FIRST-TIME OBSERVER EFFECTS IN THE NORTH AMERICAN BREEDING BIRD SURVEY

WILLIAM L. KENDALL,1,4 BRUCE G. PETERJOHN,2 AND JOHN R. SAUER3

1U.S. Fish and Wildlife Service, Office of Migratory Bird Management, Laurel, Maryland 20708, USA;
2National Biological Service, Division of Inventory and Monitoring, Laurel, Maryland 20708, USA; and
3National Biological Service, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA

ABSTRACT.—Currently the operational analysis of Breeding Bird Survey (BBS) data by the National Biological Service accounts for observer differences in estimating the trend for each route, but within-observer differences are not modeled. We tested for the existence of a form of within-observer differences in skill level, namely a change in ability to count birds of a given species after an observer's first year on a given route. An increase in ability could positively bias the trend estimate. Removal of an observer's first year of observation on each route for the period 1966 to 1991 resulted in lower average unweighted trend estimates for 415 of 459 species (90%). These reductions were statistically significant for 213 species (46%). The average reduction in trend was 1.8% change per year (SD = 5.4%). In route-regression analysis, route data are weighted by a measure of precision. Removing first-year observer counts reduced the weighted trend estimate for 275 of 416 species (66%), but differences generally were small. Received 13 July 1995, accepted 11 March 1996.

DATA FROM LONG-TERM SURVEYS such as the North American Breeding Bird Survey (BBS) (Robbins et al. 1986) have been widely used to compute temporal trends in bird populations. However, analyses are complicated by a variety of factors such as missing years of data, inconsistent route coverage, regional differences in route densities, and observer effects (Peterjohn et al. 1995). For example, observer skills significantly improved during the history of the BBS, resulting in a tendency for recent observers to count greater numbers of individuals and species than their counterparts who initiated the survey (Sauer et al. 1994). Hence, any analysis of BBS data should treat observers as covariates in order to avoid confounding their improvement in skills (as a group) with population trends.

Although differences in skill levels among observers have been documented, subtle effects within observers also can influence analyses of data from surveys like the BBS. For example, observers tend to lose some of their hearing ability as they advance in age, especially at the higher ranges of frequencies (Ramsey and Scott 1981), possibly creating the perception of population decline. Conversely, Erskine (1978) suggested the existence of a first-time observer effect along BBS routes, where observers tend to count more individuals and species in subsequent years than during their initial year of conducting the survey. This first-time effect could be due to general improvement in bird identification skills, increased efficiency in counting birds under the defined methodology, and/or learning the stop locations along the route and associating particular species with individual stops. First-time observer effects could have important consequences for data analyses. The effects may differ greatly among routes, as some routes have been surveyed by only a single observer whereas others have had as many as 11 different observers (Sauer et al. 1994). The number of first-time observers could influence estimation of population change for a route, and its effect may have to be taken into account in order to reduce bias. A first-time observer effect has two potential implications: (1) it could identify the need to train observers more thoroughly before allowing them to conduct a survey, and (2) it may justify excluding all first-year observations from the analysis of BBS data or directly modeling the effect in the analysis.

In this paper we examined the presence of first-time observer effects for species recorded
on the BBS. These effects were assessed at the level of the individual route by comparing trends estimated using all data except for those from the first year of a survey for each observer with trends estimated using all data. The consequences of first-year effects on the weighted composite trend estimates were evaluated by estimating these trends with and without first-year observer data using the methods described in Geissler and Sauer (1990).

**Methods**

Each of the 3,700 BBS routes distributed across the continental United States and Canada is randomly located along secondary roads. They are 24.5 miles (39.4 km) in length, with 50 point counts conducted at 0.5-mile (0.8-km) intervals. At each point, all birds heard or seen within a 0.25-mile (0.4-km) radius are counted during a 3-min period. Robbins et al. (1986) described the BBS methodology in greater detail.

BBS population trend estimates have been computed using a route-regression method (Geissler and Sauer 1990). For each route a linear regression over years is run on the log of the total count (adjusted by described the BBS methodology in greater detail. At each point, all birds heard or seen within a 0.25-mile (0.4-km) radius are counted during a 3-min period. Robbins et al. (1986) described the BBS methodology in greater detail.

BBS population trend estimates have been computed using a route-regression method (Geissler and Sauer 1990). For each route a linear regression over years is run on the log of the total count (adjusted by a small constant) of a species, treating observers as covariates. The estimated slope is then back-transformed to produce an estimate of trend. The route trends are averaged (weighted by their estimated precision and the marginal mean count for the route) to produce a trend estimate for a state-stratum unit (i.e. physiographic region within a state). Because of autocorrelation among counts on a route, variance estimates for state-stratum trends are derived from bootstrapped samples of route trends. Regional trend estimates are then computed as averages of the state-stratum trends, weighted by area in addition to precision and marginal mean counts. See Geissler and Sauer (1990) for details of the weighting process and variance estimation, and Link and Sauer (1994) for some innovations on the route-regression procedure. These weighting factors may mitigate consequences of first-time observer effects, because routes with more "start-ups" are given less weight in the overall analysis.

We tested for first-time observer effects from 1966 to 1991 at: (1) the level of individual survey routes; and (2) the regional level, under the weighting scheme of the route-regression method outlined above. In the route-by-route analysis, we treated each route as an experimental unit in a repeated-measures analysis. The BBS analysis using all data served as the control. The treatment consisted of removing the first year of observation for each observer for each route, then computing the new slope. Our response variable was the difference in slopes computed with and without first-time observer data. For each species, we tested for a non-zero average difference in slopes, using a nonparametric sign test (Hollander and Wolfe 1973:39) and a paired t-test. Both tests were two-tailed. The sign test considers only the direction of the difference in slopes and makes no distributional assumptions about the slopes. The paired t-test considers magnitude and direction of the difference in slopes, but assumes the slopes are distributed normally across routes.

To assess the effect of first-time observers across species (i.e. a general first-time observer effect for BBS routes), we computed the proportion of species that demonstrated a significant effect at the 5% level of significance, for both the sign test and t-test approaches. If first-time observer effects for BBS routes generally were nonexistent (null hypothesis), significant results are still expected for 5% (based on $\alpha = 0.05$) of the species by chance alone. A first-time observer effect is indicated when significant results are found for more than 5% of the species evaluated. We used a binomial test to evaluate this hypothesis. For the sign test approach we only considered species for which the estimated difference in slope was based on $\geq 10$ routes. For including a species in the t-test approach we required that $\geq 25$ routes contributed to the test for difference, in order to make the normality assumption more reasonable.

In order to assess the consistency of first-time effects, bird species were grouped based on breeding habitat type (grassland, wetland, scrub, woodland, or urban), nest type (cavity or open cup), migration strategy (permanent resident, short-distance migrant, or Neotropical migrant), and nesting strategy (ground/low or mid-story/canopy; see Peterjohn and Sauer 1993). For each of these four groupings, we used a contingency table approach to test whether significant first-time observer effect was independent of category (e.g. is this effect found more frequently with cavity nesters than open cup nesters?).

To assess the effect of first-time observers on the weighted analysis of BBS data, we analyzed the complete data set using the route-regression method of Geissler and Sauer (1990), including the weighting scheme described above. We then ran the same analysis, but restricted the input by removing the first year of data for each observer for each route. We computed the proportion of species for which this restriction resulted in a decreased trend estimate and used a binomial test to determine if it was significantly different from 0.5. We also computed the proportion of species for which this restriction resulted in a reversal of the sign of the trend estimate (e.g. the full data produced an estimated increase for a species, but the reduced data produced an estimated decline). In addition, we divided the species results into two groups, based on whether the removal of an observer's first count produced an increased or decreased trend estimate. For each group we computed the proportion of species for which the trend estimate changed in statistical significance after restriction (i.e.}
the trend estimate before restriction was significantly different from 0 and that after restriction was non-significant, and vice versa).

RESULTS

Route-specific trend results.—Of 459 species considered (each sample size ≥10), trend estimates decreased for 415 species (90%) when the first year of counts for each observer was omitted, suggesting a positive first-time observer effect. Average decrease was 1.8% change per year (SD = 5.4%). Sign tests indicated significantly different average slopes at the 5% level for 219 species (48%; see Appendix). Based on a binomial test, the number of differences is significantly ($P < 0.001$) higher than would be expected by chance. All but seven of the significant results indicated positive first-time observer effects. Paired $t$-tests indicated a significant first-time effect for 70 of 413 species (17%), and without exception the effect was positive.

Table 1 indicates how the significant results (from sign tests) were distributed within species groupings. For instance, of the 361 species classified by habitat type (Peterjohn and Sauer 1993), 28 (7.8%) were classified as grassland birds. Overall, the distributions of species indicating a significant first-time observer effect and those indicating no such effect were not different across habitat types ($\chi^2 = 3.97, df = 4, P = 0.41$) or between nesting strategies ($\chi^2 = 0.09, df = 1, P = 0.76$). The result was marginal for nest type ($\chi^2 = 3.2, df = 1, P = 0.075$) and significant for migration strategy ($\chi^2 = 24.9, df = 2, P = 0.001$). Closer inspection of the chi-square value for migration strategy indicated that most of it ($\chi^2 = 17.5$) was contributed by the cells associated with short-distance migrants (i.e. the proportion of significant results that were short-distance migrants [16%] was much lower than the proportion of non-significant results that fell into that category [40%]).

Composite regional trend results.—Trends computed under the full-weighting scheme of the BBS analysis showed that removing the first-year observer counts decreased the estimated trend for 275 (66.1%) of 416 species observed on at least 20 routes. This is significantly greater than 50% ($P < 0.001$). Of these 275 species, 14 (5.2%) exhibited a concomitant increase in significance level (i.e. the $P$-value for the test for non-zero trend changed from >0.05 to <0.05). For 21 species (7.3%), the significance level decreased (i.e. $P$ changed from <0.05 to >0.05). For the 141 species where removal of first-time observer counts increased the trend estimate, the number of species for which the significance level of the test increased or decreased was 10 (7.0%) and 10 (7.0%), respectively. Overall, the median change in trend estimates was $-0.28\%$ change per year (25th percentile = $-0.66\%$ per year; 75th percentile = 0.13% per year), whereas the median proportional change was $-15\%$ (25th percentile = $-54\%$; 75th percentile = 11%). For 31 species (7.5%), removal of an observer’s first year resulted in a change in the sign of the trend from positive to negative. For 12 species (2.9%), the trend switched from negative to positive.

DISCUSSION

Using routes as experimental units, we have demonstrated a substantial first-time observer effect on bird counts in the BBS from 1966 to 1991. With few exceptions, this effect was in a positive direction (i.e. the trend estimates decreased when the counts from an observer’s first
year were removed) and averaged 1.8% change in the population per year.

Given the enormous effort required to initiate the BBS in 1966-67, this first-time observer effect could be confounded with a survey start-up effect. To assess this possibility, we conducted our analysis again, restricting the period to 1968 to 1991, and found virtually no difference in the proportion of species indicating an effect. In addition, our analysis is based on the assumption that the logarithm of counts is linear over time for a given observer. If the logarithm of the actual number of animals on the route decreases or increases exponentially for a given observer (this would likely require long strings of years with the same observer), our analysis could indicate a first-time observer effect when, in fact, none exists. We are skeptical that this is the case, however.

Our a priori belief was that there would be a positive first-time observer effect on the trend estimates for rare, or difficult-to-identify species, due to artificially small first-time counts. We found this effect for 46% of the species. Conversely, assuming some observers tend to place less emphasis on counting numbers of abundant species in years following their initial survey, we expected a negative effect on trend estimates for some of these abundant species. We found no evidence for the latter, and several common species are among those displaying a positive effect (Appendix). The predominance of species displaying this positive effect reflects the various factors responsible for improving counts of individual birds after the first year an observer surveys a BBS route, although it precludes us from attributing the effect to any particular factor.

We found that species displaying significant first-time observer effects were disproportionately distributed among migration strategies. An effect was less likely for short-distance migrants, and more likely for Neotropical migrants and permanent residents. These results are partially counterintuitive, because we would have expected the permanent residents to include species that generally are most familiar to the first-time observers, and hence, less likely to display an effect.

Under the weighting scheme of Geissler and Sauer (1990), the positive bias resulting from the first-time effect was still apparent, but not as dramatic as the results of the route-based analysis. The trend estimates for 1966 to 1991 decreased for significantly more than 50% of the species when first-time observer counts were ignored, with a median 15% decrease in the trend estimate. The change in trend resulted in a change in significance level of the tests for non-zero trends for only 13% of the species. In addition, this restriction resulted in a reversal of the sign of the trend estimate for only 10% of species, most of which initially had trends fairly close to zero.

The existence of this first-time effect has some obvious implications for the analysis of BBS data and the operation of it and similar surveys. Analyses based on the unweighted trends of individual routes would tend to be positively biased as a result of these effects (e.g. a bias of 1.8% change per year could produce an apparently stable 20-year trend when the population had actually decreased by 40%). To reduce this bias, the first-year counts from each observer should be eliminated from analyses of individual route data. Additionally, all BBS trend estimates for relatively short periods of years relying on data from the 1960s should be viewed with considerable caution. Between 1966 and 1970, the data from all BBS routes would be influenced by these first-time effects, resulting in a positive bias in the trend estimates. These effects should be less noticeable in short-term trend estimates during subsequent years, because the proportion of first-time observers conducting surveys would be relatively small during any year, i.e. generally less than 10% of all routes surveyed (Sauer et al. 1994).

The implications of this first-time effect on the results of the weighted composite trend estimates are more complex. The weighting factors reduce the magnitude of positive bias, especially the precision weighting factor that reduces the influence of routes with relatively few years of data per observer. However, a slight positive first-time effect still exists in these analyses, as evidenced by the slightly more negative trends when the first-year counts were eliminated from the analyses. Eliminating these data can reduce the bias, but at the cost of reduced precision of the trend estimates. For both the unweighted and weighted analyses, modeling the first-time observer effect explicitly is preferable to eliminating data. Link and Sauer (unpubl. data) have modified their estimating equation approach to trend estimation (Link and Sauer 1994) to include this first-time observer effect.
Given the magnitude of first-time effects at the level of the individual BBS route, coordinators should improve the training of observers prior to their first survey for the BBS. The training should emphasize increasing familiarity with the point-count methodology employed by the survey and improving bird identification skills, particularly proficiency in identifying bird songs. Although this training could reduce the magnitude of first-time observer effects, it will never completely eliminate these effects, because factors such as familiarity with the stop locations, and the association of particular species with individual stops, cannot be controlled by training. Hence, the consequences of first-time effects always must be addressed in the analyses of the data.

ACKNOWLEDGMENTS

We thank W. Link, C. Moore, R. Dettmers, C. Robbins, and an anonymous referee for helpful comments on previous drafts of this manuscript. M. A. McKeough typed the Appendix.

LITERATURE CITED


APPENDIX. Species with significant first-time observer effects in BBS route analyses as indicated by sign tests based on ≥10 routes. Numbers following names of species are: (1) number of routes, (2) proportion of routes where removal of first-time observations reduced the trend, and (3) P-value for a test of the null hypothesis of zero average difference. Probabilities are one-sided but the test is two-tailed. Hence, P < 0.025 and P < 0.005 indicate a positive first-time observer effect significant at α = 0.05 and α = 0.01, respectively; and P > 0.975 and P > 0.995 indicate a negative first-time observer effect at α = 0.05 and α = 0.01, respectively.

Horner Grebe, Podiceps auritus, 94, 0.63, 0.01; Pied-billed Grebe, Podilymbus podiceps, 492, 0.64, 0.00; Common Loon, Gavia immer, 399, 0.55, 0.02; Great Black-backed Gull, Larus marinus, 110, 0.64, 0.00; Herring Gull, Larus argentatus, 392, 0.61, 0.00; Franklin's Gull, Larus pipixcan, 183, 0.61, 0.00; Bonaparte's Gull, Larus philadelphia, 51, 0.74, 0.00; Common Tern, Sterna hirundo, 155, 0.60, 0.01; American White Pelican, Pelecanus erythrorhynchos, 115, 0.63, 0.00; Red-breasted Merganser, Mergus serrator, 56, 0.70, 0.00; Mallard, Anas platyrhynchos, 1,640, 0.56, 0.00; American Black Duck, Anas rubripes, 362, 0.60, 0.00; American Wigeon, Anas americana, 271, 0.59, 0.00; Green-winged Teal, Anas crecca, 333, 0.61, 0.00; Blue-winged Teal, Anas discors, 669, 0.62, 0.00; Cinnamon Teal, Anas cyanoptera, 231, 0.60, 0.00; Northern Shoveler, Spatula clypeata, 299, 0.61, 0.00; Northern Pintail, Anas acuta, 450, 0.59, 0.00; Wood Duck, Aix sponsa, 115, 0.58, 0.00; Red-breasted Merganser, Mergus serrator, 56, 0.70, 0.00; Mallard, Anas platyrhynchos, 1,640, 0.56, 0.00; American Black Duck, Anas rubripes, 362, 0.60, 0.00; American Wigeon, Anas americana, 271, 0.59, 0.00; Green-winged Teal, Anas crecca, 333, 0.61, 0.00; Blue-winged Teal, Anas discors, 669, 0.62, 0.00; Cinnamon Teal, Anas cyanoptera, 231, 0.60, 0.00; Northern Shoveler, Anas clypeata, 299, 0.61, 0.00; Northern Pintail, Anas acuta, 450, 0.60, 0.00; Wood Duck, Aix sponsa, 1,115, 0.58, 0.00; Redhead, Aythya americana, 195, 0.58, 0.01; Lesser Scapu, Aythya affinis, 213, 0.58, 0.01; Common Goldeneye, Bucephala clangula, 137, 0.61, 0.00; Bufflehead, Bucephala albeola, 80, 0.62, 0.02; Ruddy Duck, Oxyura jamaicensis, 204, 0.59, 0.00; Glossy Ibis, Plegadis falcinellus, 49, 0.65, 0.02; American Bittern, Botaurus lentiginosus, 616, 0.63, 0.00; Least Bittern, Ixobrychus exilis, 49, 0.65, 0.02; Great Blue Heron, Ardea herodias, 1,797, 0.54, 0.00; Little Blue Heron, Egretta caerulea, 380, 0.55, 0.02; Green Heron, Butorides virescens, 1,463, 0.56, 0.00; Black-crowned Night-Heron, Nycticorax nycticorax, 388, 0.64, 0.00; Yellow-crowned Night-Heron, Nyctanassa violacea, 227, 0.65, 0.00; King Rail, Rallus elegans, 54, 0.76, 0.00; Virginia Rail, Rallus limicol, 147, 0.64, 0.00; Sora, Porzana carolina, 450, 0.59, 0.00; Common Moorhen, Gallinula chloropus, 136, 0.65, 0.00; American Coot, Fulica americana, 500, 0.61, 0.00; American Woodcock, Scolopax minor, 321, 0.61, 0.00; Common Snipe, Gallinago gallinago, 878, 0.58, 0.00; Least Sandpiper, Calidris minuta, 17, 0.76, 0.02; Spotted Sandpiper, Actitis macularia, 989, 0.61, 0.00; Long-billed Curlew, Numenius americanus, 179, 0.58.
APPENDIX. Continued

Prothonotary Warbler, Protonotaria citrea, 401, 0.57, 0.00; Swainson’s Warbler, Limnothlypis swainsonii, 97, 0.70, 0.00; Blue-winged Warbler Vermivora pinus, 425, 0.56, 0.01; Golden-winged Warbler, Vermivora chrysoptera, 279, 0.61, 0.00; Nashville Warbler, Vermivora ruficapilla, 596, 0.61, 0.00; Tennessee Warbler, Vermivora peregrina, 295, 0.58, 0.00; Northern Parula, Parula americana, 840, 0.56, 0.00; Cape May Warbler, Dendroica tigrina, 191, 0.60, 0.00; Yellow Warbler, Dendroica petechia, 1,929, 0.57, 0.00; Black-throated Blue Warbler, Dendroica caerulescens, 359, 0.57, 0.01; Myrtle Warbler, Dendroica c. coronata, 529, 0.56, 0.00; Audubon’s Warbler, Dendroica c. auduboni, 324, 0.58, 0.00; Magnolia Warbler, Dendroica magnolia, 445, 0.58, 0.00; Cerulean Warbler, Dendroica cerulea, 262, 0.61, 0.00; Blackpoll Warbler, Dendroica striata, 165, 0.65, 0.00; Yellow-throated Warbler, Dendroica dominica, 416, 0.57, 0.00; Black-throated Green Warbler, Dendroica virens, 544, 0.57, 0.00; Hermit Warbler, Dendroica occidentalis, 70, 0.64, 0.01; Pine Warbler, Dendroica pinus, 685, 0.55, 0.00; Prairie Warbler, Dendroica discolor, 679, 0.55, 0.01; Northern Waterthrush, Seiurus noveboracensis, 537, 0.59, 0.00; Louisiana Waterthrush, Seiurus motacilla, 529, 0.61, 0.00; Kentucky Warbler, Oporornis formosus, 598, 0.54, 0.03; Mourning Warbler, Oporornis philadelphia, 473, 0.57, 0.00; Common Yellowthroat, Geothlypis trichas, 2,129, 0.57, 0.00; Yellow-breasted Chat, Icteria virens, 1,114, 0.56, 0.00; Hooded Warbler, Wilsonia citrina, 520, 0.58, 0.00; Wilson’s Warbler, Wilsonia pusilla, 485, 0.58, 0.00; American Redstart, Setophaga ruticilla, 1,116, 0.55, 0.00; House Sparrow, Passer domesticus, 2,275, 0.55, 0.00; European Tree Sparrow, Passer montanus, 13, 0.15, 0.99; Gray Catbird, Dumetella carolinensis, 1,739, 0.55, 0.00; Brown Thrasher, Toxostoma rufum, 1,737, 0.55, 0.00; Cactus Wren, Campylorhynchus brunneicapillus, 107, 0.60, 0.02; Rock Wren, Salpinctes obsoletus, 408, 0.57, 0.00; Canyon Wren, Catherpes mexicanus, 134, 0.65, 0.00; Bewick’s Wren, Thryomanes bewickii, 505, 0.60, 0.00; House Wren, Troglodytes aedon, 1,686, 0.54, 0.00; Sedge Wren, Cistothorus platensis, 337, 0.57, 0.01; Marsh Wren, Cistothorus palustris, 342, 0.61, 0.00; Brown Creeper, Certhia americana, 484, 0.64, 0.00; Pygmy Nuthatch, Sitta pygmaea, 93, 0.62, 0.01; Golden-crowned Kinglet, Regulus satrapa, 461, 0.62, 0.00; Ruby-crowned Kinglet, Regulus calendula, 569, 0.63, 0.00; Blue-gray Gnatcatcher, Polioptila caerulea, 1,077, 0.56, 0.00; Wood Thrush, Hylocichla mustelina, 1,355, 0.54, 0.00; Swainson’s Thrush, Catharus ustulatus, 631, 0.58, 0.00; American Robin, Turdus migratorius, 2,353, 0.54, 0.00; Eastern Bluebird, Sialia sialis, 1,451, 0.46, 1.00.