and other large-scale amphibian monitoring programs are intended to provide a meaningful and relatively inexpensive method to track changes in distribution and abundance of species with applicability at a variety of scales. Volunteer observers are an integral part of these goals.

MFTS volunteers were able to correctly identify Michigan's frogs and toads by their distinctive breeding calls relatively consistently. All species were correctly identified by ≥60% of respondents, and all but 4 species were correctly identified by more than 80% of the respondents. Species that were missed or misidentified were most likely due to confusion with other species with similar calls (e.g., northern leopard frogs and pickerel frogs), inability to determine spatial location of call from CD recording and whether that species was present within the wetland site or farther away (e.g., gray treefrogs in tracks 2 and 8, spring peepers in track 7), inability to hear less conspicuous calls masked by more prominent species (e.g., western

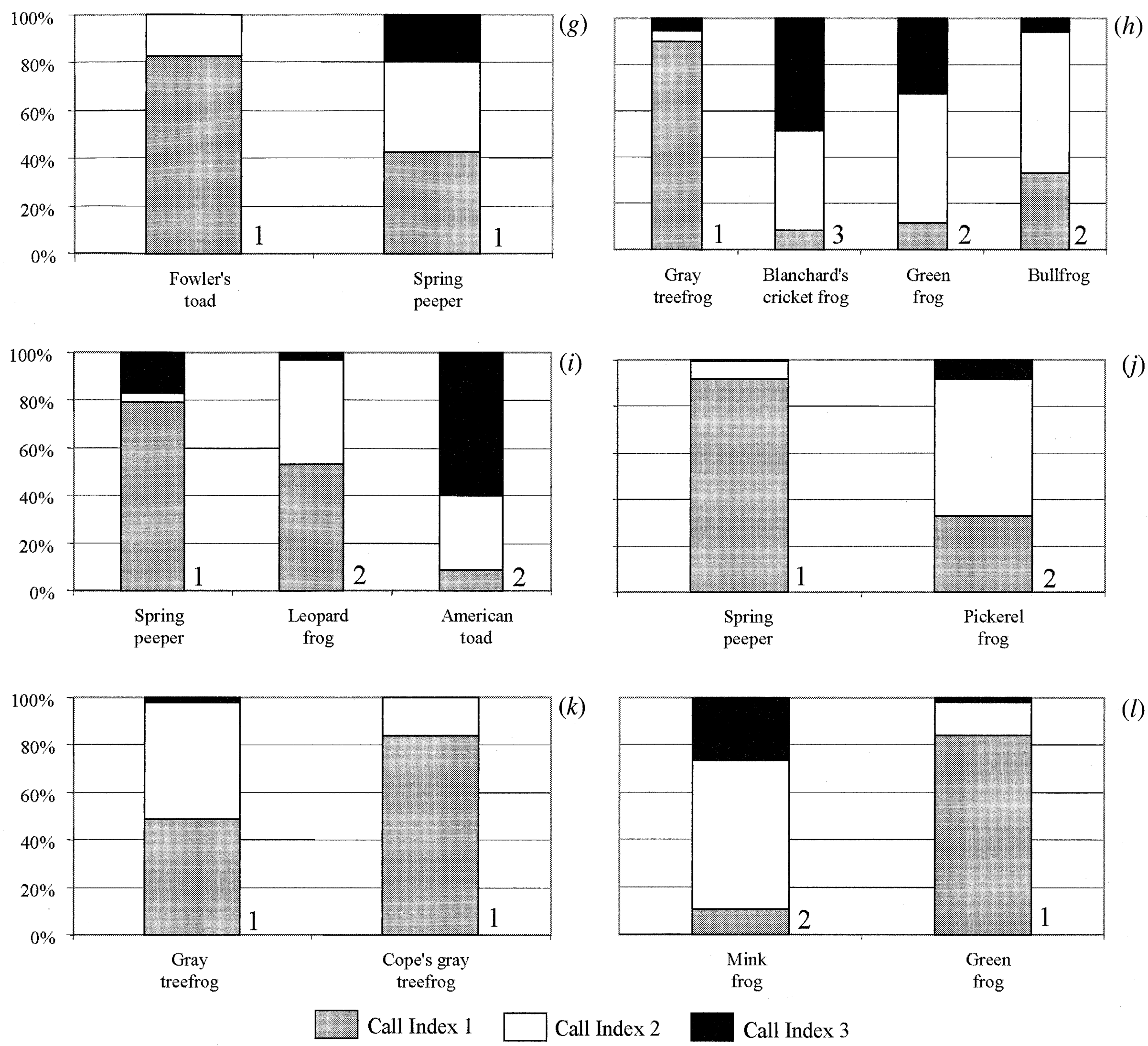

Figure 1 (*g*–*l*). Call indices assigned by MFTS respondents for each CD track in Michigan, 2001. *Y*-axis (percentage of respondents) is the same for all figures. Correct call index values are given at the base of each bar. (*g*) Track 7, (*h*) Track 8, (*i*) Track 9, (*j*) Track 10, (*k*) Track 11, (*l*) Track 12.

chorus frogs [*Pseudacris triseriata*] calling in the presence of a full chorus of spring peepers), or unfamiliarity with species not present in volunteers' survey areas (e.g., mink frogs, Fowler's toads, Blanchard's cricket frogs). In general, we view our results as positive affirmation of the data quality of the MFTS; however, several issues raise concerns that we will address in future analyses and protocol revisions. Additionally, just over half of the active volunteers in this study returned questionnaires, potentially limiting our conclusions and recommendations to the most conscientious observers.

Cases of mistaken species identification in this study probably represent a worst-case scenario. Use of the CD recording has some obvious limiting factors that need to be considered when comparing this study with field investigations. Volunteers in this study were not given any background information about the CD recordings. Had that information been available (i.e., date, geographic location, habitat type), some identification problems may have been avoided. Even so, these results indicate the need to review training materials in order to decrease any possible misidentifications. Many of the species pairs commonly misidentified do not breed at the same time (e.g., mink frogs and wood frogs) or have differences in habitat preferences (e.g., leopard frogs and pickerel frogs). Ensuring

Table 4. Percentage of MFTS respondents living within and outside ranges of the 3 species with limited distributions within Michigan who correctly identified these species, 2001. The relationship between living within a limited range and correct identification was tested using contingency table analyses.

Species	Within range	Outside range	χ^2_1	*P* value
mink frog (*Rana septentrionalis*)	75.0%	69.9%	0.2216	0.638
Fowler's toad (*Bufo fowleri*)	63.8%	59.2%	0.3265	0.568
Blanchard's cricket frog[a] (*Acris crepitans blanchardi*)	89.5%	78.6%	3.5559	0.059
Blanchard's cricket frog[b] (*Acris crepitans blanchardi*)	87.2%	84.3%	0.2711	0.603

[a] Track 2.
[b] Track 8.

that volunteers know the basic biology of each of the species in addition to their distinctive breeding calls could greatly improve the data accuracy. The relatively low proportion of respondents who correctly identified Fowler's toads is also a concern; it appears that many observers are not trained to document potential range expansions or isolated populations. Additionally, the CD was produced from field recordings, and true species diversity and abundance were not positively known for the recordings. The authors and additional experts determined the species present and their call index values, but the responses were compared to the authors' expectations and not true population sizes.

Breeding habits and call characteristics of a species may also affect probability of detection. Some species have a boisterous, unmistakable call (e.g., spring peepers) that is easier to distinguish than a call that is subtler and lower in volume (e.g., northern leopard frogs). Bishop et al. (1997) attributed the paucity of records for breeding leopard frogs in Ontario to its subtle call and lack of concentrated choruses. Our results also indicated that some species may be missed as a result of not being heard over the din of louder species in greater concentrations. Western chorus frogs frequently call at the same time and location as spring peepers, but discerning calls of western chorus frogs within the deafening chorus of spring peepers can be difficult.

Our volunteers spanned a wide range of ages and backgrounds. Their mean age was 50.52 yr, and although some hearing loss is expected with age, the 13 species of Michigan frogs typically call between 300–3,200 Hz (T. O. Matson, Cleveland Museum of Natural History, personal communication), well within the normal range of human acoustical sensitivity. Detection of frog calls also depends on their volume, and 95% of men aged 50.4 (average age of men in this study) can detect 250 Hz at 11.78 dB and 3,000 Hz at 25.22 dB (comparable values for women aged 50.6, average in this study, are 10.73 and 18.76 dB for 250 and 3,000 Hz, respectively) (G. A. Flamme, University of Iowa, personal communication). The volume at which a frog calls and the distance between it and the observer are important factors, as is additional external noise (e.g., vehicle traffic). As average age of the MFTS volunteer was middle-aged, there may be biases related to hearing loss. We will continue to investigate the potential influences of age-related hearing loss on our data quality.

Evaluation of volunteer-collected data for other amphibian monitoring programs has indicated that inter-observer agreement on presence and abundance estimation was generally high, but experience also played a role in data quality (Shirose et al. 1997, Hemesath 1998). Contrary to results from other regional amphibian monitoring programs using call surveys, observer experience appeared to play only a minor role in the MFTS. Perhaps this was a result of the fact that our volunteers had an average of 24 years of wildlife experience as well as more than 4 years of participation in the MFTS (Table 1). Observer experience level was not a significant influence in either the identification or abundance estimation for any of Michigan's anurans. Other investigators have found that inter-observer agreement on species presence was high and not influenced by experience (>96% agreement regardless of experience level, Shirose et al. 1997), but agreement on calling intensity varied with experience level (47–83% in Ontario, Shirose et al. 1997; 56–83% in Iowa, Hemesath 1998). Novice observers tended to underestimate calling intensity relative to experts (Shirose et al. 1997). Similarly, for the North American Breeding Bird Survey (BBS), observer differences were related to experience such that observer quality increased over time (Sauer et al. 1994). First-time BBS observers tended to underestimate species and individuals, and population analyses of BBS data included observers as covariates in order to avoid confounding observer quality with population trends (Erskine 1978, Kendall et al. 1996). Since experience level of the MFTS volunteers did not influence their abundance estimation, observer experience level was not likely to affect analyses of changes in abundance over time.

Although observer experience was not an influential factor, there were dramatic differences among volunteers in assignment of call index values. In many cases, almost the same proportion of survey respondents assigned 2 different call index values. In general, observers assigned a call index value of 1 with reliability and consistency but had difficulty distinguishing between call index values 2 and 3. The subjective interpretation of overlapping calls being distinguishable as individuals was the difference between call index values 2 and 3. The character of individual species' calls also influenced observers' abilities to assess their abundance and assign the correct abundance index. Shirose et al. (1997) also found that inter-observer variation in estimation of calling intensity depended on the species considered. Species with prolonged calls (e.g., American toads, *Bufo americanus*) did not appear to overlap as much as species with shorter calls (e.g., spring peepers). In other states' amphibian monitoring programs, volunteers seemed able to reliably determine the call index of 1 but often had trouble discerning between call indices of 2 and 3 (L. B. Johnson, personal communication). One solution would be to translate data into a 3-level abundance index: absence (0), low abundance (1), and high abundance (2). In addition to inter-observer differences, we also need to consider intra-observer differences in species detected and call index values assigned. With the duplicate tracks present on the CD, we found there was relatively low agreement in call index values assigned by the same volunteer. While the combination of call index values 2 and 3 would improve the agreement in abundance estimation, this is a concern we need to address in our training packet and communications with volunteers.

A related concern is the relationship between calling index and actual population size for each species. As a result of different call characteristics of the 13 species of Michigan anurans (e.g., the long trill of eastern American toads compared to the short call of the spring peeper), the same call index value recorded for different species will translate to very different population sizes. The relationship between call counts and population size has been investigated for some species (Shirose et al. 1997), although until empirical calling index-population size relationships can be developed for all species involved, call index values may be best translated into presence or absence data that can be used to track changes in populations over time (Weir and Mossman, in press).

An inevitable component of a large-scale regional monitoring program is differences among observers. There is a trade-off between the amount and extent of data and its reliability and consistency. While some monitoring programs have reported significant differences among their volunteer observers, in other programs it does not appear to be a major concern (Kline 1998, Mossman et al. 1998). Observer bias was considered minimal in some programs as a result of a combination of volunteers' experience, scientific evaluation of data, and number of observations (C. M. Francis and A. Chabot, Long Point Observatory, unpublished report, Kline 1998, Mossman et al. 1998). For amphibian call surveys in Wisconsin, increasing the number of observations (i.e., the number of routes surveyed) reduced variability in the data set more than increasing volunteer training (Kline 1998). Similarly, power to detect significant population trends in the Marsh Monitoring Project increased as more stations were surveyed (C. M. Francis and A. Chabot, Long Point Observatory, unpublished report). It appears that species that are widespread and call frequently can be adequately monitored with roadside call surveys; species that call infrequently will require more effort (i.e., more routes surveyed) to track meaningful population trends (Crouch and Paton 2002). Although we found some differences among observers in abundance estimation, we have over 400 survey routes statewide (with 10 wetland sites along each route). With such a great number of sites surveyed each year, we should have the power to track significant population trends over time. However, a formal power analysis also should be conducted.

With respect to the MFTS, we have several recommendations. We encourage our volunteers to refresh their skills prior to each survey season. Being familiar with the basic biology, phenology, range, and habitat requirements of Michigan's frogs and toads will help avoid some simple identification mistakes found in this study. Additionally, NAAMP is currently developing an online frog survey quiz for volunteers. When operational, the MFTS will consider volunteers' quiz scores when determining inclusion of their data. We also encourage volunteers to review and adhere to MFTS protocols. There was substantial variability in amount of time spent listening at survey sites, and although most species were heard within the first minute of a survey, and 3-minute stops have been recommended for call surveys (Shirose et al. 1997), we need to ensure standardization of time spent at stops. Concerning analyses of sites occupied by each species and changes in occurrence over time, we need to consider that although presence was a reliable observation by volunteers, absence at a site was not guaranteed by undetected calling males. For analyses of abundance trends, we propose combining call index values 2 and 3. While we lose some resolution by eliminating one abundance category, there is so much variability among observers in assigning those values that any differences between the 2 higher abundance categories are most likely not biologically meaningful. We also plan to continue to require verification in the form of a photo, recording, or expert observation for rare and hard-to-identify species (e.g., Blanchard's cricket frogs and Cope's gray treefrogs). In addition to these 2 species, we also encourage verification for pickerel frogs, and we may make this a requirement in the future. Finally, the training and refresher workshops offered in the past have been successful and well attended. Provided that funding and personnel are available, we plan to continue to offer these periodically.

The observer evaluation process we present here is a valuable tool that could be implemented by other regional anuran monitoring programs. One caveat is that similar studies could provide information on time of year and site characteristics with a CD recording. If we had provided that information, some misidentifications may have been reduced. Even so, our results indicated that data from the MFTS can be used to track trends reliably for most species of frogs and toads. Data quality is a major concern in large-scale long-term monitoring programs, and documenting differences among observers is essential for analysis and interpretation of the data. Detection and abundance estimation of frogs and toads are likely influenced by a variety of factors, and confidence in the data and any resulting trends is greatly improved when the influences of observer bias on the data collection process are understood and documented.

Acknowledgments. T. M. Burton, S. K. Hamilton, K. B. Joldersma, I. Kiyak, C. A. Lepczyk, P. Paton, M. J. Sanregret, A. Souza, G. C. White, and an anonymous reviewer provided valuable comments on earlier drafts. J. McGrath provided frog-call recordings. J. H. Harding and J. McGrath provided expert identifications and abundance estimation for the CD recording. We thank the hundreds of volunteers in the Michigan Frog and Toad Survey for their continuing efforts. Funding was provided by the MDNR Wildlife Division, Michigan State University chapter of Sigma Xi, and the Ecology, Evolutionary Biology and Behavior graduate program at Michigan State University.

Literature cited

Alford, R. A., and S. J. Richards. 1999. Global amphibian declines: a problem in applied ecology. Annual Review of Ecology and Systematics 30: 133-165.

Bishop, M. E., K. E. Pettit, M. E. Gartshore, and D. A. Macleod. 1997. Extensive monitoring of anuran populations using call counts and road transects in Ontario (1992-1993). Herpetological Conservation 1: 149-160.

Blaustein, A. R., D. B. Wake, and W. P. Sousa. 1994. Amphibian declines: judging stability, persistence, and susceptibility of populations to local and global extinctions. Conservation Biology 8: 60-71.

Crouch, W. B., and P. W. C. Paton. 2002. Assessing the use of call surveys to monitor breeding anurans in Rhode Island. Journal of Herpetology 36: 185-192.

Driscoll, D. 1998. Counts of calling males as estimates of population size in the endangered frogs *Geocrinia alba* and *G. vitellina*. Journal of Herpetology 32: 475-481.

Erskine, A. J. 1978. The first ten years of the cooperative Breeding Bird Survey in Canada. Canadian Wildlife Service Report Series No. 42, Hull, Quebec, Canada.

Green, D. M. 1997. Perspectives on amphibian population declines: defining the problem and searching for answers. Herpetological Conservation 1: 291-308.

Hemesath, L. M. 1998. Iowa's frog and toad survey, 1991-1994. Pages 206-216 *in* M. J. Lannoo, editor. Status and Conservation of Midwestern Amphibians. University of Iowa, Iowa City, USA.

Huff, J. 1991. Frog and toad survey 1991. Pages 114-123 *in* Madison: Wisconsin Department of Natural Resources. Technical Report. Madison, Wisconsin, USA.

Kendall, W. L., B. G. Peterjohn, and J. R. Sauer. 1996. First-time observer effects in the North American Breeding Bird Survey. The Auk 113: 823-829.

Kline, J. 1998. Monitoring amphibians in created and restored wetlands. Pages 360-368 *in* M. J. Lannoo, editor. Status and Conservation of Midwestern Amphibians. University of Iowa, Iowa City, USA.

Lepage, M., R. Courtois, and C. Daigle. 1997. Surveying calling anurans in Quebec using volunteers. Herpetological Conservation 1: 128-140.

Link, W. A., and J. R. Sauer. 1996. Extremes in ecology: avoiding the misleading effects of sampling variation in summary analyses. Ecology 77: 1633-1640.

Mossman, M. J., L. M. Hartman, R. Hay, J. R. Sauer, and B. J. Dhuey. 1998. Monitoring long-term trends in Wisconsin frog and toad populations. Pages 169-198 *in* M. J. Lannoo, editor. Status and Conservation of Midwestern Amphibians. University of Iowa, Iowa City, USA.

Pechmann, J. H. K., and H. M. Wilbur. 1994. Putting declining amphibian populations in perspective: natural fluctuations and human impacts. Herpetologica 50: 65-84.

Sargent, L. G. 2000. Frog and toad population monitoring in Michigan. Journal of the Iowa Academy of Sciences 107: 195-199.

SARKAR, S. 1996. Ecological theory and anuran declines. BioScience 46: 199-207.

SAUER, J. R., B. G. PETERJOHN, AND W. A. LINK. 1994. Observer differences in the North American Breeding Bird Survey. Auk 111:50-62.

SCOTT, N. J., JR., AND B. D. WOODWARD. 1994. Surveys at breeding sites. Pages 118-125 *in* W. R. Heyer, M. A. Donnelly, R. W. McDiarmid, L. C. Hayek, and M. S. Foster, editors. Measuring and monitoring biological diversity: standard methods for amphibians. Smithsonian Press, Washington, D.C., USA.

SHIROSE, L. J., C. A. BISHOP, D. M. GREEN, C. J. MACDONALD, R. J. BROOKS, AND N. J. HELFERTY. 1997. Validation tests of an amphibian call count survey technique in Ontario, Canada. Herpetologica 53:312-320.

TRAVIS, J. 1994. Calibrating our expectations in studying amphibian populations. Herpetologica 50: 104-108.

WEIR, L. A., AND M. J. MOSSMAN. In press. The protocol and history of the amphibian calling survey of the North American Amphibian Monitoring Program (NAAMP). *In*: M. J. Lannoo, editor. Declining amphibians: a United States' response to the global phenomenon. University of California Press, Berkeley, USA.

ZAR, J. R. 1998. Biostatistical Analysis. Fourth edition. Prentice Hall, Englewood Cliffs, New Jersey, USA.

ZIMMERMAN, B. L. 1994. Audio strip transects. Pages 92-97 *in* W. R. Heyer, M. A. Donnelly, R. W. McDiarmid, L. C. Hayek, and M. S. Foster, editors. Measuring and monitoring biological diversity: standard methods for amphibians. Smithsonian Press, Washington, D.C., USA.

Kristen S. Genet is currently a doctoral student in the Department of Zoology and Ecology, Evolutionary Biology and Behavior program at Michigan State University in East Lansing, Michigan. She received her M.S. in zoology from Michigan State University in 1999, and a B.A. in biology in 1996 from Gustavus Adolphus College in St. Peter, Minnesota. Her research has been focused on habitat fragmentation, landscape ecology, and responses of amphibian and reptile populations to impacted habitats. Currently her research is focused on the abundance and distribution of frogs and toads in southern Michigan, including analyses of population trends from the Michigan DNR Frog and Toad Survey. ***Lori Sargent*** has been a nongame wildlife biologist with the Wildlife Division of the Department of Natural Resources in Lansing, Michigan since 1994. She received her B.S. in wildlife biology from Michigan State University and M.S. in wildlife biology from Purdue University. Professional interests and expertise include woodland hawks and other raptors and herpetology. Lori has served as the membership chairperson for the Michigan Chapter of the Wildlife Society for 8 years and has served as Secretary for the same chapter as well as the Michigan State University student chapter. Her interests include not only wildlife biology but also bird hunting and training pointing dogs for field work, obedience, and other competitive events. Along with her husband, Mark, she breeds, trains, and loves German wirehaired pointers.

Associate editor: *White*

Appendix

Survey questions presented to active volunteers on questionnaire in Michigan, 2001. Responses were edited and evaluated as described in the text.

1. When did you become involved with the Frog and Toad Survey (1996, 1997, 1998, 1999, 2000, 2001)?
2. How many years have you submitted data (1, 2, 3, 4, 5, 6)?
3. How many years do you anticipate submitting data?
4. How many years have you been involved with frogs, herps, or wildlife in general as an avid hobbyist or professional (0-5, 6-10, 11-15, 16-20, >20)?
5. With respect to the Frog and Toad Survey, please rate your current level of expertise at which you perceive yourself, on a scale from 1 (low) to 5 (high).
6. Please rate the level of expertise at which you perceived yourself before getting involved with the Michigan Frog and Toad Survey, on a scale from 1 (low) to 5 (high).
7. What forms of training did you participate in before beginning to survey your routes (Please mark all that apply) (attended training workshop, listened to training tape, practice in the field with a trained observer, other—specify)?
8. How many people participate in your survey runs each year?
9. Is this number consistent each year (yes, no)?
10. Do the same people participate in the survey each year (yes, no)?
11. Is there one person designated as the primary observer (yes, no)?
12. Are you the primary observer (yes, no)?
13. Please estimate the amount of time you spend listening at each site along your route, in minutes.
14. During the time of your involvement with this program, have there been discrepancies among observers in species heard or call indices that should be assigned (yes, no)?
15. If yes to previous question, please describe how these discrepancies are resolved in the space that follows.
16. Are you also an avid birder (yes, no)?
17. If yes, do you most often identify birds by sight, song, or both?
18. Are you involved in any additional wildlife monitoring programs (yes, no)?
19. If yes, please indicate which monitoring programs in which you participate (Marsh Monitoring Program, North American Breeding Bird Survey, Christmas Bird Count, Other—specify).