RELIABILITY OF SINGING BIRD SURVEYS: CHANGES IN OBSERVER EFFICIENCY WITH AVIAN DENSITY

JONATHAN BART AND JAMES D. SCHOULTZ
Ohio Cooperative Wildlife Research Unit, Department of Zoology, Ohio State University, Columbus, Ohio 43210 USA

ABSTRACT.—Field trials in which paired observers were used and indoor simulations in which recorded bird songs were used indicated that, as the number of singing birds audible from a listening station increased from 1 to 4, the fraction of them recorded by observers declined by up to 50%. This reduction in efficiency violates one of the basic assumptions of any index—that the proportion of animals detected remains constant—and could cause surveyors who rely primarily on auditory cues to underestimate changes in population density by up to 25% for common species and 33% for abundant species. The change in efficiency, which is best regarded as measurement error, cannot be detected by a statistical examination of the data and thus may pass undetected in many field studies. It seems unlikely that any general procedure for "correcting" the error would be reliable. The results indicate that singing bird surveys of common species should be supplemented by other methods if accurate estimates of changes in density are needed.

A general conclusion of the study is that whenever animals "compete" for a place in the survey, for example by filling up traps or suppressing one another's songs, then the index tends to underestimate a change in density. If efficiency increases with density, then the survey tends to overestimate a change in density. If the sign of the bias can be determined, the survey can be used to provide a minimum or maximum estimate of a change in density even if the magnitude of the bias cannot be estimated. *Received 11 July 1983, accepted 9 January 1984.*

TWO approaches are available to estimate a difference in avian densities from transect or point-count data. The counts may be translated into density estimates, or the difference in counts may be used as a direct estimate of the difference in densities. Density estimation is not reliable in many habitats, because the critical assumption that observers record all birds near them is not valid (Mayfield 1981). In addition, many of the density-estimation methods require accurate estimates of the distance to each bird, which may be difficult to obtain. Index methods require that the same proportion of birds in each population be detected. Many observers feel that this is a more reasonable assumption than that all birds near the observer are detected. They therefore recommend the index method when a change in density, rather than density itself, is to be estimated (Dawson 1981). Historically, far more data have been analyzed by means of index methods than by means of density estimation, and indices continue to be used in numerous regional and national surveys (grouse: Rodgers 1981; quail: Robel et al. 1969; pheasant: Carney and Petrides 1957; turkey: Scott and Boeker 1972; woodcock: Duke 1966; doves: Sayre et al. 1978; other gamebirds: Sanderson 1977; Breeding Bird Survey, many terrestrial species: Bystrak 1981).

Although index methods are widely used, some observers have expressed concern over the assumption that surveyors record the same proportion of each population (Caughley 1977, Sayre et al. 1980, Dawson 1981). Surveys are usually designed so that factors such as time of day and season, weather conditions, observer skill, and habitat parameters are either standardized or have similar distributions in each population. A more difficult problem is posed by factors that respond directly to density. If, for example, song output per bird varies with density, then there is no way to design the survey so that the average song output is the same in each population. This issue, while potentially quite serious, has received little detailed study. Walankiewicz (1977) and Frochot et al. (1977) concluded, on the basis of indirect evidence, that observers missed a higher proportion of birds when density was high. Dawson (1981) and Järvinen and Väisänen (1976) examined the issue briefly but could not reach definite conclusions. Bystrak (1981) mentioned

The Auk 101: 307-318. April 1984

"overloading" as a possible problem for some uses of Breeding Bird Survey data. Scott and Ramsey (1981) compared counts of one and three target species when observers listened only for the target species and when they listened for all species. When listening for all species, they recorded 25-50% fewer individuals of the target species, suggesting a pronounced saturation effect. None of these studies provided estimates of how much observer efficiency might change with density.

These reports pertain mainly to interspecific effects: the increased likelihood of missing a bird when numerous other species are present. Intraspecific factors may also cause problems. DeSante (1981) reported, in a study of the variable circular-plot method (Reynolds et al. 1980), that at high density some birds were missed and at low density they were over-counted. This caused the variable circular-plot estimates to be inversely correlated with actual density (determined through intensive surveys), and the same would be true if the simple index results had been used. DeSante attributed the errors in part to greater movement by birds (and thus double counting) in larger territories. Other factors that may also be important are that, when several individuals of a species are audible from a listening station, a near bird may mask a far one, two individuals may be close together and therefore difficult to separate, and after the observer records one or two individuals of a species his attention may turn to other species he thought he heard or thought he ought to have heard based on the habitat. Field experience suggested to us that the problem of reduced observer efficiency when many conspecifics are present might be of significance. We undertook this project to determine whether or not observer efficiency does decrease as the number of singing conspecifics increases and, if it does, to determine how seriously this compromises the singing bird index.

METHODS

Two methods were used to measure how efficiency—the average proportion of singing birds recorded by the observer—changes with the number singing. In the first method we accompanied observers conducting surveys of breeding birds. During the first 4-5 h of daylight, observers counted all birds heard during 3 min spent at each of 50 roadside stations. The survey resembled the Breeding Bird Survey (Bystrak 1981), but the birds detected only by visual cues were excluded from the analysis. While the surveyors listened for all species, we each listened for only one species. The observers were unaware of which species we listened for. We calculated the average number of birds they recorded when we recorded one individual, two individuals, etc., and we used our results as a standard with which to estimate their efficiency at different actual densities. During 2 yr of fieldwork in east-central Ohio, we accompanied 11 observers on 16 routes and recorded more than 3,000 individual birds.

The field methods gave us considerable realism, but we could not control extraneous variables such as traffic noise, dispersion of the birds, and habitat characteristics. In addition, there was no objective measure of how accurate our counts were. We therefore designed an indoor simulation of the dawn chorus using tape-recorded bird songs. This experiment provided the needed control and an independent check on the field-study results.

In preparation for the indoor trials, two seasons were first spent studying the vocal behavior of birds in the study area. The fieldwork described above provided estimates of the number of species, and individuals of each species, that typically sing during a 3-min listening period. We also monitored song duration of individual birds at randomly selected roadside locations. During 840 15-min periods in June of 1980 and 1981, we recorded the amount of song for each individual of species selected in advance. The results were used to construct species-specific frequency distributions of song duration during 3-min intervals (Table 1). Song duration was defined as the total time in 15-s intervals during which the bird sang at least once. If in a 3-min interval the bird sang once in the first 15-s interval and once in the final 15-s interval, duration of the song would be recorded as 30-s. We used the results of this fieldwork to insure that the distribution of durations in our simulations was realistic.

The simulation consisted of 43 3-min listening periods during which recorded songs were played from loudspeakers surrounding the surveyors. Each listening period had 12 species and 20 or 21 individuals singing at typical rates. Approximately two-thirds of the species and individuals were placed in two-thirds of the circle labelled "scrubgrowth"; the remainder were placed in the "forest" segment. Volume (simulating distance) and duration varied within the listening period but had identical distributions between periods. Conspecifics were spaced out as evenly as possible over no more than one-third of the circle.

We use the term "treatment group" to designate the set of periods having the same density of a particular species. Within each species, the distributions of song volume and duration were similar in each treatment group. For example, the average duration of song by American Robins (*Turdus migratorius*) in periods with one robin present was the same as in

TABLE 1. Frequency of song durations during 3-min periods in which the individual sang at least once. Presence or absence of song in each 15-s interval was recorded.

Species	Number of periods	Number 15-s intervals with song 1–2	3–4	5–6	7–8	9–10	10–11
Mourning Dove (*Zenaida macroura*)	91	0.35	0.24	0.19	0.05	0.08	0.09
Great Crested Flycatcher (*Myiarchus crinitus*)	55	0.42	0.15	0.07	0.07	0.11	0.18
American Crow (*Corvus brachyrhynchos*)	35	0.48	0.23	0.03	0.03	0.06	0.17
House Wren (*Troglodytes aedon*)	95	0.12	0.20	0.16	0.12	0.15	0.26
Gray Catbird (*Dumetella carolinensis*)	376	0.26	0.18	0.14	0.07	0.11	0.24
Brown Thrasher (*Toxostoma rufum*)	18	0.33	0.11	0.11	0.17	0.06	0.22
American Robin (*Turdus migratorius*)	108	0.16	0.14	0.11	0.07	0.17	0.35
Wood Thrush (*Hylocichla mustelina*)	42	0.29	0.05	0.14	0.02	0.05	0.45
Red-eyed Vireo (*Vireo olivaceus*)	130	0.12	0.13	0.11	0.08	0.12	0.45
Common Yellowthroat (*Geothlypis trichas*)	304	0.16	0.13	0.09	0.08	0.08	0.46
Northern Cardinal (*Cardinalis cardinalis*)	140	0.26	0.21	0.14	0.13	0.11	0.15
Indigo Bunting (*Passerina cyanea*)	377	0.09	0.09	0.09	0.07	0.09	0.57
Rufous-sided Towhee (*Pipilo erythrophthalmus*)	28	0.25	0.21	0.11	0.07	0.14	0.21
Field Sparrow (*Spizella pusilla*)	167	0.21	0.10	0.13	0.10	0.09	0.38
Song Sparrow (*Melospiza melodia*)	360	0.24	0.12	0.16	0.11	0.09	0.28

periods with two robins present. A total of 39 species occurred during the 43 periods; the experiments were designed to study 16 of them. Each of these 16 species had 10 periods with one individual present and 8 periods with 2 individuals present. Among 7 species, densities of 3 (5 periods) and 4 (4 periods) also occurred.

A total of 20 experienced birdwatchers from Maryland, all of whom had run Breeding Bird Surveys, participated in the trials, which were held in a large high school cafeteria. Participants arrived for the simulation at 1700 and were acquainted with the layout of the experiment before they began the 4–4.5-h trial. They were not informed of the purpose of the study or of any of the parameters, such as number of species per listening period. They stood in a 1.5-m-diameter circle surrounded by 27 loudspeakers equally spaced around a circle of 20-m radius. Each loudspeaker was attached to a separate cassette player located in a control booth adjacent to the circle. Four people worked in this booth playing the cassettes at predetermined times during each 3-min listening period. Each cassette contained songs of a single species. There were no call notes, and no other birds were audible on the tapes. The surveyors used a separate sheet for each listening period to avoid recording errors; they were asked not to discuss their results until after the entire trial. Two to five people participated per night. The simulations were identical each night.

RESULTS

The fieldwork and indoor simulations both indicated, for all species studied, that efficiency declined steadily and markedly with increasing density (Fig. 1, Table 2). In only one case was there an increase in efficiency with an increase in density [Field Sparrows (*Spizella pusilla*) in the field trials between densities 1 and 2]. For all species combined, the field-trial estimates declined 32% from 0.72 at density = 1 to 0.49 at density = 4. Comparable figures from the indoor simulations were 1.1 to 0.56, a 49% decline. In the indoor simulations, the decline in efficiency was greatest between the first two density classes, but this trend was absent or weak in the field-trial data.

Although the trends shown by the two methods are similar, some differences can be seen. The greatest difference is that efficiencies from the indoor trials were higher than those from the field study, perhaps because the participants requested that we set the volumes higher than we had intended to. Indoor efficiencies also decreased somewhat faster than outdoor ones, probably because the individuals of each species were closer together in the indoor simulation, on average, than is usually true in the field.

During the fieldwork, our observers frequently recorded individuals of a species we were monitoring when we had not detected the species. We suspected, but could not prove, that these were misidentifications. This impression was confirmed by the tape recorder sessions. All observers occasionally recorded species not present during the listening period. The average number of "phantom species" re-

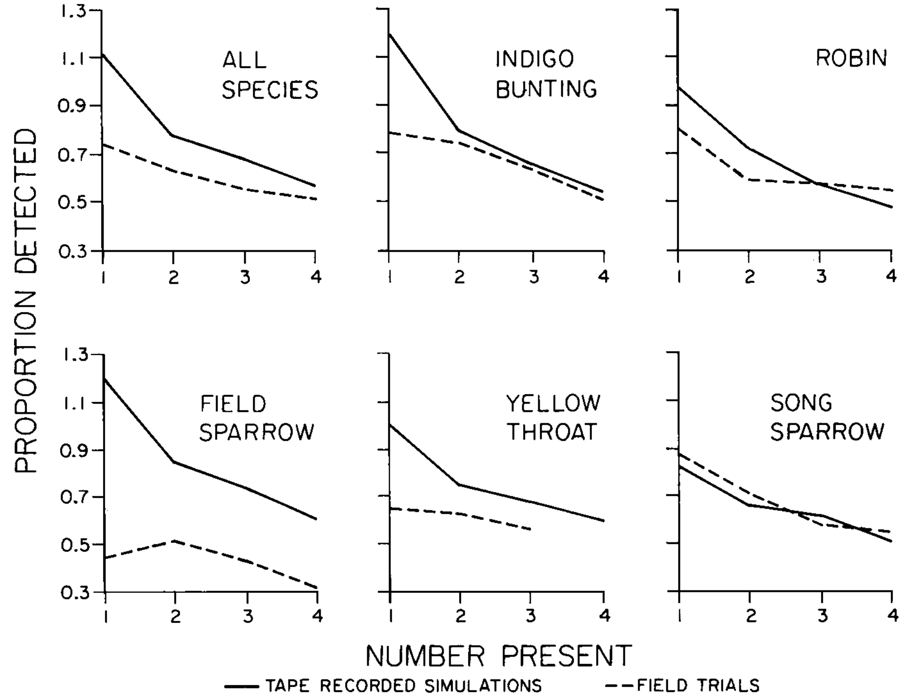

Fig. 1. Relationship between number of birds present and average proportion detected in field and indoor trials. For standard errors, see Table 2.

corded was 0.67 per listening period. The averages for individual observers varied from 0.19 to 2.02; 80% of the observers had averages between 0.37 and 0.95 inclusive. A related statistic is that, for 8 of the 16 species studied in detail, when only 1 individual was present, 2 were recorded more often than 0, producing "efficiencies" exceeding 1.0 (Fig. 1, Table 2).

Table 2. Relationship between number of birds present and average proportion detected (efficiency) in field trials (excludes species in Fig. 1) and standard errors for field and indoor trials.

	Number present			
	1	2	3	4
A. Point estimates of efficiency				
Mourning Dove	1.02	0.72	0.68	0.56
American Crow	1.08	0.93	0.72	0.62
Eastern Wood-Pewee (*Contopus virens*)	1.30	0.87	0.74	0.62
Red-eyed Vireo	1.12	0.82	0.62	0.53
Gray Catbird	0.64	0.55	—	—
Brown Thrasher	0.78	0.57	—	—
Wood Thrush	0.94	0.72	—	—
Great Crested Flycatcher	0.84	0.67	—	—
House Wren	0.84	0.72	—	—
Rufous-sided Towhee	0.73	0.63	—	—
Northern Cardinal	0.95	0.74	—	—
B. Pooled standard errors[a]				
Field trials with 1-4 present	0.069	0.051	0.041	0.035
Indoor simulations with 1-2 present	0.046	0.035	—	—

[a] Calculated by considering the results from all species as a single data set.

The finding that even highly competent observers regularly misidentify species or overestimate their numbers may be of interest to field workers, but it should not cause undue concern about the singing-bird index. If an observer always recorded twice as many birds as were present, the index would estimate change in density with perfect accuracy. The same is true of the observer who always records 10% of those present.

The change in efficiency with density was similar among species. In the indoor simulations, less common species (max. density = 2) had lower average efficiencies. Their proportional change in efficiency (the component of bias we were studying) was not consistently different from that of the more common species.

These results demonstrate clearly that efficiency does change with density, and they provide an estimate of the magnitude of the change. The next step is to determine how much this variation affects the ability of the survey to measure changes in population density. To do this, let

$$\bar{N}_1 = \sum_1^4 df_{1d} = \text{average number present/station in year 1, and}$$

$$\bar{n}_1 = \sum_1^4 df_{1d}e_{1d} = \text{average number recorded/station in year 1,}$$

where f_{1d} = proportion of stations in year 1 with d birds present, and e_d = average proportion recorded when d birds are present (efficiency). If similar terms are defined for year 2, then

$$\text{actual change} = \frac{\bar{N}_2}{\bar{N}_1} - 1,$$

$$\text{estimated change} = \frac{\bar{n}_2}{\bar{n}_1} - 1,$$

and the bias in the survey estimate of change in population density may be expressed as

$$\text{relative bias} = \frac{\textit{estimated change}}{\textit{actual change}} - 1$$

$$= \frac{\dfrac{\sum df_{2d}e_d}{\sum df_{1d}e_d} - 1}{\dfrac{\sum df_{2d}}{\sum df_{1d}} - 1} - 1. \qquad (1)$$

Bias arises from variation in the e_d. It is identical to zero if all $e_d = e$. The sign of relative bias indicates the direction of the error; relative bias = −0.11 means that the survey, on average, would underestimate the true change by 11%.

Some difficulty may be caused by our use of the phrase "number of birds present," because this number depends on how much area is included, and it is not immediately clear what size area is appropriate. If we assume that the listening stations are placed randomly with respect to birds in the population we wish to make inferences about (an assumption required for any statistical use of the data), then, in calculating relative bias, the size of the area around the listening station is of no consequence as long as it only includes birds in the population of interest. Conceptually, it is easiest to imagine that the area includes all the birds that even the best observer may record. How much additional area is included does not matter, because the relative bias is determined by the proportional change in efficiency, $(\bar{n}_1/\bar{N}_1)/(\bar{n}_2/\bar{N}_2)$. Changing the size of the area used to calculate $\bar{N}_1$ and $\bar{N}_2$ simply changes these numbers by a constant multiplier; it has no effect on the proportional change in efficiency.

The basic question addressed in this study is how closely $\bar{n}_2/\bar{n}_1$ estimates $\bar{N}_2/\bar{N}_1$. The field and indoor trials provide estimates of e_d, but to estimate relative bias the change in frequency of each density class between years must also be determined. This point can be made clear by an example. If the zero-density class increased during a population decline and all other frequencies decreased by the same proportional amount, $f_{2d}/f_{1d} = a$ for all d, then substituting $f_{2d} = f_{1d}\, a$ in equation (1) yields zero, indicating that the relative bias would be zero regardless of how much efficiency changed with density. This example is not realistic, because it is unlikely that the frequency of each density class would decline by the same proportional amount, but it shows that the change in frequency cannot simply be ignored in the calculations.

We investigated two models specifying how the frequency of each density class changes as the result of a population decline. Both models assume that density is the only change between years. Under this assumption, the frequencies in year 2 are

$$f_{2d} = \sum_{N=d}^{4} f_{1N}b_{d/N},\ d = 0, \ldots, 4, \qquad (2)$$

where $f_{1N} = f$ (stations in year 1 with N birds

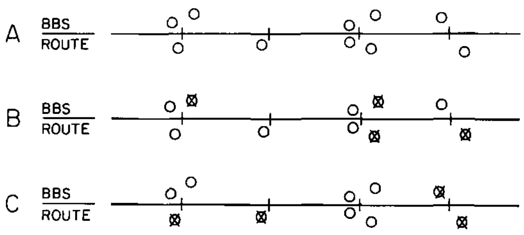

Fig. 2. Two models of how a population decline affects the distribution of bird densities at survey stations. A. Distribution of birds if no decline had occurred. B. Losses distributed randomly (constant-loss model). C. Losses concentrated at stations that would have had fewer individuals present (variable-loss model).

TABLE 3. Sensitivity of relative bias to initial frequencies and change in efficiency.[a]

A. Initial distributions and efficiencies

Parameter	Type	Density 1	2	3	4
Initial frequency	A	0.80	0.20	—	—
	B	0.60	0.30	0.10	—
	C	0.40	0.30	0.20	0.10
	D	0.20	0.30	0.40	0.10
Efficiency	Field study	0.72	0.63	0.55	0.49
	Lab study	1.10	0.78	0.67	0.56

B. Relative bias (actual decline = 0.10)

Distribution type	Evaluation	Survey result Before	Survey result After	Estimated decline	Relative bias
A	Field study	0.828	0.748	0.097	−0.030
	Lab study	1.192	1.084	0.091	−0.094
B	Field study	0.975	0.887	0.090	−0.097
	Lab study	1.329	1.220	0.082	−0.180
C	Field study	1.130	1.034	0.085	−0.150
	Lab study	1.477	1.366	0.075	−0.248
D	Field study	1.378	1.270	0.078	−0.216
	Lab study	1.716	1.601	0.067	−0.330

[a] Random-loss model (see text).

present), and $b_{d|N}$ = proportion of the stations with N birds in year 1 that have d birds in year 2. In the first model, vacancies (at locations that were occupied in Year 1) are assumed to be randomly distributed among the density classes. If we define

$$p = \bar{N}_2/\bar{N}_1,$$

then $b_{d|N}$ is the binomial probability

$$b_{d|N} = \binom{N}{d} p^d (1 - p)^{N-d}. \qquad (3)$$

This is referred to as the constant-loss model (Fig. 2). The second model recognizes that there may be a correlation between habitat quality and number of birds present in Year 1 at a station. In this case, vacancies in Year 2 might be concentrated at stations that had high or low density in Year 1. Thus, the probabilities of a vacancy ($1 - p$) would vary with density class, so we substitute p_N, $N = 1, \ldots 4$, in equation (3). Given an overall change in density, p, and the initial density class frequencies, f_{1N}, the p_N may be calculated by specifying one frequency as a dummy variable, say $p_4 = k$, and then defining the other probabilities in terms of k: $p_3 = 0.9k$, $p_2 = 0.5k$, $p_1 = 0.3k$, for example. This would indicate that vacancies were concentrated at stations having lower density in Year 1. The unknown, k, can then be calculated by iterative methods from which the p_N can be determined. They are substituted for p in equation (3). We refer to this as the variable-loss model (Fig. 2).

The results of the simulations are that the relative bias increased as the proportion of stations with more than one bird present increased (Table 3). The proportion of stations without any birds present has no effect on relative bias, as seems intuitively reasonable. Furthermore, if all stations with any birds had exactly one bird present, then relative bias would be zero, because the samples in each year would be based on only one efficiency. As the proportion of stations with only one individual present dropped to 60%, 45%, and 20%, relative bias rose steadily, reaching 33% in the worst case investigated.

Relative bias was higher for the efficiencies obtained in the laboratory study, because there was a greater change in the lab efficiencies than in those from the field. With frequency type A (Table 3), there was a three-fold difference in the estimates of relative bias, depending upon whether the efficiencies obtained in the lab or field were used. With frequency type D, this difference dropped to about one-third. Thus, when relative bias is large enough to be of con-

TABLE 4. Relative bias with three different population declines.[a]

Change in efficiency	Decline in population density		
	0.02	0.10	0.25
Field study	−0.156	−0.150	−0.126
Lab study	−0.256	−0.248	−0.216

[a] Initial frequency distribution: type C, Table 4.

cern to investigators (Types C, D), it makes little difference whether one uses the laboratory or field efficiencies to calculate the bias. This indicates that the exact value of the true efficiencies may be of little consequence.

The magnitude of relative bias is also insensitive to the magnitude of the population decline over the range of values we investigated (Table 4). Changing the size of the decline by more than an order of magnitude caused only a small change in the size of relative bias with either the laboratory or the field efficiencies. All of these results were obtained using the random-loss model.

The results produced by the variable-loss model were nearly identical, tending to be slightly smaller when high-density stations suffered lower than average losses (because in that case a larger proportion of stations changed from 1 present to 0 present). Because there is probably no way to be sure what loss model is most realistic, it is fortunate that the results are little affected by this parameter.

DISCUSSION

The magnitude of the population decline was underestimated in all cases. This would also be true if density increased, for underestimates are caused by the inverse relationship between efficiency and density. In many cases, it may be possible to guess the direction of change in efficiency. It is therefore worth stressing that if efficiency varies directly with density, then change will be overestimated, whereas if, as is probably true of many indices, the relationship is inverse, the change will be underestimated. Anytime animals "compete" for a place in the survey, for example by filling up a trap or suppressing one another's song, the survey will tend to underestimate changes in density. Conversely, if efficiency increases with increasing density, then the survey will tend to overestimate change. This trend appears to occur sometimes at hawk migration stations (Sattler and Bart in prep.). Efficiency increases with increasing numbers of raptors passing the lookout, perhaps because the observer becomes more attentive then and because high-flying birds pass in clusters. If the observer sees any one of the birds, he is likely to see them all. These examples suggest that sometimes it may be possible to predict the direction of the bias caused by changing efficiency so that the estimated change in density can be viewed as a minimum or maximum.

It can be argued that the relative bias uncovered in this study has little effect on the survey's ability to detect change in density, even though estimating its exact magnitude may be difficult. With a relative bias of −0.15 to −0.25, the survey results would be expected to decrease 7.5–8.5% if a 10% decline occurred in a common species, clearly showing that a decline had occurred. There would be some loss of statistical power, because the expected difference is up to 25% smaller than it would be in the absence of relative bias, and thus the significance of the observed difference would be lower than if relative bias were zero. But, the argument goes, extensive programs such as the Breeding Bird Survey often have very large samples. Hypothesis tests are therefore likely to establish the significance of observed declines despite the loss of power caused by relative bias—at least if the true decline is large enough to be of any biological importance.

There is considerable merit in this argument. The problem of relative bias *is* most serious for investigators trying to measure the magnitude of a change in density rather than simply establish its occurrence. A few rejoinders should be made to the points above, however. First, many investigators do not have the large samples typical of a national monitoring program. Suppose, for example, that a random sample of 1,000 3-min listening stations is taken in each of two years during which a true decline of 10% occurs. If the SD is 0.65, then standard power calculations show that the survey has a 99.8% chance of detecting the decline (with $\alpha = 0.05$) if the relative bias is zero, and this probability is 95.5% even with relative bias equal to 25%. Thus, the presence of relative bias has little impact in this case. With a sample of 300, however, the power is 77% with no relative bias, a respectable figure, but drops to 51% with a relative bias of 25%. Many investigators probably would feel that a study with a power of only 51% is not worth doing. Another approach is to contrast situations with and without relative bias. Relative bias might be reduced to nearly zero, for example, if observers avoided the dawn chorus or counted only a few species. If we wish the power to be the same in both cases, then how much larger a sample must be taken in the case with relative bias not equal to zero? The answer is $(1 + b)^{-2}$, where b is relative bias. With a relative bias of -0.25, the sample must be 78% larger than if the relative bias were 0.0. Thus, with the sample sizes typical of small studies, relative bias has a substantial impact on power. If the relative bias is assumed to be absent, then the power may be considerably overestimated, and the increase in sample size needed to compensate for the effects of relative bias may be considerable.

The second point, which must be mentioned in response to doubts about the importance of relative bias, is that even if a very large sample is taken, relative bias may still cause trouble. If two areas are being compared and a statistically significant change in sample results occurs, then two interpretations are possible: that there is a difference in density or that there is a difference in detectability. Suppose, for example, that the frequencies of stations with 1 and 2 birds actually present are 40% and 10% at one location, and 30% and 15% at the other (no difference in density). The application of the efficiencies obtained in the indoor trials indicates that a 5% difference in sample results would be expected. The bias is caused by the second location having a higher frequency of stations with lower efficiency. Even if a single location is being studied, the average efficiency may change in response to long-term changes in habitat. The United States Fish and Wildlife Service Mourning Dove [*Zenaida macroura*] Coo Count routes (Baskett et al. 1978) have shown such a trend in some areas in recent years, and wildlife managers are currently trying to decide whether a real decline has occurred or whether an artifact such as the one described here has caused the change (D. Dolton pers. comm.).

Which of the density distributions (Types A–D, Table 3) are most realistic? Type D, the most dense, would apply to few species, as more than 50% of stations with the species present had three or four individuals. Such densities are unlikely to occur in surveys like the Breeding Bird Survey where the routes are laid out randomly and thus pass through many different habitat patches. In such cases, many stations have only a small amount of favorable habitat for a given species within hearing range. This is evident in the BBS data. For example, among all BBS routes surveyed in Ohio and West Virginia in 1979–1981, Common Yellowthroats (*Geothlypis trichas*) occurred singly at about 80% of the stations (with the species present), whereas Indigo Buntings (*Passerina cyanea*) and Red-eyed Vireos (*Vireo olivaceous*) occurred singly at 50–60% of the stations (Bart unpubl. data). At these frequencies, the relative bias will be less than 15% in most cases. On the other hand, in a study with transects intentionally placed in homogeneous habitat, as occurs in many research projects, higher densities such as Types C or D may well occur. In the area of Ohio that we studied while estimating song frequencies, several species, including most of the ones we evaluated in the field study, most closely approximated the Type C distribution. A study solely in scrubgrowth or woods might well encounter species having Type D distributions. Thus, the problem uncovered in this study is unlikely to be serious in extensive surveys using randomly laid-out routes but may well be significant when routes are located primarily in homogeneous habitat.

Diurnal timing of the surveys may also be important in determining the magnitude of relative bias. During the first hour of daylight, most birds sing more actively, and some, especially robins, sing far more actively than they do later in the morning. This proliferation of song makes it difficult to separate individuals from one another and probably intensifies the decline of efficiency with density. Thus, when relative bias is a concern, surveyors should consider avoiding the early morning period. This will often decrease the number of birds detected, of course, but the effect on precision may be surprisingly small (Bart and Herrick 1984).

An unexpected finding of the study was that the contribution to relative bias from stations with two birds present is at least as great as the contribution from stations with four birds present. The bias is caused by a proportional change in efficiency, e_{d+1}/e_d, $d = 1, 2, 3$, most of which varied from 0.85 to 0.90. The indoor trials, however, yielded a great change in efficiency between one and two birds present. As an ex-

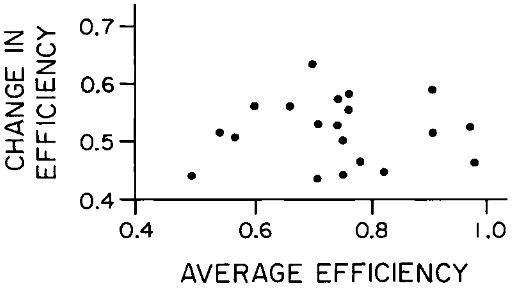

Fig. 3. Relationship between average efficiency and change in efficiency during indoor trials. Each dot represents the results for one observer. Change in efficiency = efficiency with 4 present/efficiency with 1 present.

ample of how important this is, if stations all had two birds present and the indoor efficiencies applied, then (with a 10% decline) relative bias would be −37%, higher than in any of our other examples. Thus, it is not always safe to assume, as was, for example, by Bystrak (1981), that relative bias will be negligible simply because densities do not exceed three or four.

The indoor simulations provided one of the first data sets in which it was known what birds were present, and conditions were sufficiently standardized to enable clear trends to emerge in the survey results. This makes it possible to examine several other issues related to the use of singing-bird survey data.

Efficiency vs. relative bias.—One question that may interest field workers is: Are observers with high overall efficiency most successful in estimating a change in population density? Although intuition might suggest that the answer is yes, there was no detectable relationship between observer efficiency (over the range of efficiencies in our sample) and relative bias calculated from their data (Fig. 3). This may be explained by an observation we have often made in the field. Keen observers frequently seem to concentrate on detecting rare species at the expense of getting more individuals of the common species. The senior author participated in both the field trials and the indoor trials, and his change in efficiency was about average. It thus appears that the problem of changing efficiency is difficult to overcome even if the observer is aware of the danger.

Table 5. Comparison of species lists when observers were counting only species and when they were counting species and individuals. In all cases, 12 species were actually present.

	Number species recorded when counting	
Listening period	Species only	Species and individuals
1	11.04	11.39
2	10.61	10.26
3	10.04	10.04
4	10.04	10.43
5	10.74	10.74
6	10.52	10.43
7	10.83	10.56
Average	10.55	10.55
SE	0.14	0.16

Presence/absence data.—It has sometimes been suggested that if observers only recorded species and did not try to count individuals, their data might be more accurate. Such a plan would raise formidable analysis problems, at least in trying to estimate a change in density, and might be opposed on that ground alone. Another question, which our tape-recorder trials provided a means of answering, is whether or not species lists are significantly more accurate when observers are not counting individuals. To answer this question we interspersed the listening periods with seven "species only" periods during which observers were instructed not to count individuals. Each species-only period was an exact replica of one of the regular periods and was separated from it by 10 other periods. In three cases the species-only period came first; in 4 cases the regular period came first. These steps were taken to eliminate any effects of learning or fatigue. The results (Table 5) provide no support for the hypothesis that species lists will be more accurate if only species are being surveyed. It seems likely that with more than 12 species present—as often happens in the field—it would be at least slightly easier to detect species if individuals were not counted. Our results suggest that the average difference for all stations is unlikely to be large, however.

Fatigue.—Many breeding-bird surveyors have remarked that 50 3-min periods makes a very long morning, and this has caused concern that towards the end of the experiment data collection might be less diligent. Our experiment provides no support for this suggestion either; efficiency did not vary systematically during the trials. Average efficiencies for the periods

TABLE 6. Ratio of maximum number recorded by members of a party to number actually present as a function of party size.[a]

Party size	Mourning Dove	American Robin	Common Yellow-throat	Field Sparrow	Song Sparrow	Great Crested Fly-catcher	Rufous-sided Towhee	Average	SD
1	0.797	0.734	0.819	0.802	0.683	0.813	0.715	0.77	0.05
2	0.823	0.935	0.995	0.990	0.868	0.941	0.906	0.92	0.06
3	0.926	1.110	1.219	1.169	1.076	1.062	1.062	1.09	0.09
4	1.043	1.158	1.392	1.365	1.137	1.260	1.240	1.23	0.13
5	1.058	1.186	1.339	1.317	1.140	1.260	1.257	1.22	0.10
6	1.070	1.216	1.421	1.319	1.185	1.264	1.326	1.26	0.11

[a] Entries are the means from 8 randomly selected groups of observers, 43 listening periods per observer.

(excluding the 7 species-only periods), divided into 6 groups of 7 or 8 periods each, were: 0.79, 0.75, 0.76, 0.76, 0.75, and 0.75, suggesting a possible decline in efficiency early in the trials but no decline in the later ones. There was considerable peer pressure in our experiments, which may have counteracted the effects of fatigue.

Effects of party size.—Preston (1979) suggested that the change in the maximum number of birds recorded as party size increases might be used to estimate the number of birds actually present. His model rests heavily on the assumption that all birds are equally detectable, a condition we feel is unlikely to be met in practice, but it does raise the issue of whether useful information is contained in the change in numbers reported by different individuals recording simultaneously. Without attempting to answer the question directly, we wish to point out that observers may be recording birds not actually present more commonly than is often supposed. This has a serious impact on any attempt to use the maximum number of birds recorded as party size increases. In our indoor simulation, averaging across all species, parties of three or more tended to overestimate the number of birds present when the maximum number any person recorded was used as the party estimate (Table 6). In all cases, combining the records from a party of four and using the maximum estimate as the "best" estimate led to overestimating the number present, in some cases by as much as 40%. Thus, modeling efforts that imply that there is no overcounting are probably too unrealistic.

CONCLUSION

It would be satisfying to end the study with suggestions for adjusting data to reduce relative bias. Unfortunately, this does not seem possible. Changes in observer efficiency undoubtedly depend heavily upon the number and kind of other species singing, on other noises present, on habitat, and on other features. This study suggests that the change in observer efficiency with density is unlikely to cause errors exceeding 25% for common species or 33% for abundant species, and in many cases the error is surely much smaller. As noted above, relative bias would be 0.0 for a species that never occurs more than singly at a station. Thus, no single recommendation for adjusting estimates would be appropriate for all situations.

Relative bias may be caused by several other factors than the one analyzed here. Average song duration, and the proportion of birds that sing at all, may change with density. Another possibility, often overlooked, is that the survey may be far more sensitive to a change in reproductive success than to a change in density. For example, unmated Mourning Doves sing 10 or more times as much as mated ones (Sayre et al. 1978). Thus, a decline in mating success or an increase in female mortality could lead to an *increase* in survey results. This problem can be avoided if it is possible to insure that the birds in each population being compared are, on average, at the same stage of reproduction. In practice, however, it may be impossible to determine the reproductive stage of the individuals being surveyed.

There is also a difficult problem with interactions among some species. For example, robins often mask other species during the first 30–45 min of daylight. If they declined in density, the number of them reported might not change much, because more distant individuals would be detected, but the reported numbers of less conspicuous species might increase, because they would be more detectable. This may account for Robbins' (1981) report that Northern Orioles (*Icterus galbula*) are recorded only one-half as often during the Breeding Bird Survey's first hour when robins, which have a similar song and are far more abundant, are in full chorus, than in each of the subsequent 3-h.

The general situation might be summarized by the error function,

$$\text{relative bias} = f(e_1, e_2, e_3, e_4, \ldots),$$

where e_1 = change in recorder efficiency with density, e_2 = change in proportion of birds singing with density, e_3 = change in amount of song per bird and consequent change in detectability, and e_4 = change in detectability caused by change in extraneous noises. Clearly, the total bias could be quite large, especially if the factors tend to have the same sign.

For all these reasons, it seems best not to assume that singing-bird surveys yield highly accurate estimates of a change in abundance. They are certainly capable of detecting major changes in range or density (the major purpose of large programs like the Breeding Bird Survey), and they may indicate smaller changes. Nonetheless, in the absence of more information on relative bias than now exists, it seems unwise to carry out standard statistical analyses, such as interval estimation of differences in density, without noting that the relative bias is unknown and may, for abundant species, be as large as 33%.

ACKNOWLEDGMENTS

Numerous skilled ornithologists participated in the study; in the field trials they included T. Cullinan, R. Dolbeer, R. Jezerinac, B. McLean, R. McCullough, K. McGowan, B. Peterjohn, J. Pogacnik, W. Randle, and J. Skinner, and in the indoor trials they were D. Boone, D. Bystrak, P. Bystrak, B. Dowell, G. Dremeaux, S. Duffy, F. Fallon, E. Gizzarelli, R. Greenberg, D. Holmes, M. Howe, C. Jeschke, P. Kalka, K. Klimkiewicz, W. Klockner, P. Leifer, J. Orgain, R. Patterson, R. Ringler, C. Robbins, P. Webb, R. Whitcomb, H. Wierenga, and J. Wilkinson. Technicians on the project included T. Burkhart, K. Cunningham, D. Drake, K. Gilliland, N. Heaslip, K. Komisar, T. Leyes, J. Lupfer, K. Marois, D. Morawetz, J. Skinner, M. Taylor, C. Turik, S. White, and D. Wilson. Assistance on various other phases of the project was provided by R. Arbib and P. Geissler. Particular thanks go to D. Bystrak for recruiting most of the participants in the indoor trials and for helping to conduct them in Maryland. We also thank D. Borror and L. Fairchild for use of the bird recordings and equipment in the Borror Bioaccoustical Laboratory. Helpful comments on the manuscript were made by L. Best, D. Bystrak, D. G. Dawson, R. Dolbeer, C. Robbins, R. S. Stehn, R. Whitcomb, and R. A. Väisänen.

LITERATURE CITED

BART, J. & J. HERRICK. 1984. Diurnal timing of bird surveys. Auk 101: 384-387.

BASKETT, T. S., M. J. ARMBRUSTER, & M. W. SAYRE. 1978. Biological perspectives for the Mourning Dove call-count survey. Trans. North Amer. Wildl. Nat. Res. Conf. 43: 163-180.

BYSTRAK, D. 1981. The North American Breeding Bird Survey. Pp. 34-41 *in* Estimating the numbers of terrestrial birds (C. J. Ralph and J. M. Scott, Eds.). Stud. Avian. Biol. 6.

CARNEY, S. M., & G. A. PETRIDES. 1957. An analysis of variation among participants in pheasant cock-crowing censuses. J. Wildl. Mgmt. 21: 392-397.

CAUGHLEY, G. 1977. Analysis of vertebrate populations. New York, J. Wiley & Sons.

DAWSON, D. G. 1981. Counting birds for a relative measure (index) of density. Pp. 12-16 *in* Estimating the numbers of terrestrial birds (C. J. Ralph and J. M. Scott, Eds.). Stud. Avian Biol. 6.

DESANTE, D. F. 1981. A field test of the variable circular-plot censusing technique in a California coastal scrub breeding bird community. Pp. 177-185 *in* Estimating the numbers of terrestrial birds (C. J. Ralph and J. M. Scott, Eds.). Stud. Avian Biol. 6.

DUKE, G. E. 1966. Reliability of censuses of singing male woodcocks. J. Wildl. Mgmt. 30: 697-707.

FROCHOT, B., D. REUDET, & Y. LERUTH. 1977. A comparison of preliminary results of three census methods applied to the same population of forest birds. Pol. Ecol. Stud. 3: 71-75.

JÄRVINEN, O., & R. A. VÄISÄNEN. 1976. Estimating relative densities of breeding birds by the line transect method. IV. Geographical constancy of the proportion of main belt observations. Ornis. Fennica 53: 87-91.

MAYFIELD, H. F. 1981. Problems in estimating population size through counts of singing males. Pp. 220-224 *in* Estimating the numbers of terrestrial birds (C. J. Ralph and J. M. Scott, Eds.). Stud. Avian Biol. 6.

PRESTON, F. W. 1979. The invisible birds. Ecology 60: 451-454.

REYNOLDS, R. T., J. M. SCOTT, & R. A. NUSSBAUM. 1980. A variable circular-plot method for estimating bird numbers. Condor 82: 309-313.

ROBBINS, C. S. 1981. Effect of time of day on bird activity. Pp. 275-286 *in* Estimating the numbers of terrestrial birds (C. J. Ralph and J. M. Scott, Eds.). Stud. Avian Biol. 6.

ROBEL, R. J., D. J. DICK, & G. F. KRAUSE. 1969. Regression coefficients used to adjust bobwhite quail whistle count data. J. Wildl. Mgmt. 33: 662-668.

RODGERS, R. D. 1981. Factors affecting ruffed grouse drumming counts in southwestern Wisconsin. J. Wildl. Mgmt. 45(2): 409-418.

SANDERSON, G. C. (Ed.). 1977. Management of migratory and upland game birds in North America. Washington, D.C., Intern. Assoc. Fish Wildl. Agencies.

SAYRE, M. W., R. D. ATKINSON, T. S. BASKETT, & G. H. HAAS. 1978. Reappraising factors affecting mourning dove perch cooing. J. Wildl. Mgmt. 33: 662-668.

———, T. S. BASKETT, & K. C. SADLER. 1980. Radiotelemetry studies of the mourning dove in Missouri. Terrestrial Series #9, Missouri Dept. Conservation.

SCOTT, J. M., & F. L. RAMSEY. 1981. Effects of abundant species on the ability of observers to make accurate counts of birds. Auk 98: 610-613.

SCOTT, V. E., & E. L. BOEKER. 1972. An evaluation of wild turkey call counts in Arizona. J. Wildl. Mgmt. 36: 628-630.

WALANKIEWICZ, W. 1977. A comparison of the mapping method and I.P.A. results in Bialowieza National Park. Pol. Ecol. Stud. 3: 119-125.

IMPORTANT NOTICE
CHANGE IN EDITORIAL OFFICES OF *THE AUK*

With the completion of this volume (101) of *The Auk,* John Wiens will retire as Editor. Alan Brush will begin editing the journal with volume 102 (1985). Effective immediately, all authors submitting manuscripts for consideration for possible publication in *The Auk* should send materials directly to the new Editor:

Dr. Alan H. Brush
Editor, *The Auk*
Biological Sciences Group
University of Connecticut
Storrs, Connecticut 06268 USA