Received: 14 July 2025 | Revised: 5 February 2026 | Accepted: 16 March 2026

DOI: 10.1002/jwmg.70206

THE JOURNAL OF WILDLIFE MANAGEMENT | THE WILDLIFE SOCIETY

RESEARCH ARTICLE

Pooling information across sites and years to estimate bat fatalities at wind farms

Natalia Berberian[1] | **Philip M. Dixon**[2]

[1]Departamento de Biometría, Estadística y Computación, Facultad de Agronomía, Universidad de la República, Av. Garzón 780, Montevideo, Uruguay

[2]Statistics Department, Iowa State University, 2438 Osborn Dr, Ames, IA 50011, USA

Correspondence
Philip M. Dixon, Department of Statistics, Iowa State University, 2438 Osborn Dr, Ames, IA 50011, USA.
Email: pdixon@iastate.edu

Abstract

Precise estimation of bat mortality at wind farms is essential for conservation planning. Carcass detection probabilities, however, are often low and data sparse, leading to large uncertainties in fatality estimates. Precision can be improved by pooling information across multiple sites and years. Using post-construction monitoring data from 20 wind farms in Iowa, USA, we evaluated complete pooling (CP), no pooling (NP), and partial pooling (PP) approaches for estimating searcher efficiency, carcass persistence probability, and the fatality estimates obtained by integrating these detection-probability components. For searcher efficiency and persistence probability, PP produced estimates that balanced the extremes of CP and NP, yielding narrower credible intervals than NP, especially at data-limited sites, while preserving meaningful site-level differences. Fatality estimates for 3 bat species, differing in carcass abundance, were generally similar in the PP and NP models, but both were often different from CP. Partial pooling showed gains in precision, especially for abundant species and years with less monitoring effort. A simulation study of scenarios with different amounts of heterogeneity among sites showed that PP provided more accurate predictions than CP and NP in intermediate and high heterogeneity settings. Our findings highlight the value of PP as a flexible strategy for improving detection probability estimation under ecological uncertainty and support its broader application in wildlife impact assessment.

KEYWORDS
bat monitoring, Bayesian, evidence of absence, multiple sites, multiple years, partial pooling, site-year effects

The number and size of wind farms to generate renewable energy from wind power is increasing globally. Wind farms can be a significant source of wildlife mortality, particularly for bats, with fatalities documented globally (O'Shea et al. 2016), and may increase extinction risk for certain species (Frick et al. 2017).

Estimating mortality due to inshore wind farms generally begins by searching the ground around turbines and counting the carcasses found (Huso 2011). The number of carcasses counted differs from the number killed for 3 reasons: 1) searches are usually conducted on a subset of turbines or limited area around each turbine (Smallwood 2007, Strickland et al. 2011), 2) scavengers may remove carcasses before the area is searched, and 3) carcasses present during a search may not be detected. Hence, estimating mortality requires adjusting the number of found carcasses by the characteristics of carcass searches, the proportion of carcasses persisting unscavenged and observable until the next search, and the probability that a carcass present during a search is detected. These are commonly called the search area adjustment, the average probability of carcass persistence (carcass persistence), and the searcher efficiency, respectively. The product of search area adjustment, carcass persistence, and searcher efficiency is called the detection probability, and a robust and precise estimate of it and its uncertainty is crucial to adequately estimate the number of fatalities (Huso 2011).

Statistical methods to estimate bat mortality have generally been based on the Horvitz-Thompson estimator, perhaps with modifications (Pollock 2007, Hull and Muir 2010, Huso 2011), or Bayesian models (Huso et al. 2015, McDonald et al. 2021). Huso (2011) provided a Horvitz-Thompson estimator using data on search area adjustment, searcher efficiency, and carcass persistence collected from ancillary studies. Huso et al. (2015) developed a Bayesian model that incorporates uncertainty in detection probability. While the original focus of their work was on estimating fatalities when no dead individuals were found, this framework has informed broader approaches to fatality estimation. It has been used in applied monitoring contexts (Bay et al. 2017) and has subsequently been extended into more general hierarchical models that allow fatality rates to be related to site-level characteristics (McDonald et al. 2021).

McDonald et al. (2021) proposed a hierarchical regression model that combines data from multiple sites or years to estimate fatality rates. In their approach, the number of fatalities at each site is modeled as a Poisson process, with the expected fatality rate linked to site-specific covariates through a log-linear function. The number of carcasses found in searches is then modeled as a binomial outcome that depends on both the true number of fatalities and the probability of detecting a carcass. Because this detection probability depends on factors such as searcher efficiency and carcass persistence, it is usually estimated from independent field trials at each site. These provide site-specific estimates of searcher efficiency and carcass persistence. Alternatively, detection information from all sites can be pooled into a single estimate using weighted averages as described by Huso et al. (2015) and McDonald et al. (2021). The site-specific estimates can be described as no pooling (NP), and the single pooled all-site estimate can be described as complete pooling (CP).

A third way to estimate the detection probability for a site is to combine the pooled all-site estimate and its unpooled site-specific estimate (Efron and Morris 1977). This allows each site to have a different estimate while sharing information across sites by assuming they share a common distribution (Wikle 2003, Royle and Dorazio 2008). One way to implement this approach uses a mixed model with a fixed effect for the pooled mean across all sites, a random effect for the variability among sites, and a random error for the site-specific uncertainty in the estimated detection probability. These mixed-model estimates can be described as partial pooling (PP) because they incorporate shrinkage to the mean (Efron and Morris 1977) to compromise between pooled and unpooled estimates.

We used mixed models to partially pool searcher efficiency and carcass persistence information across sites in an analysis of bat mortality at wind farms. Partial pooling takes advantage of collective data to produce more precise estimates, even when individual groups have sparse or noisy data (Gelman and Hill 2007, McElreath 2020).

We focused on searcher efficiency and carcass persistence because site-specific estimates (i.e., NP) are often estimated from small data sets. We assessed the variability of components of detection probability across sites, then evaluated the influence of the pooling method on estimates of bat mortality. Using post-construction monitoring data from 20 wind farms in Iowa, USA, we compared searcher efficiency and carcass persistence estimates from the 3 pooling methods (objective 1) and then examined how these approaches affected resulting fatality estimates (objective 2). Finally, we used simulated scenarios with varying levels of heterogeneity among sites to evaluate how these pooling specifications (i.e., CP, NP, and PP) performed under controlled conditions (objective 3).

STUDY AREA

The 20 wind energy facilities are found in central and western Iowa (Figure S1). This area is in the Western Corn Belt Plains level 3 ecoregion and has a humid continental climate, with annual precipitation of 700–900 mm and average annual temperature of 7–9°C. The natural vegetation is primarily tall grass prairie dominated by big bluestem (*Andropogon gerardii*), Indian grass (*Sorghastrum nutans*), and sweetgrass (*Panicum virgatum*), with forested areas along many rivers and creeks. Nine species of bats are regularly found in Iowa. McDonald et al. (2021) describe the procedures for monitoring bat carcasses.

METHODS

Preliminary evidence

Final reports of post-construction monitoring studies of bat fatalities are often required as part of wind energy installation projects. These requirements typically stem from regulations or agreements designed to ensure compliance with environmental protection laws and guidelines (Thaxter et al. 2017, European Commission 2011, U.S. Fish and Wildlife Service 2012). Several post-construction monitoring technical reports are publicly available and, specifically for Midwest area projects, information regarding bat fatality is summarized by year and species. These reports provide information on the searcher efficiency and carcass persistence trials at each site and year. We obtained data from post-construction monitoring reports of 20 operating wind-power projects in Iowa between 2015 and 2017 (Table 1; Figure S1). Of these 20 projects, 6 were monitored only in 2015, 9 only in 2016, 3 in both 2015 and 2016, and 2 only in 2017. We consider the 3 sites monitored twice as 6 site-year combinations.

We used the searcher efficiency data from the 12 sites monitored in 2016 to illustrate the heterogeneity among sites (Figure 1). We obtained searcher efficiency estimates as maximum-likelihood estimates from binomial models based on the number of successful detections out of the total number of trials, with 90% normal-approximation confidence intervals. The overlap among many of the confidence intervals suggested that treating sites as independent (no pooling) may not fully use the information available across sites. Conversely, the large differences observed for some sites indicated that assuming a single value for all sites (complete pooling) was not appropriate. These patterns suggested that using partial pooling to share information across sites could reduce uncertainty in fatality estimates.

Effect of pooling choice on searcher efficiency and carcass persistence

Modeling searcher efficiency

For objective 1, we first assessed how different degrees of information sharing across sites and years (i.e., CP, NP, and PP) influence searcher efficiency estimates using post-construction monitoring data. Searcher efficiency quantifies the

TABLE 1 Bat carcass monitoring data from 20 wind projects located in Iowa, USA, from 2015 to 2017. Search area adjustment (SAA) is the value obtained from post-construction reports for each site (based on the road and pad search area). Searcher efficiency (SE) columns include the number of carcasses placed (n_{SE}) and observed (OC_{SE}) during SE studies carried out at each site. Average probability of carcass persistence (PCP) columns provide data from carcass persistence trials, including the reported average carcass persistence time ($\hat{\theta}$), the number of days between 2 standardized carcass searches conducted in each project (d), the number of carcasses placed as evidence in each project area (n_{PCP}), and the estimate of the total number of carcasses actually observed in each trial ($n_{PCP_{adj}}$) based on the information for observed and censored observations. Project sites are ordered alphabetically. Data extracted from <https://www.regulations.gov/document/FWS-R3-ES-2018-0037-0107> and <https://www.regulations.gov/document/FWS-R3-ES-2018-0037-0104> (accessed 20 Feb 2024).

			SE		PCP			
Project	Year	SAA	n_{SE}	OC_{SE}	$\hat{\theta}$	d	n_{PCP}	$n_{PCP_{adj}}$
Adair	2015	0.10	50	43	2.94	7	50	45
Adams	2016	0.17	115	80	3.3	3.5	116	76
Carroll	2015	0.08	50	44	2.52	7	50	47
Century	2016	0.16	107	45	6.90	3.5	120	48
Charles City	2016	0.15	104	57	4.07	3.5	117	68
Eclipse	2015	0.09	50	49	2.59	7	50	47
Highland	2016	0.13	121	66	3.37	3.5	119	77
Ida Grove	2017	0.13	116	80	11.55	3.5	110	29
Intrepid	2016	0.07	124	59	9.43	3.5	118	37
Laurel	2016	0.17	109	44	3.90	3.5	119	70
Lundgren	2015	0.07	55	51	2.50	7	50	47
Lundgren	2016	0.16	113	49	6.6	3.5	124	51
Macksburg	2015	0.07	52	44	1.92	7	50	49
Macksburg	2016	0.35	114	49	3.20	3.5	110	73
Morning Light	2015	0.09	50	42	2.32	7	50	48
Obrien	2017	0.08	106	79	3.06	7	50	45
Pomeroy	2016	0.13	117	46	4.47	3.5	117	64
Rolling Hills	2015	0.07	50	45	5.29	7	50	37
Rolling Hills	2016	0.15	115	56	2.70	3.5	115	84
Victory	2015	0.10	50	39	2.53	7	50	47
Vienna I&II	2016	0.24	118	45	6.63	3.5	110	45
Walnut	2015	0.09	50	43	1.97	7	48	47
Wellsburg	2016	0.20	112	59	3.13	3.5	118	79

probability that a carcass on the ground is found during a search. This is estimated by placing carcasses in the search area and counting the number found during a search. Different search protocols were used in 2015, 2016, and 2017, so the mean searcher efficiency is expected to differ between years. Carcasses are assumed to be found independently with a constant probability. For site i, the model for the number of observed carcasses (OC_i) is:

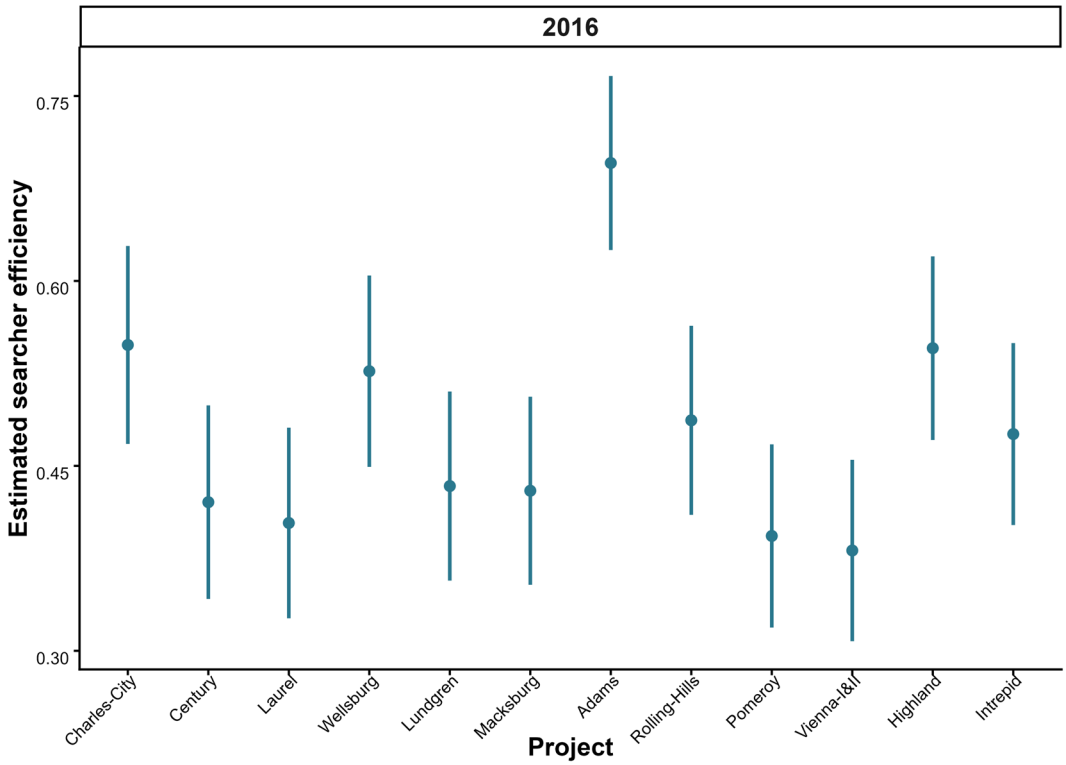

FIGURE 1 Median and 90% confidence interval limits for maximum likelihood estimates of carcass searcher efficiency in 2016, by wind energy project site in Iowa, USA. Project sites are ordered by increasing number of carcasses placed for each search.

$$OC_i \mid SE_i \sim \text{Binomial}(n_i,\ SE_i), \tag{1}$$

where n_i is the number placed in the search area, and SE_i is the searcher efficiency for site i. The CP, NP, and PP approaches use different models for SE_i. The CP model assumes the same searcher efficiency at all sites within a year, modeled as:

$$\text{logit}(SE_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_2 x_{3i}, \tag{2}$$

where x_{1i}, x_{2i}, and x_{3i} are indicator variables for whether site i was surveyed in 2015, 2016, or 2017. The NP model assumes each site has a different searcher efficiency:

$$\text{logit}(SE_i) = \beta_{0i}. \tag{3}$$

The PP model assumes the site-specific searcher efficiencies are samples from a distribution with year-specific means and a random effect for the heterogeneity among sites:

$$\text{logit}(SE_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_2 x_{3i} + \varepsilon_i, \tag{4}$$

where x_{1i}, x_{2i}, and x_{3i} are indicator variables, and $\varepsilon_i \sim \text{Normal}\left(0,\ \sigma^2_{SE}\right)$ accounts for random site-level variations in searcher efficiency, with precision $\tau_{SE} = 1/\sigma^2_{SE}$. We chose to model logit(SE_i) for CP, NP, and PP specifications to

maintain consistency across the models. We used weakly informative prior distributions for all parameters. Specifically, for β_1, β_2, and β_3 we assigned independent Normal(0, 3) priors, and $\tau_{SE} \sim \text{Gamma}(1, 1)$, using the Gamma (shape, rate) parametrization. Alternative choices for priors produced similar posterior distributions for each SE_i.

Modeling average probability of carcass persistence

For the second part of objective 1, we assessed the effect of pooling choice (i.e., CP, NP, or PP) on probability of carcass persistence. A recent bat fatality occurring between scheduled carcass searches will not be detected until the next search. During this interval, the carcass may decompose or be removed by scavengers. The average probability of carcass persistence is the probability that a carcass arriving at a random time before the next carcass search remains detectable at that search. This is commonly modeled by an exponential decay model (Huso 2011, Huso et al. 2015). Assuming carcasses disappear at a constant rate over time, the number of days carcass j at site i remains detectable follows an exponential distribution with mean carcass persistence time (θ_i): $CRT_{ij} \sim \text{Exponential}(1/\theta_i)$ (Huso 2011). Post-construction monitoring reports provide the reported average ($\hat{\theta}_i$) and the number of carcasses used to obtain that average, from which the estimation variance of $\hat{\theta}_i$ is ${\theta_i}^2/n_i$ when there are no censored observations. Because post-construction monitoring reports did not include the raw data, we chose to model $\log(\hat{\theta}_i)$ as a normal distribution. Removals were monitored for a limited period: 7 days in 2015, 3.5 days in 2016, and both 3.5 and 7 days in 2017. Carcasses not removed by the end of the study have right-censored carcass retention time. The variance of the estimated mean of an exponential distribution with right-censored values is $\theta^2/n_{PCP_{adj}}$ (Cox and Oakes 1984), where $n_{PCP_{adj}}$ is the number of non-censored observations for modeling the probability of carcass persistence (PCP). We estimated $n_{PCP_{adj}}$ using the probability that a carcass is removed before the censoring limit (i.e., 3.5 or 7 days); this is $n_{PCP_{adj}} = [n_{PCP,i}][1 - \exp(-d_i/\hat{\theta}_i)]$, where $n_{PCP,i}$ is the number of carcasses placed on the ground at site i, $\hat{\theta}_i$ is the reported average carcass persistence time, and d_i is the censoring limit (Table 1). The delta method (Cramér 1946) variance of $\log(\hat{\theta}_i)$ is then $[{\hat{\theta}_i}^2/n_{PCP_{adj},i}][(1/\hat{\theta}_i)^2] = 1/n_{PCP_{adj},i}$. We assumed that the log-transformed average carcass persistence times were normally distributed with mean ϕ_i and an estimation error variance of $1/n_{PCP_{adj},i}$:

$$\log(\hat{\theta}_i) \sim \text{Normal}\left(\phi_i, \frac{1}{n_{PCP_{adj},i}}\right). \tag{5}$$

The CP, NP, and PP approaches use different models for ϕ_i. The CP model assumes that ϕ_i is the same for all sites within a year, modeled as:

$$\phi_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_2 x_{3i}, \tag{6}$$

where x_{1i}, x_{2i}, and x_{3i} are year indicator variables. The NP model assumes each site has a different ϕ_i:

$$\phi_i = \beta_{0i}. \tag{7}$$

The PP model assumes the site-specific searcher efficiencies are samples from a distribution with year-specific means and a random effect for the heterogeneity among sites:

$$\phi_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_2 x_{3i} + \varepsilon_i, \tag{8}$$

where x_{1i}, x_{2i}, and x_{3i} are the year indicator variables and $\varepsilon_i \sim \text{Normal}(0, \sigma^2_{PCP})$ accounts for random differences among sites in ϕ_i, with precision $\tau_{PCP} = 1/\sigma^2_{PCP}$. We specified weakly informative prior distributions for all parameters. Specifically, for β_1, β_2, and β_3 we assigned independent Normal(0, 2) priors, and $\tau_{PCP} \sim \text{Gamma}(1, 1)$, using

the parametrization Gamma(shape, rate). We calculated, separately for CP, NP, and PP, the probability (PCP_i) that a carcass deposited at a uniformly distributed time between searches is still on the ground at site i and search time t, as (Huso 2011):

$$PCP_i = [\exp(\hat{\phi}_i)/t][1 - \exp(t/\exp(\hat{\phi}_i))]. \quad (9)$$

Effect of pooling choice on the estimated number of bat fatalities

For objective 2, we analyzed the estimated number of fatalities using the same 3 pooling models (i.e., CP, NP, and PP) evaluated for searcher efficiency and carcass persistence previously. We focused on 3 bat species found in Iowa, USA: little brown bat (*Myotis lucifugus*), considered Endangered by the International Union for Conservation of Nature (IUCN; Solari 2021); evening bat (*Nycticeius humeralis*), categorized as Least Concern (Solari 2019); and hoary bat (*Lasiurus cinereus*), also categorized as Least Concern (Gonzalez et al. 2016). We selected these 3 species to capture a gradient in carcass-count variability across site-year searches. Little brown bats exhibited frequent zero counts (12 out of 23 site-year combinations), evening bats showed moderate occurrence of zeros (3 out of 23), and hoary bats had no zero counts recorded. This gradient allowed us to assess model performance under differing levels of bat abundance (see Data S1 in the Supporting Information).

We estimated the site-specific number of fatalities using the N-mixture model:

$$g_i = (SAA_i)(SE_i)(PCP_i) \quad (10.1)$$

$$C_i \sim \text{Binomial}(M_i, g_i) \quad (10.2)$$

that relates the number of found carcasses (C_i) to the number of fatalities (M_i) and the detection probability (g_i). The detection probability is the product of the search area adjustment (SAA_i), searcher efficiency (SE_i), and carcass persistence (PCP_i).

We performed 3 analyses, each applying the same pooling specification (i.e., CP, NP, or PP) to both searcher efficiency and the carcass persistence. Throughout all analyses we treated search area adjustment as a fixed value. We modeled the number of fatalities (M_i) using a Poisson distribution with mean λ_i and assigned to $\log(\lambda_i)$ an independent Normal(0, 10) prior for each site-year combination. This formulation corresponds to an NP structure, with each λ_i estimated separately and no hierarchical sharing of information across site-years.

Effect of pooling choice on simulated scenarios with varying levels of heterogeneity among sites

The results from our second objective (effect of pooling choice on estimated bat fatalities) did not allow us to determine which estimator was closest to the true mortality because the true mortality was unknown. For objective 3, we addressed this limitation by simulating data for which the true mortality was known. To ensure comparability with previous analyses, we estimated mortality under 3 pooling specifications (i.e., CP, NP, and PP) using simulated searcher efficiency and carcass persistence data and the modeling approach described in objective 1. We assessed model performance in terms of predictive accuracy under scenarios with different levels of site-to-site variability in searcher efficiency and carcass persistence.

For each site, we simulated the true searcher efficiency (SE_i) and the true mean carcass persistence time (θ_i). We assumed these to be:

$$SE_i \sim \text{Beta}(\mu_{SE} v_{SE},\ [1 - \mu_{SE}] v_{SE}) \tag{11.1}$$

$$\theta_i \sim \text{Gamma}\left(\frac{\mu_{PCP}^2}{\sigma_{PCP}^2},\ \frac{\mu_{PCP}}{\sigma_{PCP}^2}\right) \tag{11.2}$$

The mean and variance of SE_i are μ_{SE} and $\mu_{SE}(1 - \mu_{SE})/(1 + v_{SE})$, and of θ_i are μ_{PCP} and σ_{PCP}^2, respectively. We considered 3 scenarios with high, intermediate, and no heterogeneity among sites. We achieved this decreasing gradient of heterogeneity among scenarios by assigning zero, intermediate, and high values to v_{SE} (eq. 11.1) and high, intermediate, and zero values to σ_{PCP}^2, respectively (eq. 11.2). We based mean values for the distributions of searcher efficiency and carcass persistence on summaries of the Iowa wind-projects data (McDonald et al. 2021). We then calculated the true probability of carcass persistence for each site (PCP_i) from ϕ_i using Equation (9). We set the search area adjustment (SAA) to a constant. The true detection probability (g_i) for each site is then calculated (eq. 10.1) as $g_i = (SAA_i)(SE_i)(PCP_i)$.

We also simulated the true number of carcasses at each site as $M_i \sim \text{Poisson}(\lambda_i)$. The site-specific rate (λ_i) varied among sites as $\log(\lambda_i) \sim \text{Normal}(\log[1{,}400],\ 0.3)$. We chose parameters for $\log(\lambda_i)$ to mimic conditions for a relatively abundant species such as hoary bat. We used the same model for M_i for all 3 scenarios.

We then simulated data for each site: the observed number of carcasses (C_i), the number of carcasses found during a searcher efficiency trial (X_i), and the carcass persistence times (T_{ij}) from a carcass persistence trial. We assumed the following distributions:

$$C_i \sim \text{Binomial}(M_i,\ g_i) \tag{12.1}$$

$$X_i \sim \text{Binomial}(n_{SE},\ SE_i) \tag{12.2}$$

$$T_{ij} \sim \text{Exponential}(\theta_i)\ j = 1, \ldots,\ n_{PCP}. \tag{12.3}$$

For each scenario, we simulated 1,000 data sets, each with 20 sites. Each data set had a different set of parameters (M_i, SE_i, and θ_i) and data (C_i, X_i, and T_{ij}). For each data set, we fit the CP, NP, and PP models and estimated site-specific $\hat{M}_i$ for each model.

Fitting the partial, complete, and no pooling models

We used the rjags package (Plummer 2023) to implement a Bayesian framework and fit models through Markov Chain Monte Carlo (MCMC) sampling in software R (R Core Team 2023) to achieve all 3 objectives (R code available in Supporting Information). We considered 4 MCMC chains for each model, and assessed convergence using the Gelman-Rubin diagnostic, also known as the $\hat{R}$ statistic (Gelman and Rubin 1992), and by visually inspecting trace plots following the guidelines outlined by Gelman et al. (2013). To reduce within-chain autocorrelation and achieve satisfactory convergence across chains, we used different combinations of burn-in, iterations, and thinning depending on the objective (objective 1: 10,000 burn-in, 5,000 iterations, no thinning; objective 2: 20,000 burn-in, 40,000 iterations, thinning of 5; and objective 3: 10,000 burn-in, 10,000 iterations, thinning of 5).

We evaluated model performance using the leave-one-out cross-validation information criterion (LOOIC). This approach provides robust measures of predictive precision and penalizes for model complexity, making it well-suited

for Bayesian model comparison (Gelman et al. 2014, Vehtari et al. 2017). For objectives 1 and 2, we evaluated the precision of posterior summary measures for the model parameters using 90% credible interval width. For objective 3, where the true number of fatalities was known, we computed 4 statistics for predicted fatalities: root mean square error of the predicted median (RMSE), width of the 90% credible interval (CI) and proportion of data sets where that CI included the true value (empirical coverage), and proportion of data sets where the upper bound of that CI was below the true value (empirical coverage of the 95% one-sided upper confidence bound).

One aspect of the worth of a statistical method is whether it provides a narrower CI, so long as that confidence interval retains the desired coverage. This worth, relative to a standard method, can be quantified by the ratio of 2 confidence interval widths. This can be expressed in practical terms by comparing sample sizes that provide the same precision (Cochran and Cox 1957). For 2 methods with CI widths of w_1 and w_2, calculated from the same sample size n, the sample sizes n_1 and n_2 that provide equally wide CIs are $n_2 = n_1(w_2/w_1)^2$. This assumes that the CI width is proportional to the standard error of the estimate, which is the case when the CI is symmetric around the estimate.

We did not set a random seed in any of the analyses conducted. We ran all models repeatedly during development, and the results showed a high degree of repeatability with different seeds; numerical values varied only minimally across runs, while the qualitative patterns, model rankings, and overall conclusions remained unchanged. If the analyses are rerun, the exact numerical results will differ slightly, but the overall patterns and results should not.

RESULTS

Effect of pooling choice on searcher efficiency and carcass persistence

Our findings showed that both searcher efficiency and carcass persistence differed between years when estimating posterior medians (Figures 2 and 3). Complete pooling estimates were equal within each year, NP estimates showed considerable variability, and PP estimates fell between the CP and NP values. To illustrate this pattern for searcher efficiency, we highlight results for Adams and Intrepid in 2016. For Adams, the estimates were 0.479 (CP), 0.694 (NP), and 0.668 (PP); for Intrepid, all 3 models yielded similar values (0.479 [CP], 0.475 [NP], and 0.475 [PP]) because NP was nearly identical to CP (Figure 2). We observed similar patterns for carcass persistence. For example, in 2016 at Intrepid, estimates were 1.407 (CP), 2.218 (NP), and 2.151 (PP), with PP estimates shrunk toward the CP mean. By contrast, estimates at Charles City that year were nearly identical across models (1.407 [CP], 1.393 [NP], and 1.401 [PP]). Overall, carcass persistence showed greater variability in posterior median estimates between sites compared to searcher efficiency (Figure 3). The impact of sample size was clear within each year, particularly for carcass persistence (Figures 2 and 3). For example, in 2016, sites with less data, such as Intrepid, showed more extreme NP estimates that deviated further from the yearly average, represented by CP (Figure 3). In contrast, PP pulled these estimates toward the overall mean, especially compared to sites like Charles City, where more data were available.

We assessed precision of the searcher efficiency and carcass persistence estimates through 90% credible interval (CI) width. As expected, the CP model produced the narrowest CIs (Figures 2 and 3), as it assumed that all data came from the same underlying population and did not account for variability among sites. Hence, we focused on the comparison between the PP and NP estimates (Tables S1 and S2). For searcher efficiency in 2015, CIs from the PP model were 22–45% narrower than those from the NP except for one site where both models yielded similar widths. Credible intervals for searcher efficiency were 2–12% narrower at sites measured only in 2016 or 2017 (Table S1). For carcass persistence, CIs under PP were narrower for all sites, but reductions ranged from 1% to 5% except at the 3 sites with 2 years of data, where PP resulted in reductions from 12% to 29% (Table S2). For example, the 90% CI widths for searcher efficiency at Adair in 2015 were 0.124 for PP and 0.160 for NP, both

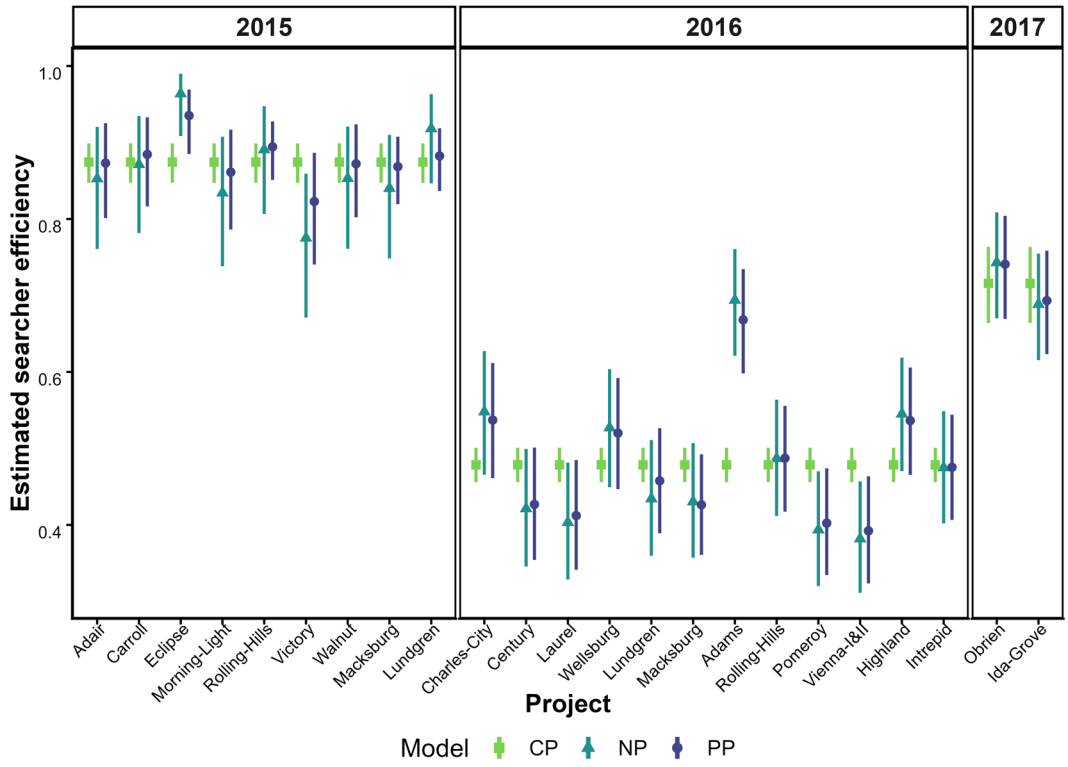

FIGURE 2 Median and 90% credible interval (CI) for carcass searcher efficiency estimates from 3 models, complete pooling (CP), no pooling (NP), and partial pooling (PP), by year (2015–2017) and wind energy project site in Iowa, USA. Within each year, project sites are ordered by increasing number of carcasses placed for each search.

based on a trial for searcher efficiency with 50 carcasses (Table S1; Figure S2). To obtain the same CI width as the NP trial with 50 carcasses, the PP model would require a searcher efficiency trial with only 30 carcasses, a decrease of 39% (Table S1; Figure S2). We observed this behavior across all other site-year combinations, except for Eclipse in 2015, with reductions in required searcher efficiency trial sample sizes ranging from 5% to 71% (Table S1). We observed a comparable gain for carcass persistence at the same site and year. The 90% CI widths were 0.47 for PP and 0.49 for NP at Adair, both based on a trial with 45 carcasses. Achieving the NP CI width with the PP model would require a carcass persistence trial with 42 carcasses, corresponding to a 6% reduction in sample size (Table S2; Figure S2). Across the remaining site-year combinations, PP consistently reduced the number of carcasses needed to attain a given CI width, with reductions ranging from 2% to 51% (Table S2).

We evaluated the predictive performance of each model using the LOOIC, an aggregate measure that integrates model fit across all sites, along with its associated standard error (Table 2). The relative performance of pooling strategies differed between searcher efficiency and carcass persistence probability. The results suggest that the PP model provided improved predictive performance relative to both CP and NP for searcher efficiency. Although the comparison between PP and CP was associated with higher uncertainty, the overall pattern consistently favored the PP approach (Table 2). In contrast, model performance for carcass persistence varied across pooling strategies. While the PP model performed better than CP, it did not outperform the NP model. The comparison between PP and CP was characterized by substantial uncertainty (Table 2). Overall, these results indicate that PP improves predictive performance for searcher efficiency, whereas carcass persistence appears to be better represented by site-specific estimates.

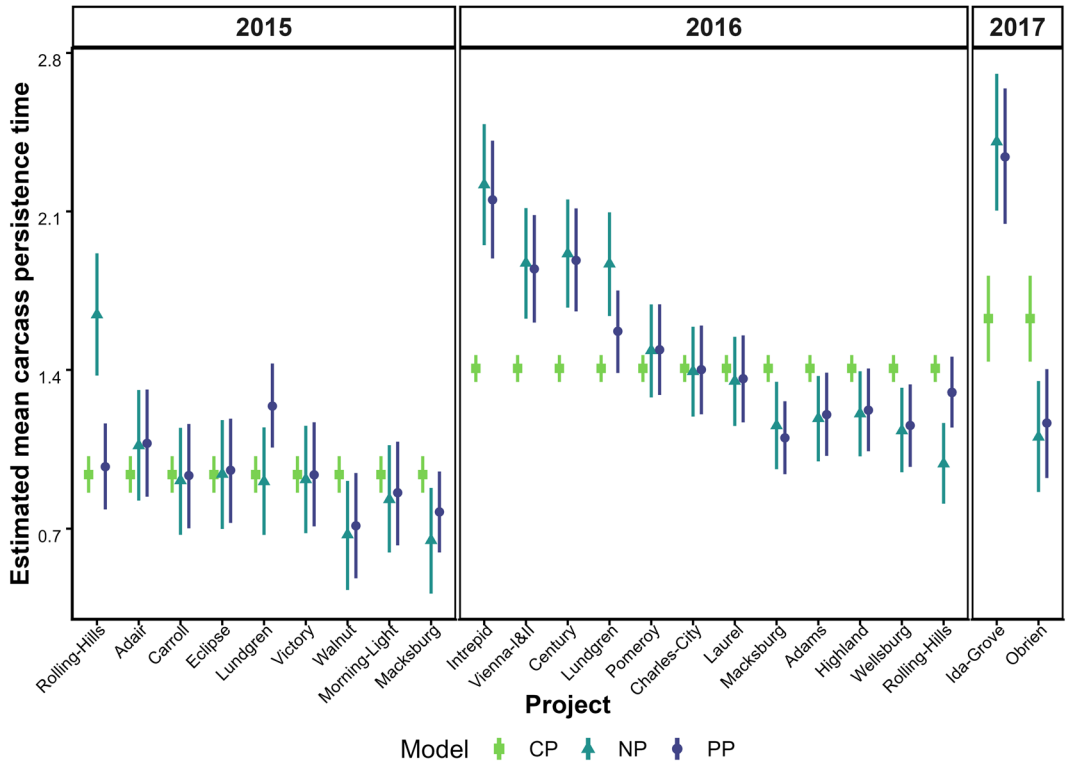

FIGURE 3 Median and 90% credible interval (CI) for mean carcass persistence time (days) estimates from 3 models, complete pooling (CP), no pooling (NP) and partial pooling (PP), by year (2015–2017) and wind energy project site in Iowa, USA. Within each year, project sites are ordered by the increasing adjusted number of carcasses placed per search.

TABLE 2 Leave-one-out information criterion (LOOIC), with standard error in parentheses, used to compare the predictive performance of 3 modeling approaches, complete pooling (CP), no pooling (NP) and partial pooling (PP), using bat carcass data from 20 wind projects in Iowa, USA (2015-2017). We evaluated models based on their ability to estimate the detection probability components of searcher efficiency (SE) and average probability of carcass persistence (PCP) and the total number of fatalities for 3 bat species: little brown (LBBA; *Myotis lucifugus*), evening (EVBA; *Nycticeius humeralis*), and hoary (HOBA; *Lasiurus cinereus*). Smaller LOOIC values indicate better model predictive performance.

	LOOIC or ΔLOOIC (standard error)				
Model	**SE**	**PCP**	**LBBA**	**EVBA**	**HOBA**
CP	168.6 (24.8)	138.2 (47.1)	67.0 (14.4)	113.7 (10.2)	193.9 (3.3)
NP	157.8 (4.4)	7.5 (1.4)	66.8 (14.2)	115.4 (11.4)	190.2 (3.1)
PP	146.3 (5.8)	74.1 (36.7)	66.6 (14.3)	112.7 (10.3)	191.2 (3.4)
PP vs. CP[a]	22.3 (10.9)	64.1 (47.2)	0.4 (0.8)	1.0 (1.6)	2.7 (2.6)
PP vs. NP[a]	11.5 (1.8)	−66.6 (36.6)	0.2 (1.0)	2.7 (3.6)	−1.0 (2.4)

[a]For comparison rows, we present the difference in LOOIC (ΔLOOIC) between models (and standard error for these differences); positive values favor the first model in each comparison. Differences in LOOIC < 2 indicate negligible differences, 4–7 moderate support, and >10 strong support for the model with the lower LOOIC (see Vehtari et al. 2017 for interpretation).

Effect of pooling choice on the estimated number of bat fatalities

For most sites, the estimated median number of hoary bat fatalities was relatively similar for the NP and PP models, while the CP results were often different, especially in 2016 and 2017 (Figure 4). For little brown bats (Figure S3) and evening bats (Figure S4), the 3 pooling models gave more similar estimates of median fatalities, although when there were differences, CP was frequently the most different. For example, at Laurel (2016), median estimated fatalities for hoary bat were similar for NP (420 fatalities) and PP (413 fatalities) and more different for CP (348 fatalities). Little brown bat had a similar pattern: estimated median fatalities of 31 and 30 for NP and PP, respectively, and 26 for CP, while estimates were nearly identical for evening bat (12, 11, and 10 for NP, PP, and CP, respectively).

The 90% credible intervals for hoary bat fatalities from the PP model were generally shorter than those for the NP model (Figure S5). This pattern was more evident when searcher efficiency and carcass persistence trials had smaller sample sizes. In 2015, when monitoring effort was lowest, PP credible intervals were narrower than those for NP in 5 out of 9 sites. In contrast, in 2016 and 2017, years with more intensive sampling schemes, the differences between the 2 models decreased, resulting in very similar interval widths. The CP model had the smallest CI widths, but it ignored the heterogeneity among sites.

Model comparisons based on the LOOIC revealed minimal differences in predictive precision across CP, NP, and PP models for all species. The small differences observed were within the margin of error, indicating no clear advantage of one model over another in terms of overall predictive performance (Table 2).

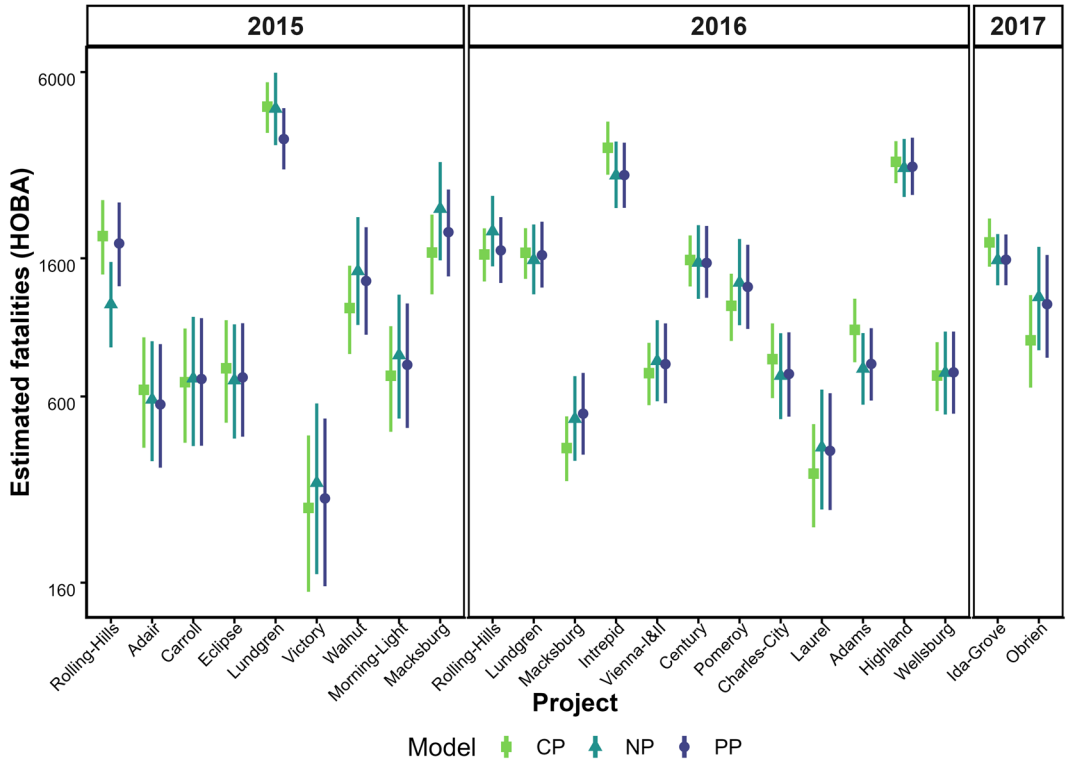

FIGURE 4 Median and 90% credible interval (CI) for total number of hoary bat (HOBA; *Lasiurus cinereus*) fatalities estimated from 3 models, complete pooling (CP), no pooling (NP), and partial pooling (PP), by year (2015–2017) and wind energy project site in Iowa, USA. Within each year, projects are ordered alphabetically.

Effect of pooling choice on simulated scenarios with varying levels of heterogeneity among sites

In the high and intermediate heterogeneity scenario, the PP model had the lowest RMSE and the CP model the highest (Figure 5). In the high heterogeneity scenario, the NP model had slightly larger RMSE than the PP model. When the variability in searcher efficiency and carcass persistence was twice that in the high heterogeneity scenario, the RMSE values for PP and NP were nearly identical (results not shown). In the no heterogeneity scenario, the CP model had the smallest RMSE, the NP model had the largest, and the PP model was in between. In all 3 simulated scenarios, median CI widths were larger for NP (1,100–1,171) than for PP (987–1,075; Table 3). The standard errors of the CI width estimates (all between 1.2 and 3.9) were much smaller than the differences between NP and PP models. Despite the narrower CIs for PP, empirical coverages of nominal 90% CIs for both PP and NP were close to 90% (89.0–89.6% for NP and 89.2–91.3% for PP). Empirical coverages of 95% upper bounds were slightly below the nominal 95% for both NP (93.4–93.7%) and PP (92.6–94.1%). In contrast, median CI widths for the CP model were consistently smaller than those for NP and PP models (Table 3), but CP intervals no longer had the desired coverage when there was heterogeneity among sites, i.e., for the intermediate (empirical coverage of 62.4%) and high (empirical coverage of 42.9%) scenarios.

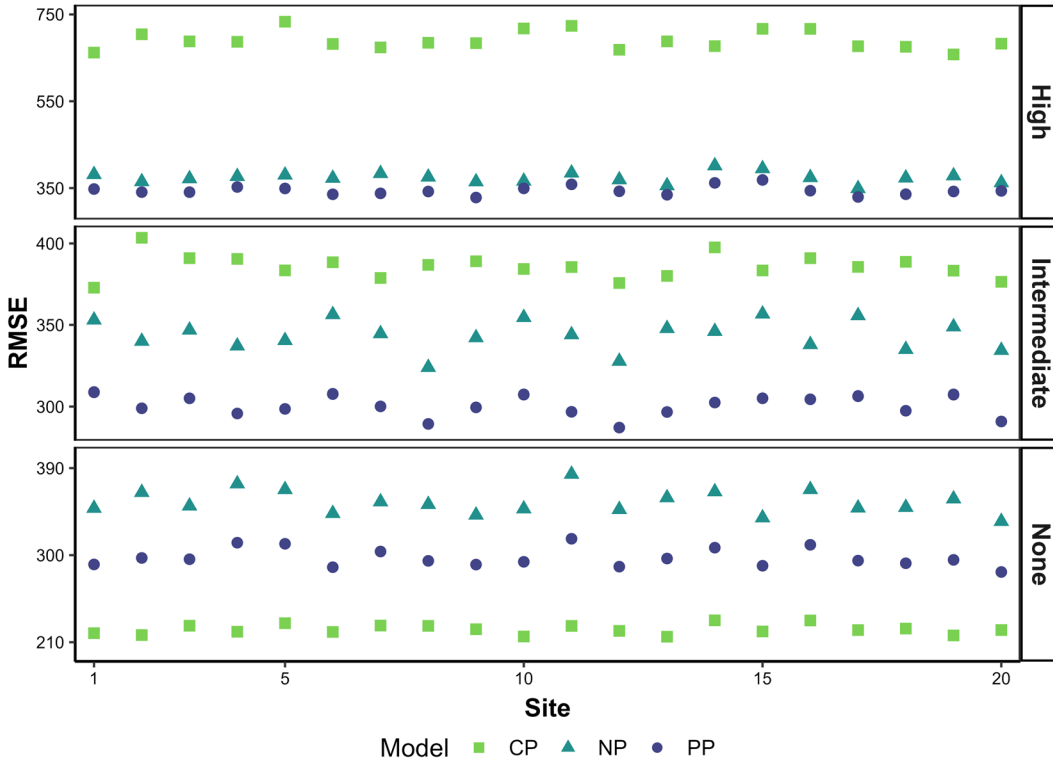

FIGURE 5 Root mean square error (RMSE) of the estimated true number of bat fatalities (M_i) in 20 simulated wind energy sites under 3 pooling specification models, complete pooling (CP), no pooling (NP), and partial pooling (PP), across 3 simulated scenarios representing decreasing levels of site-level heterogeneity: high, intermediate, and none.

TABLE 3 Median 90% credible interval (CI) width for the estimated true number of simulated bat fatalities at wind projects (M_i), the proportion of simulated data sets where the CIs included the true value (EC; empirical coverage), and the proportion of simulated data sets where the upper bound of the CIs was below the true value (EUB; empirical coverage of the 95% 1-sided upper confidence bound), under 3 pooling specifications, complete pooling (CP), no pooling (NP), and partial pooling (PP), across 3 simulated scenarios representing decreasing levels of site-level heterogeneity (high, intermediate and none). Standard error of each CI width estimate is <3.9; standard errors for EC and EUB are <1% and 0.7%, respectively.

Scenario	Model	CI width	EC (%)	EUB (%)
High	CP	728	42.9	71.4
	NP	1171	89.3	93.4
	PP	1075	89.2	92.6
Intermediate	CP	648	62.4	77.8
	NP	1100	89.0	93.6
	PP	987	89.8	93.2
None	CP	694	88.3	91.2
	NP	1139	89.6	93.7
	PP	1018	91.3	94.1

DISCUSSION

Partial pooling (PP) consistently balanced the extremes of complete pooling (CP) and no pooling (NP), reducing excess variability in the estimates while retaining meaningful differences among sites. In the estimation of searcher efficiency and carcass persistence, CP produced equal estimates within each year, although site-specific estimates varied considerably across years. In contrast, the NP model often generated highly variable site-level estimates, particularly when data were limited. Partial pooling moderated these extremes, providing narrower credible intervals than NP, especially where monitoring effort was lower.

When estimating fatalities from the post-construction monitoring data, the NP and PP models frequently had similar median estimated fatalities. The gains in precision using PP instead of NP were largest for the most abundant species (hoary bat) in the year with smaller sample sizes for searcher efficiency and carcass persistence trials (2015). Simulation results reinforced these empirical patterns. Across the simulated scenarios, PP consistently produced narrower credible intervals than NP while both maintained empirical coverage close to the nominal level and had slightly conservative upper bounds.

Complete pooling generated the narrowest intervals, but those intervals had reduced empirical coverage for both the 90% 2-sided interval and 95% upper bound. Both the NP and PP intervals had the appropriate coverage for 2-sided intervals and both had slightly conservative 1-sided upper bounds. Together, these results highlight the PP model as a flexible strategy that effectively shares information across sites while preserving biologically meaningful variability in bat fatality estimation.

The partial pooling estimates are best linear unbiased predictions (BLUPs; Robinson 1991). As such, they are a compromise between the overall mean (CP estimate) and the site-specific (NP) estimate, a statistical phenomenon called shrinkage (Efron and Morris 1977, McElreath 2020). The relative weight given to the site-specific estimate depends on the precision of the site-specific estimate and the heterogeneity among the sites (Efron and Morris 1977). When the site-specific estimates are precisely estimated and very different among sites, BLUPs are very close to the site-specific estimates and the shrinkage is minimal (Robinson 1991). These patterns are illustrated

by the searcher efficiency data (Figure 2). Conversely, when site-specific estimates have large standard errors and sites have very similar estimates, the BLUPs lie very close to the overall mean, resulting in considerable shrinkage. Sample sizes for searcher efficiency trials were higher in 2016 and 2017 than in 2015 (Table 1). Consequently, PP estimates of searcher efficiency in 2015 were both further from and more precise than the NP estimates.

Our data set has monitoring data from 1 year at most sites, so our models emphasized the variability among sites. There are 2 years of data from Lundgren, Macksburg, and Rolling Hills. Our partial pooling model ignored any variation between years at those sites, so the PP estimates are shrunk towards each site's average over 2 years. This is reasonable when conditions at a site are similar across years, but this may not be appropriate for carcass persistence estimates at Lundgren (Figure 3). The PP estimate for 2015 at Lundgren is much larger than the NP estimate, whereas that for 2016 is below the NP estimate. With more years of data from more sites, it would be possible to fit more complicated mixed models to account for 3 sources of variability: consistent effects of sites, consistent effects of years, and the idiosyncratic variability among site-year combinations.

Collectively, when BLUPs are compared to site-specific averages, the BLUPs are more precise and closer to the true population values (Robinson 1991). However, BLUPs are not always more precise at each site; some BLUPs can be less precise than the associated site-specific averages (Efron and Morris 1977).

Partial pooling of detection probability components complements the evidence of absence regression model (McDonald et al. 2021). Both use data from multiple sites, or years, but focus on different components of the fatality estimation. The partial pooling model for searcher efficiency or carcass persistence here combines a regression model to account for known variation in searcher efficiency and carcass persistence with a random effect to account for unobserved variation. In the models for searcher efficiency and carcass persistence, the sample sizes (number of bats used in searcher efficiency or carcass persistence trials) are known. Evidence of absence regression develops regression models for the unobserved mean number of fatalities. A random effect could be included in an evidence of absence regression model, but it would have a different effect because the sample size for the mortality model is unknown. Instead, introducing a random effect changes the assumed response distribution in the evidence of absence model from Poisson to negative binomial or some other over-dispersed distribution.

Other methods could be used to combine information from multiple sites, years, and seasons. When there are multiple years of fatality monitoring, Bayesian updating can be used to combine information across years. The data from the first year is used to estimate the posterior distribution of the fatality rate (λ) for that year. This distribution is then used as the prior distribution for the analysis of the data from the second year (Huso et al. 2015). Similar Bayesian updating could be applied to searcher efficiency or carcass persistence distributions. However, updating relies on a temporal order and would not be applicable to combining information from multiple sites.

A second approach is to use a statistical test of no differences among sites to decide whether to use CP or NP. This is an example of pretesting, i.e., using a hypothesis test to decide how to analyze a data set. Two common applications are testing normality (Schucany and Ng 2006) or equal variances (Zimmerman 2010). Post-construction monitoring at the sites we used is typically conducted across 3 seasons, each with searcher efficiency and carcass persistence trials. Whether the seasons are pooled depends on a hypothesis test of equality in the 3 seasons. The properties of this approach for fatality estimation have not been studied, but the general consensus among statisticians is to not use pretesting. It leads to incorrect *P*-values (Moser and Stevens 1992, Rasch et al. 2011) and confidence intervals (Olshen 1973) in the primary analysis. When the estimates based on a pre-testing approach are compared to those from a PP approach, the PP estimates are closer to the true values (Sclove et al. 1972).

The PP model used here considers fixed effects that account for among-year differences in survey protocols and independent random effects that account for additional variability among sites. Both components could be extended to incorporate further sources of variability. For example, additional fixed effect terms could be added to capture regional or vegetation-related differences in searcher efficiency, or latitudinal differences in carcass persistence. More complicated random effect structures could also be adopted to account for temporal variation among years or spatial dependence among sites. A time series correlation structure such as autoregressive, order 1

(Steel 2010), could be used with data from multiple years, or multiple seasons within a year. When site coordinates are available, spatial correlation structures such as spherical or Matérn could be used to allow nearby sites to be more highly correlated than more distant ones (Lawson and Banerjee 2009, Banerjee et al. 2025). In this study our focus was on modeling the components of g, the detection probability. McDonald et al. (2021) describe regression models for λ_i, the mean number of fatalities at site i. Although random effects could also be incorporated to induce PP for λ_i, we did not explore that extension here.

Small area estimation in survey sampling is an application based on PP concepts (Jiang and Rao 2020). The context for small area estimation is a survey that has a sufficiently large sample size to provide a precise estimate of a national average but small sample sizes for individual state estimates. Partial pooling, called borrowing strength in the survey literature (Jiang and Rao 2020), provides more accurate estimates for states, the small areas. In bat fatality monitoring, the small areas are individual sites. The model used for carcass persistence, in which both site variability and estimation error have normal distributions is known as the Fay-Herriot model in the survey literature (Fay and Herriot 1979). This model has been extended in many ways including robust estimation methods (Jiang and Rao 2020) and models for spatial, temporal, or spatio-temporal data (Marhuenda et al. 2013). All could be applied to partial pooling of the components of g when there are sufficient numbers of sites or times.

Partial pooling can be used with the GenEst software (Dalthorp et al. 2017) for mortality estimation. This software inputs information from a searcher efficiency trial, a carcass persistence trial, the search area adjustment, and carcass counts and produces the posterior distribution of fatalities. It can be adapted to implement partial pooling because it allows the user to specify the parameters of the beta distribution that describes the detection probability, g. The code provided in the supporting information fits a partial pooling model and produces samples from the posterior distributions of searcher efficiency and carcass persistence. Samples from the posterior distribution of g can be computed by multiplying posterior samples of searcher efficiency, posterior samples of carcass persistence, and the search area adjustment. The parameters needed by the GenEst software are then obtained by fitting a beta distribution to the samples of g.

We suggest that partial pooling provides 2 useful benefits. First, PP avoids the need to rely on hypothesis tests or model-comparison procedures to choose between CP and NP. Second, CIs using PP require smaller sample sizes than NP to achieve the same precision. For searcher efficiency and carcass persistence, which are each functions of a single sample size, the reduction in sample size can be calculated from the reduction in CI width. We show that partial pooling provides more precise estimates of mortality, but relating that to sample size is more complicated because the mortality estimate is based on 3 sample sizes (carcass searches, search efficiency trials, and carcass persistence trials). The contribution of each sample size to the uncertainty in the estimated mortality can be assessed using statistical sensitivity analysis (Saltelli et al. 2004).

The benefits of hierarchical modeling and partial pooling (Gelman and Hill 2007) extend to many other ecological contexts. The evidence of absence framework and the PP model are well-suited for rare species (Kéry and Schaub 2012), open populations (Royle and Dorazio 2008), and scenarios with low detection probabilities (MacKenzie et al. 2002), making them valuable tools not only for post-construction fatality estimation but also for broader applications in conservation and ecological inference (Ellison 2004, Bürkner 2017).

MANAGEMENT IMPLICATIONS

Precise estimates of bat fatalities are essential for informed decision-making regarding wind farm operations and mitigation strategies. These estimates depend on correctly accounting for imperfect detection, a challenge that becomes more complex when data are sparse or vary across sites. Our findings show that partial pooling improves the precision of key detection parameters and fatality estimates, offering a reliable and practical tool for managers working across multiple sites or with limited data.

ACKNOWLEDGMENTS

We thank 2 reviewers and especially the Associate Editor for their insightful comments that substantially improved the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ETHICS STATEMENT

This study analyzes data derived from previous studies that measured bat fatalities at wind power projects. No live animals were involved in this research. The data used in this study were obtained from publicly available sources and published reports, where the respective studies adhered to the ethical guidelines and regulations regarding animal welfare. The studies from which the data were derived followed the necessary ethical approvals, including permits for data collection in accordance with local and international guidelines for wildlife research. Specific protocol numbers can be found in the original publications cited.

DATA AVAILABILITY STATEMENT

These data were derived from the following resources available in the public domain: Reports in Final Habitat Conservation Plan, Appendices, and Addendum (https://www.regulations.gov/document/FWS-R3-ES-2018-0037-0107) and MEC Final HCP Addendum Vol II (https://www.regulations.gov/document/FWS-R3-ES-2018-0037-0104). They are summarized in BatData.xlsx in the Supporting Information.

ORCID

Natalia Berberian https://orcid.org/0000-0003-4795-2925
Philip M. Dixon https://orcid.org/0000-0002-1778-0686

REFERENCES

Banerjee, S., A. E. Gelfand, and B. P. Carlin. 2025. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC Press, London, United Kingdom.

Bay, K., E. Baumgartner, J. Studyvin, and R. McDonald. 2017. 2015–2016 post-construction fatality monitoring: bat-focused surveys. MidAmerican Energy Company Iowa Wind Energy Portfolio: Adams, Century, Charles City, Highland, Intrepid, Laurel, Lundgren, Macksburg, Pomeroy, Rolling Hills, Vienna I, Vienna II, and Wellsburg. Western EcoSystems Technology, Inc., Cheyenne, Wyoming, USA. Report prepared for MidAmerican Energy Company, Urbandale, Iowa, USA. <https://www.regulations.gov/document/FWS-R3-ES-2018-0037-0107>. Accessed 20 Feb 2024.

Bürkner, P.-C. 2017. brms: an R package for Bayesian multilevel models using Stan. Journal of Statistical Software 80:1–28. https://doi.org/10.18637/jss.v080.i01.

Cochran, W. G., and G. M. Cox. 1957. Experimental designs, second edition. Wiley, New York, New York, USA.

Cox, D. R., and D. Oakes. 1984. Analysis of survival data. Chapman and Hall/CRC, New York, New York, USA.

Cramér, H. 1946. Mathematical methods of statistics. Volume 9. Princeton Mathematical Series. Princeton University Press, Princeton, New Jersey, USA.

Dalthorp, D. H., M. M. P. Huso, and D. Dail. 2017. Evidence of absence (v2.0) software user guide. U.S. Geological Survey Data Series 1055:109. https://doi.org/10.3133/ds1055.

Efron, B., and C. Morris. 1977. Stein's paradox in statistics. Scientific American 236(5):119–127.

Ellison, A. M. 2004. Bayesian inference in ecology. Ecology Letters 7:509–520. https://doi.org/10.1111/j.1461-0248.2004.00603.x

European Commission. 2011. Wind energy developments and Natura 2000: guidance document. Publication Office of the European Union, Luxembourg City, Luxembourg. <https://op.europa.eu/en/publication-detail/-/publication/65364c77-b5b8-4ab6-919d-8f4e3c6eb5c2/language-en>. Accessed 8 Dec 2025.

Fay III, R. E., and R. A. Herriot. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association 74:269-277. https://doi.org/10.1080/01621459.1979.10482505.

Frick, W. F., E. F. Baerwald, J. F. Pollock, R. M. R. Barclay, J. A. Szymanski, T. J. Weller, A. L. Russell, S. C. Loeb, R. A. Medellin, and L. P. McGuire. 2017. Fatalities at wind turbines may threaten population viability of a migratory bat. Biological Conservation 209:172–177. https://doi.org/10.1016/j.biocon.2017.02.023.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian data analysis. Third edition. Chapman and Hall/CRC, New York, New York, USA.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, United Kingdom.

Gelman, A., J. Hwang, and A. Vehtari. 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24:997–1016. https://doi.org/10.1007/s11222-013-9416-2

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7: 457–472.

Gonzalez, E., R. Barquez, and J. Arroyo-Cabrales. 2016. *Lasiurus cinereus*. The IUCN Red List of Threatened Species 2016:e.T11345A22120305. <https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T11345A22120305.en>. Accessed 8 Dec 2025.

Hull, C. L., and S. Muir. 2010. Search areas for monitoring bird and bat carcasses at wind farms using a Monte-Carlo model. Australasian Journal of Environmental Management 17:77–87. https://doi.org/10.1080/14486563.2010.9725253.

Huso, M. M. P. 2011. An estimator of wildlife fatality from observed carcasses. Environmetrics 22:318–329. https://doi.org/10.1002/env.1052.

Huso, M. M. P., D. Dalthorp, D. Dail, and L. Madsen. 2015. Estimating wind-turbine-caused bird and bat fatality when zero carcasses are observed. Ecological Applications 25:1213–1225. https://doi.org/10.1890/14-0764.1.

Jiang, J., and J. S. Rao. 2020. Robust small area estimation: an overview. Annual Review of Statistics and its Application 7: 337-360. https://doi.org/10.1146/annurev-statistics-031219-041212

Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press, Waltham, Massachusetts, USA.

Lawson, A. B., and S. Banerjee. 2009. Bayesian spatial analysis. Pages 361–387 *in* A. S. Fotheringham, and P. A. Rogerson, editors. The SAGE handbook of spatial analysis. SAGE, Thousand Oaks, California, USA.

MacKenzie, D. L., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Marhuenda, Y., I. Molina, and D. Morales. 2013. Small area estimation with spatial-temporal Fay-Herriot models. Computation Statistics and Data Analysis 58:308-325. https://doi.org/10.1016/j.csda.2012.09.002.

McDonald, T., K. Bay, J. Studyvin, J. Leckband, A. Schorg, and J. McIvor. 2021. Evidence of absence regression: a binomial N-mixture model for estimating fatalities at wind energy facilities. Ecological Applications 31:e2408. https://doi.org/10.1002/eap.2408.

McElreath, R. 2020. Statistical rethinking: a Bayesian course with examples in R and Stan. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Moser, B. K., and G. R. Stevens. 1992. Homogeneity of variance in the two-sample means test. American Statistician 46: 19-21. https://doi.org/10.2307/2684403.

Olshen, R. A. 1973. The condition level of the F-test. Journal of the American Statistical Association 68:692-698.

O'Shea, T. J., P. M. Cryan, D. T. Hayman, R. K. Plowright, and D. S. Streicker. 2016. Multiple mortality events in bats: a global review. Mammal Review 46:175–190. https://doi.org/10.1111/mam.12064.

Plummer, M. 2023. Rjags: Bayesian graphical models using MCMC. https://cran.r-project.org/web/packages/rjags/index.html.

Pollock, R. 2007. Recommended formulas for adjusting fatality rates. Appendix F *in* California Energy Commission and California Department of Fish and Game, editors. California guidelines for reducing impacts to birds and bats from wind energy development. Final commission report CEC-700-2007-008-CMF. California Energy Commission, Sacramento, USA.

R Core Team. 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Rasch, D., K. D. Kubinger, and K. Moder. 2011. The two-sample *t* test: pre-testing its assumptions does not pay off. Statistical Papers 52:219-231. https://doi.org/10.1007/s00362-009-0224-x.

Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. Statistical Science 6:15-32.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, Amsterdam, Netherlands.

Saltelli, A., S. Trantola, F. Campolongo, and M. Ratto. 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Wiley, Chichester, England.

Schucany, W. R., and H. K. T. Ng. 2006. Preliminary goodness of fit tests for normality do not validate the one-sample Student t. Communications in Statistics: Theory and Methods 35:2275-2286. https://doi.org/10.1080/03610920600853308.

Sclove, S. L., C. Morris, and R. Radhakrishnan. 1972. Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. Annals of Mathematical Statistics 43:1481-1490.

Smallwood, K. S. 2007. Estimating wind turbine–caused bird mortality. Journal of Wildlife Management 71:2781–2791. https://doi.org/10.2193/2007-006.

Solari, S. 2019. *Nycticeius humeralis*. The IUCN Red List of Threatened Species 2019: e.T14944A22015223. <https://doi.org/10.2305/IUCN.UK.2019-1.RLTS.T14944A22015223.en>. Accessed 8 Dec 2025.

Solari, S. 2021. *Myotis lucifugus*. The IUCN Red List of Threatened Species 2021: e.T14176A208031565. Amended version of 2018 assessment. <https://www.iucnredlist.org/species/14176/208031565>. Accessed 8 Dec 2025.

Steel, M. F. J. 2010. Bayesian time series analysis. Pages 35–45 *in* S. N. Durlauf, and L. E. Blume, editors. Macroeconometrics and time series analysis. The New Palgrave Economics Collection. Palgrave Macmillan, London, United Kingdom. https://doi.org/10.1057/9780230280830_4.

Strickland, M. D., E. B. Arnett, W. P. Erickson, D. H. Johnson, G. D. Johnson, M. L. Morrison, J. A. Shaffer, and W. Warren-Hicks. 2011. Comprehensive guide to studying wind energy/wildlife interactions. Prepared for the National Wind Coordinating Collaborative, Washington, D.C., USA. <https://tethys.pnnl.gov/sites/default/files/publications/Comprehensive-Guide-to-Studying-Wind-Energy-Wildlife-Interactions.pdf>.

Thaxter, C. B., G. M. Buchanan, J. Carr, S. H. M. Butchart, T. Newbold, R. E. Green, J. A. Tobias, W. B. Foden, S. O'Brien, and J. W. Pearce-Higgins. 2017. Bird and bat species' global vulnerability to collision mortality at wind farms revealed through a trait-based assessment. Proceedings of the Royal Society B: Biological Sciences 284:20170829. https://doi.org/10.1098/rspb.2017.0829.

U.S. Fish and Wildlife Service. 2012. Land-based wind energy guidelines. U.S. Department of the Interior, Washington, D.C., USA. <https://www.fws.gov/sites/default/files/documents/land-based-wind-energy-guidelines.pdf>.

Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432. https://doi.org/10.1007/s11222-016-9696-4.

Wikle, C. K. 2003. Hierarchical models in environmental science. International Statistical Review 71(2):181–199. https://doi.org/10.1111/j.1751-5823.2003.tb00192.x.

Zimmerman, D. W. 2010. A note on preliminary tests of equality of variances. British Journal of Mathematical and Statistical Psychology 57:173-181. https://doi.org/10.1348/000711004849222.

Associate Editor: Christopher Rota.

SUPPORTING INFORMATION

Additional supporting material may be found in the online version of this article at the publisher's website.

How to cite this article: Berberian, N., and P. M. Dixon. 2026. Pooling information across sites and years to estimate bat fatalities at wind farms. Journal of Wildlife Management 90:e70206. https://doi.org/10.1002/jwmg.70206