GOALS AND STRATEGIES FOR ESTIMATING TRENDS IN LANDBIRD ABUNDANCE

JONATHAN BART,[1] USGS Forest and Rangeland Ecosystem Science Center, Snake River Field Station, 970 Lusk Street, Boise, ID 83706, USA
KENNETH P. BURNHAM, Colorado Cooperative Fish and Wildlife Research Unit, Room 201, Wagar Building, Colorado State University, Fort Collins, CO 80523, USA
ERICA H. DUNN, Canadian Wildlife Service, National Wildlife Research Centre, 100 Gamelin Boulevard, Hull, PQ K1A 0H3, Canada
CHARLES M. FRANCIS, Canadian Wildlife Service, National Wildlife Research Centre, 100 Gamelin Boulevard, Hull, PQ K1A 0H3, Canada
C. JOHN RALPH, Redwood Sciences Lab, U.S. Forest Service, 1700 Bayview Drive, Arcata, CA 95521, USA

Abstract: Reliable estimates of trends in population size are critical to effective management of landbirds. We propose a standard for considering that landbird populations are adequately monitored: 80% power to detect a 50% decline occurring within 20 years, using a 2-tailed test and a significance level of 0.10, and incorporating effects of potential bias. Our standard also requires that at least two-thirds of the target region be covered by the monitoring program. We recommend that the standard be achieved for species' entire ranges or for any area one-third the size of the temperate portions of Canada and the United States, whichever is smaller. We applied our approach to North American Breeding Bird Survey (BBS) data. At present, potential annual bias for the BBS is estimated at ±0.008. Further, the BBS achieves the monitoring standard for only about 42% of landbirds for which the BBS is considered the most effective monitoring approach. Achieving the proposed monitoring target for ≥80% of these species would require increasing the number of BBS—or similar survey—routes by several-fold, a goal that probably is impractical. We suggest several methods for reducing potential bias and argue that if our methods are implemented, potential bias would fall to ±0.003. The required number of BBS or similar routes would then be 5,106, about 40% more than in the current BBS program. Most of the needed increases are in 15 states or provinces. Developing a comprehensive landbird monitoring program will require increased support for coordination of the BBS (currently 2 people) and new programs for species that are poorly covered at present. Our results provide a quantitative goal for long-term landbird monitoring and identify the sample sizes needed, within each state and province, to achieve the monitoring goal for most of the roughly 300 landbird species that are well suited to monitoring with the BBS and similar surveys.

JOURNAL OF WILDLIFE MANAGEMENT 68(3):611–626

Key words: birds, monitoring, North American Breeding Bird Survey, population trends, surveys, trend estimation.

Population size monitoring has provided critical information for nearly all major—and thousands of minor—wildlife issues during the past several decades. Examples in which knowledge of trends in population size have been critical to the success of management programs include (1) identification of pesticides as a serious threat to wildlife (Carson 1962); (2) recovery of species from pesticide impacts (Sheail 1985); (3) declines of individual species such as spotted owls (*Strix occidentalis*; Gutiérrez et al. 1995) and sage grouse (*Centrocercus urophasianus*; Schroeder et al. 1999); (4) declines of groups of species such as those in eastern thickets, grasslands, and western riparian habitats (Askins 2000); and (5) the recovery of species under management such as the peregrine falcon (*Falco pereginus*; White et al. 2002), Kirtland's warblers (*Dendroica kirtlandii*; Mayfield 1992), and many waterfowl (Williams et al. 2002).

Despite the success of many past monitoring programs, much opportunity exists for making programs even more useful (Downes et al. 2000, Williams et al. 2002). We propose a quantitative goal for landbird monitoring programs and describe strategies for achieving the goal for North American landbirds. We believe that a quantitative goal is needed to design a comprehensive program, identify needed resources, and measure progress. Our work is based on Butcher et al. (1993), who proposed a monitoring goal, and on recent work by Partners in Flight (Downes et al. 2000, Pashley et al. 2000).

Substantial literature (summarized by Williams et al. 2002) discusses the importance of relating monitoring methods to management goals. For example, quantitative models may be used to assess costs and probabilities of different management errors and to identify optimal survey methods and sample sizes in accordance with the results. We are concerned with multispecies landbird surveys, and the integration of monitoring

[1] E-mail: jbart@eagle.boisestate.edu

and management on such a broad scale is necessarily incomplete. We have identified broad management goals, such as detecting declines, and we believe that surveys can and should be designed to help achieve these goals. No specific management programs can be envisaged, however, because the surveys cover hundreds of species and management actions will not—and cannot—be designed until declines, and their causes, have been identified.

We addressed 3 questions: (1) What should the accuracy target be in programs to estimate trends in landbird population size? (2) Which landbird species and populations in North America warrant coverage by monitoring programs, and which of these species are best monitored by the BBS (Sauer et al. 2001) and similar programs? (The phrase "and similar programs" means multispecies landbird surveys, using point counts conducted at about the same time that the BBS is conducted, which can be used to supplement the BBS). (3) How many of the populations best monitored by the BBS are adequately covered now, and how many routes would be needed to provide adequate coverage for most of these populations? We also briefly discuss other measures needed to develop a comprehensive program for landbird monitoring.

METHODS

Expression for Power

Butcher et al. (1993) suggested that a reasonable accuracy target for trend-monitoring programs is 80% power to detect a 50% decline occurring within 20 years. We evaluated this suggestion by deriving an expression for power, estimating its components, and exploring alternatives to the recommendation by Butcher et al. (1993). Power usually is calculated under the assumption that bias is zero. With trend estimates, however, this is often not a reasonable assumption (Williams et al. 2002). One approach for incorporating potential bias in the analysis is to establish lower and upper limits (i.e., b_l and b_u) for potential bias and to use these in deciding whether an observed trend provides a reliable indication of the direction of change in the population. For example, if bias in the estimated annual rate of change is assumed to be as much as ±1%, then $b_l = -0.01$ and $b_u = 0.01$. If the estimated annual rate of change in population size is <1, we only conclude that the population has declined if the observed decline is significantly less than $1 + b_l$ (0.99 in the example). If the trend estimate is >1, we only conclude that the population has increased if the observed increase is significantly greater than $1 + b_u$ (1.01 in the example). Incorporation of potential bias into the analysis in this manner affects power to detect a change. We derived an expression for power that incorporates limits for potential bias (Appendix A). We present the expression in terms of the required sample size (n; e.g., BBS routes), with simple random selection of routes, to ensure that power is at least $1 - \beta$. The equation is

$$n = \left(\frac{sd(Z_{\alpha/2} + Z_{\beta})}{|(C+1)^{1/d} - 1| - (b_u - b_l)} \right)^2, \quad (1)$$

where sd is the sample standard deviation of within-route trends, α is the level of significance, d is the duration of the survey in years, and C is the change in population size during the d years (e.g., for a 40% decline, $C = -0.4$).

Three Major Questions

Question (1): What should the accuracy target be in programs to estimate trends in landbird population size?—We addressed this question using BBS data. We estimated sd in Expression (1) for $d = 5$, 10, 15, and 20 years, for each of the 133 species recorded on >500 BBS routes during 1982–2001. For $d < 20$, we estimated sd for all sequential series of d years and then calculated the average of these values for each species. Periods that are too short may yield misleading estimates of the long-term trend due to cycles in the counts about the long-term trend. We investigated this issue by calculating the 85, 95, and 99% confidence intervals (CI) for each trend. We used a version of the trend estimation method described in Bart et al. (2003) suitable for stratified sampling and recorded the frequency with which these CIs included the trend estimated from the 20-year data set. We restricted the investigation to 2-tailed procedures on the assumption that detecting increases, as well as decreases, often is important. Reasonable values for C were identified by determining the frequency distribution of 20-year trends for the 133 selected species. We established values for b_l and b_u by reviewing studies that identified sources of potential bias and estimating the magnitude of these biases. We then assessed the degree to which different sources would be additive or complementary.

In many long-term monitoring programs, large, continuous portions of the area of interest lay outside the sampling frame (i.e., the set of locations that might be surveyed). For example, the range of many North American species extends into the boreal region, where few BBS routes are conducted. We refer to this situation as incomplete "coverage." We were unable to identify a completely realistic model to explore the likely magnitude of bias with incomplete coverage, but we derived an approximate expression (Appendix A). We analyzed BBS data from the 6 U.S. Fish and Wildlife Service (USFWS) regions that together cover the coterminous United States. We assumed that abundance for a given species was equal across regions, and that all or none of each region was covered by the surveys. In this case, the probability that bias exceeds any given level G, is approximately

$$P(bias > G) = P\left(Z > \frac{n_1 G}{\sigma^3}\right), \tag{2}$$

where Z is a standard normal variable, n_1 is the number of USFWS regions covered by the survey, and σ^2 is the variance of the true trends among the 6 regions. We derived an expression for σ^2 that separated and removed sampling error (Appendix A) and estimated its value for each of the 133 selected species. We then plotted bias as a function of how much of the range was covered, and we used the results to develop a guideline for how much of the study region needs to be covered by the surveys to avoid substantial bias.

Question (2): Which landbird species and populations in North America warrant coverage by monitoring programs, and which species are best monitored by the BBS and similar programs?—We identified landbirds that warrant monitoring using the general principle that we should monitor species we would try to conserve if we knew they were declining or changing in other undesirable ways. According to this principle, we would ideally monitor all landbird species that occur regularly in North America, but we would not attempt to monitor species that only occur rarely in North America. An initial list was prepared that included all landbird species with range maps in a popular field guide (National Geographic Society 1999). A committee of landbird specialists, appointed by the Partners in Flight Monitoring Working Group, then revised the list in accordance with the general principle above. Many of these species have substantial breeding populations in both temperate and northern (boreal and arctic) regions. Different survey methods must be used in boreal areas. We therefore distinguished separate temperate and northern populations for species that breed in both areas. Based on results from our analysis of Question (1), we defined separate temperate and northern populations if the temperate and northern regions each contained >33% of the species' range.

Many species that do breed in temperate regions are not well suited to monitoring with the BBS. We identified these species by defining 6 landbird survey methods and identifying species that will be best monitored using the BBS and similar—rather than other—methods. This analysis produced a list of species that warrant monitoring and an identified subset of these species that will be best monitored with the BBS and similar programs.

Question (3): How many of the populations best monitored by the BBS are adequately covered now, and how many routes would be needed to provide adequate coverage for most of these populations?—The proposed monitoring target was expressed as the standard error of the estimated trend. Trend estimates are often needed for smaller areas than the entire range, especially for species with wide ranges. We considered different sized areas and concluded that meeting the accuracy target for regions one-third the size of the temperate regions of Canada and the United States represented a reasonable trade-off between the difficulty of obtaining estimates for small areas and the need for estimates within parts of the range. We obtained rangewide estimated standard errors for population trends for each landbird from Sauer et al. (2001). Next, we converted these to the standard errors that would have been achieved with a survey area one-third the size of the temperate region of Canada and the United States (unless the species range was smaller than this area). Finally, we recorded whether the resulting standard error was less than or greater than the threshold standard error required to achieve the proposed monitoring objective. This process produced a list of landbird species that currently are adequately monitored by the BBS. We also reported the number of rangewide estimates that met the proposed accuracy target to provide an indication of the extra cost of meeting the accuracy target at the smaller spatial scale.

Using Expression (1), we estimated how many additional BBS or similar routes would be needed to achieve the monitoring objective for most landbirds best monitored by the BBS. The term *sd* in Expression (1) varies among species, and we did not have sufficient data to estimate this quan-

Table 1. Frequency distribution for standard deviation (*sd*) based on 133 species well surveyed by the North American Breeding Bird Survey.[a]

sd	No. of species	Proportion	Cumulative proportion
<0.060	2	0.01	0.01
0.059–0.080	19	0.14	0.15
0.081–0.140	85	0.65	0.80
0.141–0.200	18	0.13	0.93
>0.200	9	0.07	1.00

[a] Data collected during 1980–1999 were used to estimate *sd*.

tity for most species. We therefore calculated values of *sd* for a sample of well-surveyed species and used the eightieth quantile as the value for *sd*. This gave us a conservative value for *sd*. The other terms in Expression (1) were specified in the proposed monitoring goal and did not vary among species. We were thus able to calculate a single, estimated sample size (no. of BBS or similar routes) needed to achieve the accuracy target.

For the i^{th} species, the density (y_i) of routes needed to achieve the monitoring objective may be expressed as

$$y_i = \frac{n}{A_i p_i}, \quad (3)$$

where n is the needed number of routes (not species specific); A_i is the area covered by the range of species i, or the portion of the range for which the trend estimate is obtained; and p_i is the proportion of routes, within the area A_i, on which the i^{th} species is recorded at least 4 times (the minimum for estimating trends recommended by Bart et al. 2003). We selected 14 species with small to medium-sized ranges, estimated the size of each species' breeding range using maps provided by Project WILDSPACE (Welsh et al. 1999), and determined p_i by analyzing BBS data using the trend estimation method of Bart et al. (2003). We then used Expression (3) to estimate the needed density of routes for each of these species with A_i equal to the species' entire range or one-third the temperate region of Canada and the United States (approx 3,000,000 km^2), whichever was smaller. Achieving the accuracy target for all species using only the BBS and similar programs would be difficult and probably is not a wise use of resources. We selected a threshold value (y) as the y_i such that the monitoring objective would be achieved for 80% of the species best monitored using the BBS and similar programs. This density of routes was multiplied by the area of each province and state to obtain the target, minimum number of BBS or similar routes. We summed province- and state-specific values to obtain the needed total number of routes.

Table 2. Frequency distribution of trends during 1980–1999 for 133 landbird species well surveyed by the North American Breeding Bird Survey.

Change[a] (%)	No. of species	Proportion	Cumulative proportion
<25	77	0.58	0.58
26–50	37	0.28	0.86
51–75	14	0.11	0.96
>75	5	0.04	1.00

[a] Estimated increase or decrease during 1980–1999.

RESULTS AND DISCUSSION

Question (1): What should the accuracy target be in programs to estimate trend in landbird population size?

With d = 20 years, 80% of the values of *sd* were ≤0.14 and 93% were ≤0.20 (Table 1). Approximately 86% of the values of C were ≤50%, and 96% of the values were ≤75% (Table 2). In choosing the specifications for power, examining the trade-off between probabilities of Type I and Type II errors may be useful. Type I errors occur when a true null hypothesis is rejected ("false alarm" errors); Type II errors occur when a false null hypothesis is not rejected (declines are missed). Managers can protect themselves against 1 type of error by increasing the risk of the other error. For example, suppose the conditions described in Table 3 applied (d = 20, sd = 0.14, $b_l = b_u = 0$, n = 100), and that managers were equally concerned with a false alarm and with not detecting a 50% decline. Setting the significance level at 0.15 would result in these 2 errors being about equally likely. If a false alarm was viewed as more serious than missing a decline, then the sig-

Table 3. Probabilities of Type I (C = 0) and Type II (C > 0) errors in relation to level of significance for detecting population declines. Standard deviations (*sd*) estimated from data on landbirds collected in the North American Breeding Bird Survey from 1980 to 1999.[a]

Significance level	20-yr decline (*C*) in survey results: 0%	25%	50%	75%
0.05	0.05	0.83	0.32	0.00
0.10	0.10	0.73	0.22	0.00
0.15	0.15	0.66	0.16	0.00
0.20	0.20	0.60	0.13	0.00

[a] From $n = \left(\frac{sd(Z_{\alpha/2} + Z_\beta)}{|(C+1)^{1/d} - 1| - (b_u - b_l)}\right)^2$ with d = 20, sd = 0.14, $b_l = b_u = 0$, and n = 100.

Table 4. Probabilities of Type I ($C = 0$) and Type II ($C > 0$) errors in relation to number of routes (n) for detecting declines in landbird species. Standard errors (se) estimated using North American Breeding Bird Survey data from 1980 to 1999.[a]

No. of	20-yr decline (C) in survey results			
routes	0%	25%	50%	75%
100	0.10	0.82	0.48	0.04
200	0.10	0.74	0.22	0.00
300	0.10	0.66	0.10	0.00
400	0.10	0.59	0.04	0.00

[a] From $n = \left(\frac{sd(Z_{\alpha/2} + Z_{\beta})}{|(C+1)^{1/d} - 1| - (b_u - b_l)}\right)^2$ with $d = 20$, $sd = 0.20$, $b_l = b_u = 0$, and $\alpha = 0.10$.

Table 6. Samples sizes required for 80% power to detect a 50% population decline in relation to standard deviation (sd) of the trend; sd estimated using North American Breeding Bird Survey data form 1980 to 1999.[a]

	Significance level (α)			
sd	0.05	0.10	0.15	0.20
0.08	43	34	29	25
0.14	133	104	88	76
0.20	271	213	179	156
0.30	609	480	404	350

[a] From $n = \left(\frac{sd(Z_{\alpha/2} + Z_{\beta})}{|(C+1)^{1/d} - 1| - (b_u - b_l)}\right)^2$ with $d = 20$ and $b_l = b_u = 0$. Note that n increases as sd^2.

nificance level might be set at 0.10 or 0.05. Note, however, that with the 0.05 level, the chance of missing a 50% decline is approximately 1 in 3 so the survey would not be very effective in identifying declining species.

Sample size has a major effect on the relative probabilities of Type I and Type II errors (Table 4). As the number of routes increases from 100 to 400, the probability of missing a 50% decline declines from 48 to 4%.

The value of *sd* also has a substantial influence on the probability of Type I and Type II errors. For example, under the conditions described in Table 5 ($d = 20$, $n = 300$, $b_l = b_u = 0$, $\alpha = 0.05$), the probability of missing a 50% decline increases from 1% with $sd = 0.08$ to 80% with $sd = 0.30$. Sample sizes required to achieve 80% power to detect a 50% decline increase rapidly with increasing *sd* (Table 6). Values of *sd* as low as 0.08 were rare in the BBS dataset, so structuring a goal for power around this level would not be useful. Similarly, attempting to achieve high power for all species ($sd = 0.30$) would be extremely difficult.

Table 5. Probabilities of Type I ($C = 0$) and Type II ($C > 0$) errors in relation to the standard deviation (sd) of the trends as estimated for landbird species using North American Breeding Bird Survey (BBS) data from 1980 to 1999.[a]

	Proportion of species with smaller	20-yr decline (C) in survey results			
sd	values of sd[b]	0%	25%	50%	75%
0.08	0.15	0.05	0.57	0.01	0.00
0.14	0.05	0.05	0.83	0.32	0.00
0.20	0.05	0.05	0.89	0.60	0.08
0.30	0.99	0.05	0.93	0.80	0.39

[a] From $n = \left(\frac{sd(Z_{\alpha/2} + Z_{\beta})}{|(C+1)^{1/d} - 1| - (b_u - b_l)}\right)^2$ with $d = 20$, $n = 300$, $b_l = b_u = 0$, and $\alpha = 0.05$.

[b] Based on 133 species considered to be well studied by BBS.

The change to be detected also has a major influence on required samples sizes (Table 7). For example, under the conditions in Table 7 ($d = 20$, $sd = 0.20$, $b_l = b_u = 0$, $\beta = 0.20$), nearly 6 times the sample size is required to detect a 25% change as to detect a 50% change.

The length of time during which a given change (e.g., 50%) occurs also affects required sample sizes. In Expression (1), for a given value of C, as d becomes smaller, the term $|(C + 1)^{1/d} - 1|$ increases, which tends to reduce the required sample size. For example, with $C = -0.50$, $|(C + 1)^{1/d} - 1| = 0.034$ with $d = 20$ but is 0.045 with $d = 15$, an increase of 32%. However, less data are collected with a shorter duration, and this tends to increase the *sd* and offset the reduction in $|(C + 1)^{1/d} - 1|$. The average increase in the *sd* for a 15-year survey, compared to a 20-year survey, was 16% and for a 10-year survey was 33% (Table 8). The net effect of these 2 changes is that for fixed values of C, the required sample size declines as duration decreases. For example, with $sd = 0.20$, significance = 0.10, and $C = -0.50$, the sample size required for power of 80% is 213 for $d = 20$ and 166 for $d = 15$, a decline of 22%. This tendency

Table 7. Sample sizes required for 80% power to detect population declines of 25–75% in relation to level of significance. Standard deviations (sd) estimated using data on landbird species in the North American Breeding Bird Survey from 1980 to 1999.[a]

Significance	20-yr decline (C) in survey results		
level	25%	50%	75%
0.05	1,540	270	70
0.10	1,212	213	55
0.15	1,022	180	46
0.20	885	155	40

[a] From $n = \left(\frac{sd(Z_{\alpha/2} + Z_{\beta})}{|(C+1)^{1/d} - 1| - (b_u - b_l)}\right)^2$ with $d = 20$, $sd = 0.20$, $b_l = b_u = 0$, and $\beta = 0.20$.

Table 8. Relative standard deviation (*sd*) and confidence interval (CI) coverage as a function of survey duration (*d*; in years). Analysis based on data for landbird species in the North American Breeding Bird Survey data from 1980 to 1999.

d	Relative *sd* (sd_d/sd_{20})	Proportion of CIs that included the 20-yr trend		
		0.85 CI	0.95 CI	0.99 CI
5	2.15	0.48	0.63	0.74
10	1.33	0.52	0.66	0.77
15	1.16	0.71	0.81	0.90

also means that a smaller change can be detected with a given sample size when the survey period is shorter. For example, with $sd = 0.20$, significance = 0.10, $d = 20$, and $n = 213$, the power to detect a 50% decline would be 80%. With $d = 15$, and assuming *sd* was 0.234 (a 16% increase), power would be 80% to detect a decline of 45.5%. Thus, for a given sample size, level of significance, and power, a slightly smaller decline can be detected if the decline occurs within 15 years rather than within 20 years. Also, a given decline will be detected with higher power if it occurs in fewer years.

Reducing the duration of a survey increases the risk that the estimated trend during the survey period will be a poor estimator of the 20-year trend due to cycles. This risk rises sharply as duration decreases (Table 8). For example, only 71% of the calculated 85% CIs (based on 15 years), included the 20-year trend, and only 52% of the calculated 85% CIs (based on 10 years), included the 20-year trend. Thus, trends based on 5 or 10 years were poor predictors of the 20-year trend, and even 15-year trends often were quite different from 20-year trends.

Guidelines.—Selecting a reasonable target for power depends on making several choices. We suggest the following guidelines:

(1) The duration of the survey should be 20 years, in part because trends during shorter periods often are poor estimators of the longer trend due to cycles (Table 8). Further, the gain achieved by using a shorter period is small (e.g., 80% power to detect a decline of 45.5% rather than 50% in the previous example), and large (e.g., 50%) declines seldom occur in <20 years.

(2) We should assume the *sd* will be about 0.14. Approximately 80% of the species had smaller values of *sd* (Table 6). Thus, a sample size that achieves the accuracy target with $sd = 0.14$ will achieve the target for most of the species. Achieving the target for ≥90% of the species, while desirable, approximately doubles the required sample size.

(3) The goal of the survey should be to detect a decline of about 50%. Detecting smaller declines (e.g., 25%) with power high enough to be useful would be extremely expensive (Table 6), and detecting these declines could only be justified if explicit management objectives required that level of precision and could justify the costs. If the goal of detecting a 50% decline with 80% probability is achieved, then we will have even higher power to detect larger declines.

Potential Bias.—We have placed much emphasis in this analysis on acknowledging bias. Bias usually is difficult to estimate rigorously and often has been ignored in the past. We believe, however, that analyses of avian trend data should include an explicit discussion of bias and that an upper limit for its effects on accuracy should be established. Our rationale is that any use of trend (or any other) estimates requires an assumption about bias.

A recent review of the BBS (O'Connor et al. 2000) identified 3 major potential sources of bias: differences between regionwide and roadside population trends, changes in observer detection rates, and bias due to analytic methods. We discuss each of these sources of bias.

Two studies have compared regionwide and roadside trends, both using change in habitat as a surrogate for avian population change. Bart et al. (1995) studied change in proportion of an Ohio, USA, landscape covered by forest and found little difference between the roadside and regionwide trends. Further, Bart et al. (1995) found no significant differences between the annual rates of change in forest cover regionwide and within 280 m of roads, suggesting that (in this study area) bias due to restricting surveys to roadsides was probably was <0.005.

Keller and Scallan (1999) studied regionwide and roadside changes between 1963 and 1988 in 6 habitats and 12 habitat features (e.g., single family homes) in Ohio and Maryland, USA. In Ohio, none of the trends in habitats differed significantly (with $\alpha = 0.05$) between the on-road and off-road study areas. In Maryland, urban habitat increased significantly faster along roads (annual rate = 0.069) than off roads (0.039), and agricultural habitats decreased faster along roads (0.014 vs. 0.004). Of the habitat features measured, houses, buildings, and associated features (e.g., driveways) increased significantly faster along roads in both Ohio and Maryland than off roads, but no other differences were significant. The lack of significant results for most habitats and habitat features suggests that bias would be small except for birds dependent on human developments. Thus, 2 studies that estimated the bias in trend estimates due to surveys being restricted to roadsides both found annual bias probably was <0.005 for most species. While more studies of this sort are needed and cases with larger bias can undoubtedly be found, absolute bias from this source presently appears to be <0.005.

Change in the skill of BBS observers was studied by Sauer et al. (1994) and James et al. (1996). In both studies, the investigators concluded that a trend has occurred in average observer detection rates. This finding led to the development of methods that remove certain trends in detection rates (Link and Sauer 1994). Link and Sauer (1994) noted that their model assumes that observer-specific detection rates do not change through time, whereas evidence exists for both a learning effect (numbers detected in the first year tend to be lower than numbers detected in subsequent years [Kendall et al. 1996]) and for senescence (numbers decline as observers' hearing and perhaps vision decline). Both problems are potentially serious. For example, if the population was stable, detection rates were 20% lower the first year, and observers collected data for 7 years, then the estimated annual rate of change would be 1.02 and bias would be 0.02, a figure high enough to cause serious errors in estimates. Bias caused by observer senescence would equal the long-term trend in the detection rate for all observers, which also could be substantial if senescence is affecting a large fraction of the observers. Both sources of bias could be reduced by training and evaluation programs and perhaps by developing models that incorporate estimates of their effects. Thus, while these problems potentially are serious at present, they probably can be reduced to low levels (e.g., absolute bias <0.003) with the implementation of improved analytic methods.

Bias due to analytic methods can be divided into 2 sources: statistical bias that exists even if all assumptions of the model are met, and bias due to using a model whose assumptions are not fully met by the data. Link and Sauer (1994) studied statistical bias in a route regression approach. Bias increased as trends diverged from 1.0 and with decreasing abundance. Bias exceeded 0.005 only for positive trends and for species recorded on average less than once per BBS route. Statistical bias thus appears to be quite small for this approach.

The effects of model failure are harder to assess. Link and Sauer (1994) pointed out that their model assumed constant within-observer detection rates, but this assumption may be false. They also identified nonlinear trends within routes and weighting factors as possible additional sources of bias. These factors deserve additional study (in part to increase precision), but they should not produce substantial bias. Bart et al. (2003) studied bias in a "linear method" for trend estimation and in the route regression approach of Link and Sauer (1994) by using real datasets to establish actual trends and estimating these trends by sampling with replacement (which simulates a large population having the same actual trend as the sample). Bart et al. (2003) studied 2 datasets, 1 from the BBS and 1 from the International Shorebird Survey (Brown et al. 2001). With the BBS data, bias was negligible (<0.002) with both methods (Bart et al. 2003). With the shorebird dataset, bias was negligible with the linear method but was substantial (exceeding 0.01 for some species) with the route regression approach, probably because of large variation in numbers recorded within sites in the shorebird dataset (Bart et al. 2003). The general conclusion from these studies is that analytic methods with negligible bias due to the estimation methods probably can be found for most datasets.

Another source of bias in the trend estimate that probably has been of some significance in the past—and may become much more significant in the future—is change in phenology. Hundreds of studies during the past decade have indicated that phenologies of many species are changing (reviewed in Parmesan and Yohe 2003). In general, species are starting their breeding season earlier, though sometimes the reverse is true. Changes in phenology change the frequency distribution of detection rates since most birds have very different detection rates at different stages in their nesting period. While this effect could cause a substantial bias if global changes continue to accelerate, adjusting the timing of counts or estimating the change due to phenology should be possible based on intensive surveys at a subset of locations (e.g., participating national wildlife refuges). A related source of bias is change in survey dates (regardless of change in phenology). Little investigation has addressed whether the distribution of dates has been stable throughout the history of the BBS.

In summary, more work is needed to establish upper limits for bias in trend estimates based on the BBS and other surveys. At present, however, the few studies that have been completed suggest that the bias probably will be small in most cases if careful attention is given to discovering and

Table 9. Influence of unacknowledged bias on the probability of a Type I error (rejecting the null hypothesis when it is true).[a]

|Bias|	Probability			
	Assumed	Actual	Assumed	Actual
0.000	0.05	0.050	0.10	0.100
0.005	0.05	0.079	0.10	0.142
0.010	0.05	0.170	0.10	0.264

[a] From
$P(\text{Type I error} \mid b_l, b_u \text{ undefined}) = P\left(Z < -X_{\alpha/2} - \frac{B}{se(r)}\right) + P\left(Z < -X_{\alpha/2} - \frac{B}{se(r)}\right)$, with $sd = 0.20$ and $n = 400$.

Table 10. Influence of acknowledging bias on power to detect a decline[a] in landbird species.

$(b_u - b_l)$	Decline		
	25%	50%	75%
0.000	0.53	1.00	1.00
0.004	0.31	0.99	1.00
0.008	0.14	0.96	1.00
0.016	0.01	0.73	1.00
0.024	0.00	0.30	1.00

[a] Probability of obtaining a significant estimated trend that is <1. From
$\text{Power with } b_l, b_u \text{ defined} = P\left(Z > Z_{\alpha/2} - \frac{|R-1| - |B - b_P|}{se(r)}\right)$, with $d = 20$, $sd = 0.14$, $n = 400$, and $\alpha = 0.05$.

reducing sources of potential bias. We can reasonably assume that restricting the surveys to roadsides, unacknowledged change in observer skill, and bias due to the analytic method might each contribute bias of up to 0.005. However, the direction of the bias should not be the same in all cases, so these biases should cancel out each other to some extent. This suggests that a reasonable estimate for the upper limit of annual bias in BBS estimates of trend at present is ±0.008. This estimate is admittedly crude, and bias will obviously differ between species, but we feel that making an estimate and using it when establishing the monitoring goal is better than ignoring the issue, which amounts to assuming that bias is zero—a value even less defensible than ±0.008. We therefore suggest using ±0.008 as the limits for bias in BBS estimates until better information leads to new, perhaps species-specific, values. If the bias-reduction methods we identified are implemented, then we can reasonably assume that absolute bias could be reduced to <0.003, and perhaps to even lower levels.

The presence of bias also affects tests, confidence intervals, and power. Unacknowledged bias has a major effect on tests when the null hypothesis is true (i.e., the population is stable) over the study period. Bias tends to make rejecting the null hypothesis much more likely than the nominal significance level. For example, if the significance level is set at 0.05 but bias is ±0.01, then the actual probability of committing a Type I error is 0.17, more than 3 times the nominal rate (Table 9). Power to detect a decline depends on whether the bias is negative or positive. Negative bias makes rejecting the null hypothesis more likely because the effects of the decline are exaggerated. With positive bias, however, power is lessened because the bias tends to mask the decline. These examples illustrate why we recommend establishing limits for bias and including the potential bias in tests, and therefore in the power calculations.

Acknowledging potential bias reduces power. For example, under the conditions in Table 10, power to detect a 50% decline falls from essentially 100% when potential bias is 0.0 to only 73% when $b_u - b_l = 0.016$, as occurs if $b_l = -0.008$ and $b_u = 0.008$, the limits we suggested for BBS data. If $b_u - b_l = 0.024$, power is only 30% to detect a 50% decline.

Combinations of the variables in Expression (1) that would result in 80% power to detect a 50% decline during 20 years are shown in Table 11. Reducing potential bias from ±0.008 to ±0.003 reduces required sample sizes by about 58%. Thus, if this reduction could be achieved by allocating some resources to the bias reduction, higher power would be achieved as long as sample size was not reduced more than 58%. Some of the bias reduction methods involve only the analysis (e.g., eliminating first-year effects and using habitat-based models to reduce roadside bias), and thus should have low costs. Other methods, such as double sampling to estimate detection rates, will reduce sample size (assuming resources are fixed), and their feasibility is unknown at present. Nonetheless, reducing bias to ±0.003 seems at least possible.

Table 11. Sample size required to achieve 80% power to detect a 50% population decline occurring within 20 years, in relation to standard deviation (*sd*) of the trends, potential bias, and significance level. Values used for *sd* were identified by analyzing landbird data in the North American Breeding Bird Survey from 1980 to 1999.

Standard deviation (*sd*)	Potential bias (b_u, b_l)	Significance level	
		0.10	0.15
0.14	–0.008, 0.008	370	313
0.14	–0.003, 0.003	154	130
0.20	–0.008, 0.008	757	638
0.20	–0.003, 0.003	314	265

Table 12. Probability that bias of an estimated rangewide trend for a landbird species, surveyed by the North American Breeding Bird Survey, exceeds 0.005, 0.01, 0.02, and 0.03 when portions of the range are not covered by the survey.

No. of regions used[a]	Proportion of range excluded	Bias			
		0.005	0.01	0.02	0.03
5	0.17	0.30	0.04	<0.01	<0.01
4	0.33	0.42	0.10	<0.01	<0.01
3	0.50	0.54	0.22	0.01	<0.01
2	0.67	0.68	0.42	0.10	0.01
1	0.83	0.84	0.68	0.42	0.22

[a] Total number of U.S. Fish and Wildlife Service regions = 6.

We did not consider a significance level of 0.05 on the basis that false alarm mistakes then become much less common than failing to detect declines, which seems inappropriate. Even with a significance level of 0.1 and potential bias of ±0.003, the actual probability of a false alarm error is about 0.06 (range = 0.056–0.065 depending on actual bias), well below the nominal level. With the level of significance set at 0.15 the actual probability of a false alarm error is about 0.09, still much smaller than the probability of missing a decline (0.20). This result might suggest that the level of significance be set at 0.15; however, setting the significance level at 0.10 is much more common and we prefer 0.10 simply for that reason. The proposed accuracy standard is thus 80% power to detect a 50% decline within 20 years, acknowledging potential bias, setting the level of significance at 0.10, and using a 2-tailed test. If the suggested limits for potential bias with the BBS (±0.008) are accepted, then up to 370 routes (Table 11) with sufficient data to estimate trends are required for species with $sd \leq 0.14$ (80% of the species in our sample), and up to 757 routes are required for species with $sd \leq 0.20$ (93% of the species in our sample). If potential bias can be reduced to ±0.003, then about 150 routes are needed to achieve the accuracy target for approximately 80% of the focal species.

Proportion of the Study Area Covered.—The average estimate of the variance in true trends among USFWS regions for numerous species was 0.0243. We solved Expression (2) for a range of values for G and n_1 (Table 12). This analysis is approximate for several reasons, especially because it does not distinguish between species. Some species probably show little variation in trend across regions, whereas in other species, the variation probably is substantial. At present, however, little basis exists for predicting the variation for a particular species other than using methods that predict the average variation. A practical guideline for how much of a species' range should be covered by surveys might be constructed as follows. Our interpretation is that bias as small as 0.005 is tolerable (Tables 10, 11). As bias rises above 0.01, however, a much larger sample size is needed for high confidence of achieving the accuracy target. If the surveys cover at least two-thirds of the range (rows 1 and 2 in Table 12), then the probability of bias >0.01 is <0.10. If only half the range is covered, then the probability of bias >0.01 rises to 0.22. Thus, while covering all of the range clearly is desirable, missing up to a one-third of the range does not appear critical.

In theory, one could estimate the potential bias due to incomplete coverage for each species and add this component to other sources of bias. Given the uncertainties we described, however, we feel that a simpler rule may be more helpful. We suggest that species for which survey coverage is less than two-thirds of the range, or other area of interest, be considered inadequately monitored regardless of how small the estimated standard error of the trend is. As we noted earlier, areas not covered are large, *continuous* portions of the range.

Our proposed accuracy standard is thus that surveys cover at least two-thirds of the region of interest and that they achieve 80% power to detect a 50% decline within 20 years, using a 2-tailed test, setting the level of significance at 0.10, and acknowledging potential bias. Adopting this standard means that in Expression (1), $d = 20$, $Z_{\alpha/2} = 1.645$, $Z_{\beta} = 0.84$, and $C = -0.5$. For BBS data, as noted earlier, $sd = 0.14$, $b_l = -0.008$, and $b_u = 0.008$ are reasonable values.

The accuracy target may need modification for some landbird species. For example, some managers or researchers may feel that smaller declines need to be detectable for long-lived species because these species have less ability to recover. Care would be needed when applying our goal to species that suffer large, natural declines after harsh winters. Additionally, defining range for colonial species is somewhat arbitrary but could affect whether surveys covered two-thirds of a species' range. Another caveat is that the standard error must be reliable. Two colonies might have very similar trends, and the standard error of the overall trend might thus be very small. However, a sample of 2 colonies generally is too small to use in estimating a species-wide trend. Thus, attempts to use our monitoring goal in applied situations may show that our goal needs revision for some species. We suspect, however, that our goal will be appropriate for most species and that it can be modified fairly easily when necessary.

Question (2): Which landbird species and populations in North American warrant coverage by monitoring programs, and which species are best monitored by the BBS and similar programs?

The initial list of landbird species that warrant monitoring had 425 species. Review by the committee of landbird specialists resulted in relatively few changes. The final list had 431 species. The breeding ranges of 264 species are largely (>67%) within the temperate region, 39 of the species' ranges are primarily in the northern (arctic and boreal) regions, and 128 of the species' ranges are in both the temperate and northern regions. The total number of populations was 559 (Table 13). The committee identified 297 populations (53% of all the landbird populations; 76% of the temperate, landbird populations) that they believed can be best monitored with the BBS and similar programs.

Table 13. Populations of North American landbirds judged to warrant monitoring programs.

Description	Number	Percent
Temperate-nesting populations	392	70
Best monitored by the BBS[a] and similar programs	297	53
Other populations[b]	95	17
Northern-nesting populations	167	30
Total	559	

[a] North American Breeding Bird Survey.
[b] Raptors (24), nocturnal species (24), hummingbirds (14), uplands game birds (11), and others (largely southwestern species, 22).

Question (3): How many of the populations best monitored by the BBS are adequately covered now, and how many routes would be needed to provide adequate coverage for most of these populations?

In Expression (1), which is based on a simple random sample, the standard error of r (the annual rate of change) is $sd/\sqrt{n}$. Rearranging Expression (1) yields

$$se(r) = \frac{|(C+1)^{1/d} - 1| - (b_u - b_l)}{Z_{\alpha/2} + Z_\beta}. \qquad (4)$$

With $d = 20$, $C = -0.5$, $b_u - b_l = 0.016$, $Z_{\alpha/2} = 1.645$, and $Z_\beta = 0.84$, as we suggested for the accuracy target, $se(r) = 0.0073$. With $b_u = 0.003$, $b_l = -0.003$, $b_u - b_l = 0.006$, and $se(r) = 0.0113$. Thus, with the current potential bias of ±0.008, estimated trends with standard errors ≤0.0073 achieve the accuracy target. We noted above that achieving the accuracy target in areas one-third the size of the temperate regions of Canada and the United States seems reasonable. A species whose range covers the entire area needs to have a rangewide $se(r)$ such that with one-third of the rangewide data, the $se(r)$ would still be ≤0.0073. Assuming an even distribution of routes, so that the sampling plan may be viewed as simple random, the rangewide standard error may be expressed as $c/\sqrt{n_r}$, where c is a constant, independent of sample size, and n_r is the rangewide sample size. The requirement is thus that the regional $se(r) = c/\sqrt{0.33 n_r} \leq 0.0073$ or that the rangewide standard error $c/\sqrt{n_r} = (\sqrt{0.33}) = 0.0042$. Thus, the threshold $se(r)$ for a species whose range covers the entire study area is 0.0042. By the same rationale, a species whose range covers two-thirds of the BBS survey area would need a $se(r) = \sqrt{0.67})*0.0073 = 0.0060$. We categorized each of the species with rangewide standard errors <0.0073 into 1 of 3 groups depending on whether their range covered <33 (group 1), 33–67 (group 2), or >67% (group 3) of the BBS survey area. The threshold standard errors for these groups were 0.0073, 0.0060, and 0.0042, respectively. We then determined how many of these species had observed, rangewide standard errors less than the threshold for their group. This analysis identified 124 species that are adequately monitored according to the proposed accuracy target. This number is 42% of the 297 species for which the BBS is considered a suitable monitoring program.

If potential bias is reduced to ±0.003, then the target rangewide standard errors are 0.0113 for species in group 1, 0.0092 for group 2, and 0.0065 for group 3. Using these target standard errors, the number of adequately monitored species is 183, or 62% of the species considered well suited to monitoring with the BBS. Thus, reducing potential bias from ±0.008 to ±0.003 would increase the number of adequately monitored species by about 50% (from 124 to 183). In contrast, doubling the number of BBS routes would increase the number of adequately monitored species by only about 30% (from 124 to 162). This analysis, like the analysis described above of how many routes are needed, shows that achieving a

Table 14. Estimated number of North American Breeding Bird Survey (BBS), or similar, routes needed to achieve adequate coverage of ≥80% of the North American landbirds that warrant monitoring, if potential bias is reduced to ±0.003 and if bias remains at its current level of ±0.008. The analysis assumes equal density of routes throughout the BBS survey area and that the goal is achieving the accuracy target for a species' entire range or in any area one-third the size of the temperate portions of Canada and the United States, whichever is smaller.

Province/state	Current no. of routes[a]	Needed if bias = ±0.003	Needed if bias = ±0.008	Province/state	Current no. of routes[a]	Needed if bias = ±0.003	Needed if bias = ±0.008
Alabama	83	72	172	New Brunswick	25	34	82
Alberta	90	189	454	New Hampshire	23	13	31
Arizona	65	158	379	New Jersey	31	10	25
Arkansas	36	73	176	New Mexico	62	169	406
B.C.[b]	81	178	427	New York	110	67	162
California	185	219	526	Newfoundland	14	40	97
Colorado	107	145	347	North Carolina	60	68	163
Connecticut	18	7	17	North Dakota	45	98	236
Delaware	11	3	7	Nova Scotia	27	27	64
Florida	93	78	186	Ohio	76	57	137
Georgia	65	81	195	Oklahoma	61	97	233
Idaho	59	116	278	Ontario	105	148	354
Illinois	82	78	188	Oregon	109	135	324
Indiana	46	51	121	Pennsylvania	108	63	151
Iowa	35	78	188	P.E.I.[c]	4	3	7
Kansas	38	114	274	Quebec	73	110	65
Kentucky	43	56	134	Rhode Island	5	1	3
Louisiana	44	64	153	Saskatchewan	45	120	289
Maine	61	45	107	South Carolina	24	43	103
Manitoba	46	95	229	South Dakota	50	107	257
Maryland	64	14	32	Tennessee	48	58	140
Massachusetts	23	11	27	Texas	170	367	881
Michigan	81	80	193	Utah	73	118	283
Minnesota	77	117	282	Vermont	24	13	32
Mississippi	35	66	159	Virginia	72	55	133
Missouri	54	97	233	Washington	89	93	224
Montana	62	205	491	West Virginia	50	34	81
Nebraska	43	107	258	Wisconsin	72	78	187
Nevada	27	154	369	Wyoming	103	136	326

[a] Indicates routes conducted by BBS.
[b] British Columbia.
[c] Prince Edward Island.

general increase in the number of BBS routes is not a very effective way to increase the number of adequately monitored species.

With potential bias of ±0.008, the number of routes required for adequate coverage of ≥80% of the landbirds that warrant coverage was 370 (Table 11). With potential bias reduced to ±0.003, the number of required routes was 150. We used these values (n = 370 and n = 150), along with species-specific values for A_i and y_i, in Expression (3) to estimate the density of routes needed for the sample of species best suited to monitoring using the BBS. The eightieth quantiles for the route density were 12 routes/10,000 km^2 with potential bias of ±0.008 and 5 routes/10,000 km^2 with potential bias of ±0.003. We used these densities to estimate the minimum number of BBS routes that would be needed in each province and state to achieve the monitoring objective for ≥80% of the species best suited to monitoring with the BBS.

The results (Table 14) showed that 5 provinces or states have the required number of routes at present (potential bias of ±0.008); but if potential bias can be reduced to ±0.003, then 19 provinces and states have the required number of routes. The total number of surveyed routes currently is about 3,640. If potential bias can be reduced to ±0.003, the needed number is 5,106 (a 40% increase); whereas if bias is not reduced, then >12,000 routes would be needed to meet the accuracy target for 80% of the species that warrant monitoring. If bias were reduced to ±0.003, then nearly enough routes are being surveyed in most states and provinces, but the number of BBS or similar routes needs to be more than doubled in 15 states and provinces (Table 12).

In summary, at present, only about 42% of the species that are considered suited for monitoring with the BBS and similar programs are adequately monitored under the standard we propose.

Increasing the number of routes will not by itself greatly increase the number of adequately monitored species. However, if measures to reduce potential bias are implemented, then a 40% increase in the number of BBS routes, concentrated in 15 states and provinces, would result in adequate coverage for approximately 80% of the species that are suitable for monitoring with the BBS and similar programs.

MANAGEMENT AND RESEARCH IMPLICATIONS

Reducing Bias

The difficulty of setting limits on bias should be an impetus to developing and using methods that have little or no bias. The review by O'Connor et al. (2000) identified many modifications to the BBS that would reduce potential bias. Several general methods have been proposed for reducing potential bias, including distance methods (Buckland et al. 2001), double-observer methods (Nichols et al. 2000), removal methods (Farnsworth et al. 2002), and double sampling (Bart and Earnst 2002). Review of these approaches is beyond the scope of this report, but we recommend that program designers give careful consideration to whether these or other methods to reduce potential bias may yield estimates of higher accuracy than unadjusted counts.

Increased Coordination

Many state and regional programs exist or are being planned to monitor landbirds during the breeding season. For example, monitoring programs are scheduled for 9 of the 11 westernmost states (excluding Alaska and Hawaii), and at least 6 large, long-term programs are coordinated by the U.S. Forest Service (C. Hargis, U.S. Forest Service, personal communication). Although these programs generally use different methods than those used by the BBS, if the trend estimates they produce are reliable, then they could be combined with the BBS trend estimate. Thus, many states and regions already have—or soon will have—sufficient data to meet the accuracy target. The challenge is to integrate all of these efforts into a single, comprehensive landbird monitoring program.

Other Surveys

More than one-third of the ranges of 167 landbird species are within the northern boreal regions. These species cannot be adequately monitored solely with temperate breeding-season surveys. Breeding-season surveys in northern regions and/or surveys at other times of year must be implemented. A major effort is thus needed to determine which approach should be implemented and to design the needed new surveys.

Approximately 95 temperate-nesting species are unlikely to be covered either by increased BBS (or similar) surveys or by surveys in northern areas (Table 13). Extensive surveys exist for many upland game species and raptors (mainly at migration stations), and several regional programs exist for nocturnal and colonial species. Much more work is needed, however, to develop comprehensive programs for species not suited to coverage by the BBS.

Increased Funding for Centralized Programs

The BBS is coordinated by 2 federal employees, 1 in Canada and 1 in the United States. Improvements, such as developing ways to reduce potential bias and implementing new surveys, will require a substantial expansion of the resources that support these centralized programs. In designing long-term surveys, biologists can help implement these ideas by: (1) adopting clear accuracy targets, such as we suggested; (2) insisting that potential bias be estimated in survey design and analysis, and that effects of bias be acknowledged when making inferences; and (3) encouraging preparation and peer review of survey protocols before the surveys are implemented.

ACKNOWLEDGMENTS

We thank D. Krueper, T. Leukering, B. Peterjohn, and K. Rosenberg for reviewing and revising the list of species that warrant monitoring and for identifying the surveys most suitable for each species, and B. Peterjohn and 2 anonymous reviewers for extensive comments on past drafts of this manuscript. D. McNicol provided the range maps from Project WILDSPACE. This project was supported by the U.S. Geological Survey and by Bird Studies Canada.

LITERATURE CITED

ASKINS, R. A. 2000. Restoring North America's birds. Yale University Press, New Haven, Connecticut, USA.

BAIN, L. L., AND M. ENGELHARDT. 1987. Introduction to probability and mathematical statistics. Doxbury Press, Boston, Massachusetts, USA.

BART, J., B. COLLINS, AND R. I. G. MORRISON. 2003. Trend estimation using a linear model. Condor 105:367–372.

———, AND S. L. EARNST. 2002. Double sampling to estimate bird density and population trends. Auk 199:36–45

———, M. HOFSCHEN, AND B. G. PETERJOHN. 1995. Reliability of the Breeding Bird Survey: effects of restricting surveys to roads. Auk 112:758–761.

BROWN, S., C. HICKEY, B. HARRINGTON, AND R. GILL, editors. 2001. United States shorebird conservation plan. Second edition. Manomet Center for Conservation Sciences, Manomet, Massachusetts, USA.

BUCKLAND, S. T., D. R. ANDERSON, K. P. BURNHAM, J. L. LAAKE, D. L. BORCHERS, AND L. THOMAS. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, United Kingdom.

BUTCHER, G. S., B. G. PETERJOHN, AND C. J. RALPH. 1993. Overview of national bird population monitoring programs and databases. Pages 192–203 *in* D. M. Finch and P. W. Stangel, editors. Status and management of Neotropical migratory birds: proceedings of the 1992 Partners in Flight National Training Workshop, 21–25 September, Estes Park, Colorado. U.S. Forest Service Rocky Mountain Forest and Range Experiment Station General Technical Report RM-229.

CARSON, R. C. 1962. Silent spring. Houghton Mifflin, New York, New York, USA.

COCHRAN, W. G. 1977. Sampling techniques. Third edition. John Wiley & Sons, New York, New York, USA.

DOWNES, C. M., E. H. DUNN, AND C. M. FRANCIS. 2000. Canadian landbird monitoring strategy: monitoring needs and priorities into the new millenium. Partners In Flight, Canada, Ottawa, Canada.

FARNSWORTH, G. L., K. H. POLLOCK, J. D. NICHOLS, T. R. SIMONS, J. E. HINES, AND J. R. SAUER. 2002. A removal method for estimating detection probabilities from point-count surveys. Auk 119:414–425.

GUTIÉRREZ, R. J., A. B. FRANKLIN, AND W. S. LAHAYE. 1995. Spotted owl (*Strix occidentalis*). Number 179 *in* A. Poole and F. Gill, editors. The birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania, USA.

JAMES, F. C., C. E. MCCULLOCH, AND D. A. WIEDENFELD. 1996. New approaches to the analysis of population trends in land birds. Ecology 77:13–27.

KELLER, C. M. E., AND J. T. SCALLAN. 1999. Potential roadside biases due to habitat changes along Breeding Bird Survey routes. Condor 101:50–57.

KENDALL, W. L., B. G. PETERJOHN, AND J. R. SAUER. 1996. First-time observer effects in the North American Breeding Bird Survey. Auk 77:13–21.

LINK, W., AND J. SAUER. 1994. Estimating equations estimates of trends. Bird Populations 2:23–32.

MAYFIELD, H. F. 1992. Kirtland's warbler. Number 19 *in* A. Poole and F. Gill, editors. The birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania, USA.

NATIONAL GEOGRAPHIC SOCIETY. 1999. Field guide to the birds of North America. Third edition. National Geographic Society, Washington, D.C., USA.

NICHOLS, J. D., J. E. HINES, J. R. SAUER, F. W. FALLON, J. E. FALLON, AND P. J. HEGLUND. 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117:393–408.

O'CONNOR, R. J., E. DUNN, D. H. JOHNSON, S. L. JONES, D. PETIT, K. POLLOCK, C. R. SMITH, J. L. TRAPP, AND W. WELLING. 2000. A programmatic review of the North American Breeding Bird Survey. Available at [http://www.mp2-wrc.usgs.gov/bbs/index.htm].

PARMESAN, C., AND G. YOHE. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42.

PASHLEY, D. N., C. J. BEARDMORE, J. A. FITZGERALD, R. P. FORD, W. C. HUNTER, M. S. MORRISON, AND K. V. ROSENBERG. 2000. Partners in Flight: conservation of land birds of the United States. American Bird Conservancy, The Plains, Virginia, USA.

SAUER, J. R., J. E. HINES, AND J. FALLON. 2001. The North American Breeding Bird Survey, results and analysis 1966–2001. Version 2001.2. U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland, USA.

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

SCHROEDER, M. A., J. R. YOUNG, AND C. E. BRAUN. 1999. Sage grouse (*Centrocercus urophasianus*). Number 425 *in* A. Poole and F. Gill, editors. The birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania, USA.

SHEAIL, J. 1985. Pesticides and nature conservation. Clarendon Press, Oxford, United Kingdom.

STEEL, R. G. D., AND J. H. TORRIE. 1980. Principles and procedures of statistics. Second edition. McGraw-Hill, New York, New York, USA.

WELSH, D. A., L. A. VENIER, D. R. FILLMAN, J. MCKEE, D. PHILLIPS, K. LAWRENCE, I. GILLESPIE, AND D. W. MCKENNEY. 1999. Development and analysis of digital rangemaps of birds breeding in Canada. Information Report ST-X-17. Science Branch, Canadian Forest Service, Natural Resources Canada, Ottawa, Ontario, Canada.

WHITE, C. M., N. J. CLUM, T. J. CADE, AND W. G. HUNT. 2002. Peregrine falcon (*Falco peregrinus*). Number 660 *in* A. Poole and F. Gill, editors. The birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania, USA.

WILLIAMS, B. K., J. D. NICHOLS, AND M. J. CONROY. 2002. Analysis and management of animal populations. Academic Press, New York, New York, USA.

Received 9 June 2003.
Accepted 7 April 2004.
Associate Editor: Flaspohler.

(Appendix on next page)

APPENDIX A: DERIVATIONS AND ADDITIONAL EXPLANATIONS

Formulas for Power

We follow the notation of Steel and Torrie (1980:114–119): α and β are the Type I and Type II errors, respectively (power = $1-\beta$), and Z_α satisfies the expression $P(Z > Z_\alpha) = \alpha$, where Z is a standard normal variable. Initially, we assume large sample sizes so that Z-values may be used. Let R equal the true rate of change in population size, r equal the estimate of R, and $se(r)$ be the estimated standard error of r. Let bias in r be $E(r) - R = B$. Since r and R are expressed as annual rates of change, B is a small number such as –0.01 or 0.005. We assume that $b_l < B < b_u$.

If the trend estimate is <1, we only wish to conclude that the population is declining if the observed decline is significantly less than 1 minus the lower endpoint for B; that is, if

$$r + Z_{\alpha/2}se(r) < 1 + b_l\,, \qquad \text{(A.1)}$$

where $Z_{\alpha/2}$ is the critical value. We would thus consider the result significant if and only if

$$\frac{r-1-b_l}{se(r)} < -Z_{\alpha/2}\,. \qquad \text{(A.2)}$$

By the same reasoning, if the point estimate is >1, we would only consider it significant if

$$\frac{r-1-b_u}{se(r)} > Z_{\alpha/2}\,. \qquad \text{(A.3)}$$

The probability of rejecting the null hypothesis that $R = 1$ is thus the sum of the probabilities of the events described by Expressions (A.2) and (A.3). First, consider

$$P\left(\frac{r-1-b_l}{se(r)} < -Z_{\alpha/2}\right). \qquad \text{(A.4)}$$

To calculate this probability, we manipulate the inequality so that the left side becomes a standard normal variable. Because the value of b_l does not depend on the sample, $E(r - 1 - b_l) = R - 1 + B - b_l$. For the same reason, $SE(r - 1 - b_l) = se(r)$. Thus, for $r < 1$, from Expression (A.2), we obtain

$$P\left(\frac{r-1-b_l}{se(r)} < -Z_{\alpha/2}\right) = P\left(Z < -Z_{\alpha/2} - \frac{R-1+B-b_l}{se(r)}\right),$$
$$= P\left(Z < -Z_{\alpha/2} + \frac{1-R-(B-b_l)}{se(r)}\right) \qquad \text{(A.5)}$$

where Z is a standard normal variable. By the same rationale, for $r > 1$, from Expression (A.3),

$$P\left(\frac{r-1-b_u}{se(r)} > Z_{\alpha/2}\right) = P\left(Z > Z_{\alpha/2} - \frac{R-1+B-b_u}{se(r)}\right)$$
$$= P\left(Z > Z_{\alpha/2} + \frac{1-R+(b_u-B)}{se(r)}\right). \qquad \text{(A.6)}$$

Omitting b_l and b_u is equivalent to ignoring the bias. In this case, the probability of a Type I error (rejecting the null hypothesis when it is true), since $R = 1$, is

$$P(\textit{Type I error} \mid b_l, b_u \textit{ undefined}) =$$
$$P\left(Z < -Z_{\alpha/2} - \frac{B}{se(r)}\right) + P\left(Z > Z_{\alpha/2} - \frac{B}{se(r)}\right). \qquad \text{(A.7)}$$

Power is usually *defined* as the probability of rejecting the null hypothesis when it is false, and then 1 of the outcomes, Expression (A.5) or (A.6) in our case, is usually *assumed* to be sufficiently unlikely to be ignored. Thus, if $R > 1$ and power is high enough to be interesting, the usual assumption would be that a significant $r < 1$ would hardly ever occur, so the probability of this event could be ignored. This is not necessarily true when bias is present but is ignored. We therefore defined power as the probability of obtaining a significant $r < 1$ if $R < 1$ or a significant $r > 1$ if $R > 1$. This distinction only mattered in a few analyses, but when it did, we referred to the power to detect a decline (if $R < 1$) or an increase (if $R > 1$). Thus,

$$\textit{Power with } R < 1, b_l, b_u \textit{ undefined} =$$
$$P\left(Z < -Z_{\alpha/2} - \frac{R-1+B}{se(r)}\right) \qquad \text{(A.8)}$$

and

$$\textit{Power with } R > 1, b_l, b_u \textit{ undefined} =$$
$$P\left(Z > Z_{\alpha/2} + \frac{1-R-B}{se(r)}\right). \qquad \text{(A.9)}$$

When b_l and b_u are defined, then a further simplification is possible. In Expression (A.5) for $r < 1$, $b_l < B$ so $B - b_l = |B - b_l|$, and we may write

$$P\left(\frac{r-1-b_l}{se(r)} < -Z_{\alpha/2}\right) =$$
$$P\left(Z < -Z_{\alpha/2} - \frac{R-1+|B-b_l|}{se(r)}\right). \qquad \text{(A.10)}$$

In Expression (A.6) for $r > 1$, $b_u > B$, so $(b_u - B) = |b_u - B|$ or $|B - b_u|$ because the order in an absolute difference does not matter. Thus,

$$P\left(\frac{r-1-b_l}{se(r)} > Z_{\alpha/2}\right) = P\left(Z > Z_{\alpha/2} + \frac{1-R+|B-b_u|}{se(r)}\right). \quad (A.11)$$

Expressions (A.10) and (A.11) may be used to calculate a general expression for the probability of making a Type I error when bias is acknowledged:

$$\begin{aligned} &P(\textit{Type I error} \mid b_l, b_u \textit{ defined}) \\ &= P\left(Z < -Z_{\alpha/2} - \frac{|B-b_l|}{se(r)}\right) + P\left(Z > Z_{\alpha/2} + \frac{|B-b_u|}{se(r)}\right) \\ &= P\left(Z > Z_{\alpha/2} + \frac{|B-b_l|}{se(r)}\right) + P\left(Z > Z_{\alpha/2} + \frac{|B-b_u|}{se(r)}\right) \end{aligned} \quad (A.12)$$

The maximum occurs when $B = b_l$ or $B = b_u$, in which case

$$P(\textit{Type I error} \mid b_l, b_u \textit{ defined}) = P(Z > Z_{\alpha/2}) + P\left(Z > Z_{\alpha/2} + \frac{(b_u - b_l)}{se(r)}\right). \quad (A.13)$$

The minimum occurs when $B = (b_u + b_l)/2$, in which case $|B - b_l| = |B - b_u| = (b_u - b_l)/2$, so

$$P(\textit{Type I error} \mid b_l, b_u \textit{ defined}) = 2P\left(Z > Z_{\alpha/2} + \frac{(b_u - b_l)}{2se(r)}\right). \quad (A.14)$$

Expressions (A.13) and (A.14) thus provide the limits for the probability of a Type I error when bias is between b_l and b_u and these limits are acknowledged in the analysis.

If power is fairly high (e.g., <0.50) then the usual assumption that 1 of the alternatives, significant $r < 1$ or significant $r > 1$, can be ignored is reasonable. If $R < 1$, then from Expression (A.10), we may write

$$\begin{aligned} &\textit{Power with } R < 1, b_l, b_u \textit{ defined} \\ &= P\left(Z > Z_{\alpha/2} + \frac{R-1+|B-b_l|}{se(r)}\right) \\ &= P\left(Z > Z_{\alpha/2} - \frac{1-R-|B-b_l|}{se(r)}\right) \\ &= P\left(Z > Z_{\alpha/2} - \frac{|1-R|-|B-b_l|}{se(r)}\right) \end{aligned} \quad (A.15)$$

the last line holding because $R < 1$. If $R > 1$, then from Expression (A.11), we may write

$$\begin{aligned} &\textit{Power with } R > 1, b_l, b_u \textit{ defined} \\ &= P\left(Z > Z_{\alpha/2} - \frac{R-1-|B-b_u|}{se(r)}\right) \\ &= P\left(Z > Z_{\alpha/2} - \frac{|R-1|-|B-b_u|}{se(r)}\right) \end{aligned} \quad (A.16)$$

because $R > 1$. Thus, a general formula may written for power,

$$\textit{Power with } b_l, b_u \textit{ defined} = P\left(Z > Z_{\alpha/2} - \frac{|R-1|-|B-b_p|}{se(r)}\right), \quad (A.17)$$

where b_p (p for potential) $= b_l$ for $R < 1$ and $= b_u$ for $R > 1$. The maximum value of $|B - b_p|$ is $b_u - b_l$, so this value should be used to ensure that power will be as high as predicted by Expression (A.17). We thus use

$$\textit{Power with } b_l, b_u \textit{ defined} \geq P\left(Z > Z_{\alpha/2} - \frac{|R-1|-(b_u-b_l)}{se(r)}\right). \quad (A.18)$$

Continuing with the derivation, if power is to be at least $1 - \beta$, then it must be true that

$$Z_{1-\beta} = -Z_\beta = Z_{\alpha/2} - \frac{|R-1|-(b_u-b_l)}{se(r)}. \quad (A.19)$$

If simple random sampling is employed to select survey locations, and the trend estimation method of Bart et al. (2003) is used to calculate r, then the $se(r)$ may be written

$$se(r) = \sqrt{\frac{v}{n}}, \quad (A.20)$$

where

$$v=\frac{1}{\bar{y}_{mid}^{2}}\left(v(b_i)+\left(\frac{\bar{b}}{\bar{y}_{mid}}\right)^2 v(y_{mid,i})-2\left(\frac{\bar{b}}{\bar{y}_{mid}}\right)cov(b_i,y_{mid,i})\right), \quad (A.21)$$

b_i and $y_{mid,i}$ are the slope and midpoint of the regression for site i (Bart et al. 2003). The BBS locations are a stratified sample, not a simple random sample, but we used a method described by Cochran (1977:136) to estimate the population variances and covariances.

Expressing R in terms of total change makes the results above more general. Let C be the total change during the survey's d years. If the survey result declined 25%, then C would be –0.25. The relationship between C and R is $C = R^d - 1$, and thus $R = (C + 1)^{1/d}$.

In Expression (A.19), substituting $(V/n)^{0.5}$ for $se(r)$, $(C + 1)^{1/d}$ for R, and solving for n yields

$$n=\left(\frac{sd(Z_{\alpha/2}+Z_\beta)}{|(C+1)^{1/d}-1|-(b_a-b_i)}\right)^2, \quad (A.22)$$

where $sd = v^{0.5}$. Expression (A.22) is Expression (1) from the body of the paper.

When t-values will be used in the analyses, Steel and Torrie (1980:118) recommend multiplying the estimated power, obtained assuming Z-values are used, by $(df + 3)/(df + 1)$, where df = degrees of freedom. If $df > 20$, then the multiplier is <1.1 so it has little effect in cases of practical interest.

Derivation of Expression (2)

For a given species, let R_i = the true trend in region i and r_i = the estimate of this trend. The variance of the r_i may be expressed as

$$V(r_i)=E[Var(r_i)]+V[E(r_i)]. \quad (A.23)$$

For this analysis, we defined R_i as $E(r_i)$. Making this substitution and rearranging Expression (A22),

$$V(R_i)=V(r_i)-E[Var(r_i)]\ . \quad (A.24)$$

An unbiased estimate of $V(r_i)$ is provided by the sample analogue, $v(r_i)$. An unbiased estimate of $Var(r_i)$ is provided by $[SE(r_i)]^2$, where $SE(r_i)$ is obtained from BBS data. The mean of the $[SE(r_i)]^2$ of the values for regions with estimates is $E[Var(r_i)]$. Thus, an unbiased estimate of $V(R_i)$ is

$$\hat{V}(R_i)=v(r_i)-\sum_i^{n_i}[SE(r_i)]^2/n_i\ , \quad (A.25)$$

where n_i = the number of regions with estimated trends for the species. The standard deviation of the true, region-specific trends may be estimated as the square root of $\hat{V}(R_i)$.

We calculated $\hat{V}(R_i)$ using BBS data for species recorded on >50 routes, having rangewide SE(trend)s of <0.01, and for which trend estimates from at least 4 of the 6 U.S. Fish and Wildlife Service regions were available. These criteria yielded 89 species. Estimates of $V(R_i)$ were made using Expression (A.3) for each of the 89 species. We found 1 clear outlier (house finch [*Carpodacus mexicanus*]; variance >5 times larger than the next smaller value). Excluding this value, the mean estimated variance was 0.024; the estimated standard deviation was 0.016.

Suppose that a species was equally abundant in n regions, and the standard deviation of the true, region-specific trends was 0.016. How much bias would be caused by estimating the rangewide trend using data from fewer than all n regions? Let $n_1 + n_2 = n$, where we use trends from n_1 regions, and ignore the other n_2 trends in calculating the mean trend. Let R_n = the mean of all n regional trends, and define this as the population-wide trend (since we assume the species to be equally abundant in the regions). Also, let R_1 be the mean of the n_1 regional trends and R_2 be the mean of the other n_2 regional trends. The bias arising from using R_1 to estimate R_n is $R_1 - R_n$, so we need an expression for this difference that allows us to calculate the probability that the bias exceeds any given amount (G) based on the estimated variation among the regional values, which is $SD(R_i)$ = 0.016. If we randomly select the n_1 regions to use in calculating R_1, then R_1 is the mean of n_1 normal random variables. The expected value is R_n and its variance (e.g., Bain and Engelhardt 1987:213) is σ^2/n_1, or $0.0243/n_2$ in our case. We may therefore write

$$P(R_1-R_n)>G=P\left(\frac{R_1-R_n}{\sigma^2/n_1}>\frac{G}{\sigma^2/n_1}\right)$$
$$=P\left(Z>\frac{n_1G}{0.0243}\right), \quad (A.26)$$

where Z is a standard normal variable.