Original Article

Title: Validation of Japanese Collision Risk Models Using Post-Construction Carcass Survey Data

Yasuo SHIMADA[1], Hiroyuki MATSUDA[2]

Corresponding author: Yasuo SHIMADA

Email: yas.shimada35@gmail.com

[1] Environmental Energy Division, Japan Weather Association, 3-1-1, Higashi-ikebukuro, 170-6055, Tokyo, Japan

[2] Faculty of Environment and Information Sciences, Yokohama National University, 79-1, Tokiwadai, 240-0067, Yokohama, Japan

Running title: Validation of collision predictions

Number of tables: 8

Number of figures: 2

ABSTRACT

Collision risk models (CRMs) are widely used to predict bird collisions with wind turbines; however, few studies have evaluated their performance using post-construction monitoring data. In this study, we assessed the validity of two collision risk models commonly used in Japan, the Ministry of the Environment model (CRM_{moe}) and the Yui and Shimada (2013) model (CRM_{yui}), by comparing model predictions with carcass survey results from two wind power facilities. A total of 14 validation cases were recorded during post-construction monitoring. Assuming that the cumulative number of detected carcasses follows a Poisson distribution, we estimated 95% confidence intervals for cumulative collision numbers using three interval-estimation methods: the log-transformed Wald method, the Garwood method, and the likelihood-ratio method. Predicted collision numbers from the CRMs were compared with these intervals after accounting for effective survey coverage. The proportions of model predictions that fell within the estimated 95% confidence intervals were 62.5% (5/8 cases) for CRM_{moe} and 71.4% (10/14 cases) for CRM_{yui}, both substantially lower than the nominal coverage probability of 95%. All predictions falling outside the confidence intervals corresponded to underestimation of collision numbers inferred from carcass survey data. A generalized linear model incorporating survey coverage as an offset further indicated that observed carcass counts were approximately 5.4 times higher than CRM predictions under the reference condition, suggesting a systematic tendency toward underestimation.

The avoidance rate, one of the few CRM parameters not estimated directly from site-specific observations, was evaluated using its original definition based on the ratio of observed to predicted collisions assuming no avoidance. Although avoidance rate was considered a plausible source of the discrepancy, the estimated scaling factor may also reflect uncertainty in other CRM components and the cumulative propagation of uncertainty throughout the model. Our results demonstrate the limitations of evaluating CRM performance using point estimates alone and highlight the importance of interval-based validation, long-term post-construction monitoring, and explicit treatment of uncertainty in collision-risk assessments for wind power facilities.

Keywords: avoidance rate, bird collision, carcass survey, interval estimation, wind power

INTRODUCTION

Collision risk models (CRMs) are mathematical frameworks used to estimate the number of bird collisions with wind turbines. Tucker (1996) first established the foundations of CRM-based collision assessment by formulating the probability of bird collisions with rotating turbine blades within a mathematical framework.

Building on this work, Band (2000) developed a more comprehensive collision risk framework that integrated bird flight activity and rotor-swept airspace use into practical collision risk estimation for wind farms. In 2012, the Strategic Ornithological Support Services (SOSS) project compiled revised and extended versions of the Band (2000) model, which provided a standardized framework for collision risk assessment. Since then, CRM approaches have been further developed and applied in a variety of environmental assessment contexts.

Masden & Cook (2016) compared representative CRMs available at the time and identified several key issues: (1) a general similarity in mathematical structure across models, (2) avoidance probability as the dominant source of uncertainty, (3) limited availability of input data such as flight height and nocturnal behavior, and (4) discrepancies between the assumptions required by CRMs and the quality and quantity of data typically obtainable in environmental impact assessments.

More recently, Cook et al. (2025) compiled 52 CRM frameworks and 75 input parameters in a comprehensive review. They emphasized that, despite increasing model complexity, standardization in practical applications and post-construction validation remain insufficient. They called for further sensitivity analyses, benchmarking, and validation using post-construction monitoring data.

Few peer reviewed studies have directly evaluated the accuracy of CRMs by comparing pre-construction predictions with post-construction carcass survey data. One of the few peer-reviewed studies, Ferrer et al. (2012) compared 53 predictions with carcass data from 20 sites and reported a weak relationship between the predicted and observed numbers of fatalities. To our knowledge, no such validation studies have been conducted in Japan.

In Japan, the Environmental Impact Assessment Act was enacted in 1997 and came into force in 1999. Wind power projects were added to the scope of the Act in 2012, after which the associated regulatory procedures were subsequently established. In 2011, the Ministry of the Environment issued technical guidance for the appropriate siting of wind power facilities with respect to birds and other wildlife and including analytical and evaluation methods for collision risk models.

The guidance also includes standardized protocols for post-construction carcass surveys, enabling direct comparison between predicted collision numbers and carcass survey results.

Among the parameters used in collision risk models, the avoidance rate is unique in that it is generally not estimated directly from site-specific field observations but instead assigned a recommended default value (Scottish Natural Heritage 2010; NatureScot 2025). Because the avoidance rate strongly influences predicted collision numbers, its validity should ideally be verified using post-construction monitoring data. However, such validation has rarely been attempted using carcass survey results.

In this study, we validated collision risk model (CRM) predictions reported in environmental impact assessment reports using post-construction carcass survey data. Because CRM predictions are reported as annual collision estimates, whereas post-construction carcass surveys are typically conducted over multiple years, the comparison was made using cumulative CRM predictions over the corresponding survey period.

MATERIALS AND METHODS

Throughout this paper, quantities predicted by collision risk models are denoted without hats, whereas quantities estimated from carcass survey data are denoted with hats. The subscript "cum" denotes cumulative values over the survey period.

For parameters defined on a yearly basis, the subscript yr denotes a particular survey year (see Table 1 for details)..

1) Collision risk models and cumulative collision estimates

In wind power environmental assessment, the annual predicted number of collisions obtained from the CRM_{moe} and the CRM_{yui} are denoted by λ_{moe} and λ_{yui}, respectively.

Because the carcass surveys analyzed in this study were conducted over multiple years, these annual predictions were converted to cumulative predicted collision numbers corresponding to the survey period before comparison with collision estimates corrected for carcass survey coverage.

In Japan, two collision risk models are commonly used: the model proposed by Yui and Shimada (2013) (hereafter referred to as the CRM_{yui}), and the model described in the Ministry of the Environment guidelines (2011, 2015) for the appropriate siting of wind power facilities with respect to birds and other wildlife (hereafter referred to as the CRM_{moe}).

Both models are based on simplified frameworks derived from the Band (2000) model (Yui &

Shimada 2013; Shimada 2021), and there is no formal guidance in Japan regarding which model should be used. According to Yui (personal communication), the predicted number of collisions from the CRM_{yui} is approximately three times that from the CRM_{moe}.

Predicted collision numbers were obtained from wind power environmental impact assessment documents. Predictions from both the CRM_{moe} (λ_{moe}) and the CRM_{yui} (λ_{yui}) were available for Site A, whereas only the CRM_{yui} prediction was available for Site B.

For both models, the annual predicted number of collisions (λ_m, where *m* denotes either "moe" or "yui") is calculated as the product of the predicted number of collisions per day (θ_m) and the annual presence period (P_{rd}).

$$\lambda_m = \theta_m \cdot P_{rd}. \qquad (1)$$

For both models, the predicted number of collisions per day (θ_m) is expressed as:

$$\theta_m = F_{req,m} \cdot P_{sg,m} \cdot (1 - A_{vd}) \cdot P_{cnt,m} \cdot O_{tm,m} \qquad (2)$$

where, $F_{req,m}$ is daily frequency of entry into the rotor-swept zone, $P_{sg,m}$ is probability of passing through the rotor-swept zone without collision, conditional on entry, A_{vd} is avoidance rate, $P_{cnt,m}$ is probability of collision during passage through the rotor-swept zone, and $O_{tm,m}$ is operating time ratio.

Because carcass surveys were conducted over multiple years, the annual CRM predictions (Yui & Shimada 2013) were converted to cumulative predictions corresponding to the survey period for validation, thereby placing predicted and observed collision numbers on the same temporal scale.

In the CRM_{yui}, the bird is represented as a rhomboid reflecting wing shape, and oblique passage through the rotor plane is assumed when calculating the swept area. As a result, the contact probability in the CRM_{yui} is generally larger.

The operating time ratio, $O_{\_tm,m}$, represents the proportion of time during which the wind turbine is operating. In the CRM_{moe}, it is defined as the ratio of the total operating time to the total time in a year and can be estimated from the cumulative frequency distribution of wind speeds between the cut-in and cut-out wind speeds, based on Rayleigh distribution on wind atlas.

In contrast, the CRM_{yui} employs a modified operation time ratio that accounts for variations in blade rotational speed depending on wind speed classes by wind observation on site.

The avoidance rate (A_{vd}) was assumed to be common to both models, and values reported by NatureScot (2025) were used. The default value is 0.98. However, NatureScot (2025) also proposes species-specific avoidance rates. For example, a value of 0.95 is recommended for the White-tailed eagle.

Following the definition of Scottish Natural Heritage (2010) and NatureScot (2025), the avoidance rate was defined as

$$A_{vd} = 1 - \frac{\text{Observed collisions}}{\text{Predicted collisions without avoidance}} \quad (3)$$

Under the default avoidance rate of 0.98, the annual predicted collision number assuming no avoidance was obtained by multiplying the annual predicted collision number by 50 (= $1/(1 - 0.98)$). For White-tailed Eagle, for which an avoidance rate of 0.95 was applied, the corresponding multiplier was 20 (= $1/(1 - 0.95)$).

As previously mentioned, the two predicted values λ_m represent the number of collisions per year. Since the carcass surveys addressed in this study spanned multiple years, these predicted values required analysis as cumulative values $\lambda_{cum,m}$. by multiplying λ_m by the number of years corresponding to the carcass survey period

In contrast, cumulative collision estimates ($\hat{\lambda}_{\text{cum}}$) were derived from carcass survey data after correcting for survey coverage, as described below.

Let S_{cum} denote the cumulative number of detected carcasses and W_{cum} the effective cumulative survey coverage. The cumulative carcass count was assumed to follow a Poisson distribution with mean $\mu_{cum} = \lambda_{cum} \cdot W_{cum}$, where λ_{cum} denotes the cumulative number of collisions during the survey period:

$$S_{cum} \sim Poisson(\mu_{cum}), \qquad \mu_{cum} = \lambda_{cum} \cdot W_{cum} \tag{4}$$

2) Prediction of collision numbers and validation using carcass survey data.

This section describes the estimation of the cumulative collision number from carcass survey data. The observed cumulative carcass count (S_{cum}) was converted to a cumulative collision estimate by accounting for cumulative survey effort (E_{cum}) and annual carcass detection probability in year yr (p_{yr}).

Detection probability (F_{nd}) and search coverage ratio (E_{cov}) were assumed to be constant throughout the study period. In contrast, carcass persistence probability ($P_{per,yr}$) was allowed to vary among years because the interval between successive carcass surveys differed by year. Therefore, the annual detection probability was defined as

$$p_{yr} = F_{nd} \cdot P_{per,yr} \cdot E_{cov} \tag{5}$$

Cumulative effective survey effort (E_{cum}) was therefore defined as

$$E_{cum} = \sum_{yr=1}^{n} T_{eff,yr} \cdot p_{yr} \qquad (6)$$

where $T_{eff,yr}$ represents the effective observation period, i.e., the portion of the target species' annual presence period covered by carcass surveys.

In this study, cumulative effective survey effort (E_{cum}) was used as the baseline measure of carcass survey effort. Although carcass persistence probability ($P_{per,yr}$) is not itself a component of survey effort, it was incorporated into E_{cum}, because shorter survey intervals increase the probability that carcasses remain detectable before removal.

Finally, the cumulative number of detected carcasses (S_{cum}) is treated as a random variable and assumed to follow a Poisson distribution with mean μ_{cum}

$$S_{cum} \sim \text{Poisson}(\mu_{cum}). \qquad (7)$$

The framework in which detection probability (F_{nd}), persistence probability ($P_{per,yr}$), and search coverage (E_{cov}) are evaluated separately and then combined to estimate the actual number of collisions can be traced back to Gauthreaux (1996). Subsequently, the approach of multiplying these correction factors to obtain an estimate was widely adopted in studies of bird collisions at wind power facilities in North America in the 2000s (e.g., Erickson et al. 2000).

The adjusted cumulative number of collisions ($\hat{\lambda}_{cum}$) becomes larger than the cumulative detected carcass count (S_{cum}). This reflects scaling from the cumulative survey effort to the cumulative seasonal presence period of the target species during the study period.

We define the cumulative survey coverage as

$$W_{cum} = \frac{E_{cum}}{n \cdot P_{rd}} \qquad (8)$$

where n is the number of survey years and P_{rd} is the mean annual presence period (days) of the target species.

The survey-adjusted cumulative collision estimate was defined as $\hat{\lambda}_{cum} = \frac{S_{cum}}{W_{cum}}$. This estimate is based on the observed carcass count (S_{cum}) and the effective cumulative survey coverage (W_{cum}). Because the observed carcass count represents one realization of a Poisson process, the point estimate alone does not represent the range of collision numbers consistent with the observed data. Therefore, confidence intervals for λ_{cum} were estimated using three interval-estimation methods.

The first method was the log-transformed Wald method, which uses the cumulative point estimate on the logarithmic scale (i.e., $\log \lambda_{cum}$) together with its standard error. The second and third methods were based on the cumulative observed carcass count (S_{cum}) and derived confidence intervals directly from the underlying probability distribution, namely the

Garwood interval and the likelihood-ratio (LR) interval, respectively. A brief overview of each method is provided below.

Finally, the cumulative CRM prediction ($\lambda_{cum,m}$) was compared with the adjusted cumulative collision number ($\hat{\lambda}_{cum}$). Because in wind power environmental assessment, CRM predictions represent the expected number of collisions during one annual presence period of the target species, they were converted to cumulative predictions over the survey period by multiplying by the number of carcass survey years (n). The cumulative CRM prediction was regarded as consistent with the carcass survey results when it fell within the 95% confidence interval of the adjusted cumulative collision number ($\hat{\lambda}_{cum}$); otherwise, it was regarded as inconsistent.

3) Interval estimation methods

The cumulative survey coverage W_{cum}, defined in Equation (8), was used to convert confidence limits for the expected number of detected carcasses to confidence limits for the cumulative collision number. Under this formulation, the expected cumulative number of detected carcasses is

$$\mu_{cum} = \lambda_{cum} \cdot W_{cum} \qquad (9)$$

where W_{cum} denotes cumulative survey coverage serving as an offset term. The cumulative carcass count is assumed to follow a Poisson distribution

$$S_{cum} \sim Poisson(\mu_{cum}) \qquad (10)$$

In this study, the 95% confidence intervals for the estimated cumulative number of collisions were calculated using the three methods described below.

a) Log-transformed Wald method (interval estimation using an offset GLM)

This approach finds application when $S_{cum} > 0$, signifying the occurrence of collisions. In the context of a Poisson generalized linear model (GLM), taking the logarithm of Equation (9) yields:

$$\log \mu_{cum} = \log \lambda_{cum} + \log W_{cum} \qquad (11)$$

This relationship can be expressed using a Poisson GLM with a log link by treating $\log W_{cum}$ as an offset term. In this formulation, the intercept parameter β_0, corresponds to $\log \lambda_{cum}$, i.e.,

$$\log \mu_{cum} = \beta_0 + \log W_{cum} \qquad (12)$$

The log-transformed Wald method was used to obtain an interval estimate for β_0. The survey-adjusted cumulative collision estimate was defined as

$$\hat{\lambda}_{cum} = \frac{S_{cum}}{W_{cum}} \qquad (13)$$

Because $\beta_0 = \log \lambda_{\mathrm{cum}}$, replacing the unknown parameter $\lambda_{\mathrm{cum}}$with its estimate $\hat{\lambda}_{\mathrm{cum}}$given by Equation (13) yields the corresponding estimate of the intercept:

$$\hat{\beta}_0 = \log \hat{\lambda}_{cum}. \tag{14}$$

Based on the estimated intercept $\hat{\beta}_0$ and its standard error $\mathrm{SE}(\hat{\beta}_0)$, the 95% confidence interval was constructed using the log-transformed Wald method

$$CI_{\mathrm{Wald}} = [\exp(\hat{\beta}_0 - 1.96\, SE(\hat{\beta}_0)), \exp\left(\hat{\beta}_0 + 1.96\, SE(\hat{\beta}_0)\right)]. \tag{15}$$

When $S_{cum} = 0$, $\hat{\beta}_0 \to -\infty$; therefore, the log-transformed Wald method cannot be applied.

b) Garwood interval

Assuming that the cumulative observed carcass count S_{cum} follows a Poisson distribution, the confidence interval for the Poisson mean μ can be derived using the chi-square (χ^2) distribution (Garwood, 1936). The degrees of freedom of the χ^2 distribution depend on the observed count, denoted by S_{cum} in this study, and the confidence limits are given as below.

When $S_{cum} = 0$:

$$\mu_L = 0,$$

$$\mu_U = 0.5 \cdot \chi^2_{2(S_{cum}+1),\,0.95}$$

When $S_{cum} > 0$:

$$\mu_L = 0.5 \cdot \chi^2(2S_{cum}, 0.025), \tag{16}$$

$$\mu_U = 0.5 \cdot \chi^2(2(S_{cum} + 1), 0.975). \tag{17}$$

Here, $\chi^2(v, q)$ denotes the $q-th$ quantile of the chi-square distribution with v degrees of freedom. The degrees of freedom are defined based on the cumulative observed carcass count S_{cum} with the lower and upper limits given by $2S_{cum}$ and $2(S_{cum} + 1)$, respectively. The lower limit excludes values less than S_{cum}, whereas the upper limit excludes values greater than or equal to $S_{cum}+1$.

If the lower limit were set at the 2.5% quantile when $S_{cum}=0$, the resulting interval would exclude the maximum likelihood estimate ($\mu=0$), which is undesirable. Therefore, when $S_{cum}=0$, the lower limit is set to 0 and the upper limit to 95% quantile.

From Equation (9), the confidence interval for λ_{cum} is obtained as

$$CI_{Gar}=\left[\frac{\mu_L}{W_{cum}},\frac{\mu_U}{W_{cum}}\right] \qquad (18)$$

The Garwood method was adopted because it can appropriately estimate confidence intervals even when no carcasses are detected.

c) Interval estimation using the likelihood ratio (LR) method

From the Poisson model defined in Section 2.2, the likelihood function is given by

$$L(\lambda_{cum})=\exp(-\lambda_{cum}\cdot W_{cum})\cdot\frac{(\lambda_{cum}\cdot W_{cum})^{S_{cum}}}{S_{cum}!} \qquad (19)$$

Taking the logarithm of this expression yields:

$$\log L(\lambda_{cum})=S_{cum}\cdot\log(\lambda_{cum}\cdot W_{cum})-\lambda_{cum}\cdot W_{cum}-\log(S_{cum}!) \qquad (20)$$

The confidence interval for λ_{cum} was estimated using the likelihood ratio statistic. The maximum likelihood estimate of λ_{cum} is given by Equation (13). The likelihood ratio statistic $\mathrm{LR}(\lambda)$ is then defined as:

$$\mathrm{LR}(\lambda_{cum}) = 2(l(\hat{\lambda}_{cum}) - l(\lambda_{cum})) \quad (21)$$

By Wilks' theorem, $\mathrm{LR}(\lambda_{cum})$ asymptotically follows a chi-square distribution with 1 degree of freedom. Therefore, the 95% confidence interval CI_{LR} is given by:

$$CI_{LR} = \{\lambda_{cum}: \mathrm{LR}(\lambda_{cum}) < \chi^2\ (1, 0.95)\} \quad (22)$$

In practice, the values of λ_{cum} satisfying $\mathrm{LR}(\lambda_{cum}) = \chi^2\ (1, 0.95)$ were numerically obtained, and the corresponding lower and upper bounds were taken as the confidence limits. When $S_{cum} = 0$, the lower bound becomes 0, resulting in a one-sided interval with only an upper limit. Unlike the Wald and Garwood methods, the likelihood-ratio method does not necessarily allocate 2.5% to each tail of the confidence interval, although the overall coverage probability remains 95%.

4) Study sites, target species, and carcass survey methods

a) Carcass survey data

Table 2 summarizes the target species and study sites (A and B), the annual predicted collision numbers from the CRM$_{moe}$ (λ_{moe}) and the CRM$_{yui}$ (λ_{yui}), the cumulative observed carcass count (S_{cum}), the number of carcass surveys (N_{um}) conducted during the study period, the classification group (G_{rp}), the search coverage ratio (E_{cov}), the persistence probability ($P_{per,yr}$), the cumulative effective survey effort (E_{cum}), the number of survey years (n), and the annual presence period of the target species (P_{rd}).

b) Study sites

The study sites consisted of two wind power plants, referred to as Site A and Site B. The site names were anonymized. Both wind farms comprise turbines with a rated capacity of approximately 2.3 MW and commenced operation in 2021 and 2018, respectively.

c) Post-construction monitoring reports

The environmental impact assessment documents and post-construction monitoring reports for Sites A and B were reviewed, and information on the predicted annual collision numbers (λ_{moe} and λ_{yui}), carcass survey methods, and survey results were compiled.

d) Target species

Target species were defined as bird species classified as large birds in the environmental impact assessment documents for Sites A and B (Table 2).

The species included in the analysis were selected from the post-construction monitoring report for Site A, and from both the post-construction monitoring report and the environmental impact assessment report for Site B.

For Site A, the post-construction monitoring report provided the species included in the CRM, their predicted collision numbers, and the numbers of carcasses detected during carcass surveys. Species for which the predicted collision number was zero because they were not recorded within the grid cells used for the CRM calculation, and for which no carcasses were detected during the carcass surveys, were excluded from the analysis. A total of four such cases were excluded from the analysis.

For Site B, the post-construction monitoring report did not provide the species included in the CRM or their predicted collision numbers, and contained only information on carcasses detected during the surveys. Because predicted collision numbers for raptors were provided in the environmental impact assessment report, these raptor species were included in the analysis.

At Site B, the target species included oriental honey-buzzard (*Pernis ptilorhynchus*), sparrowhawk (*Accipiter nisus*), northern goshawk (*Accipiter gentilis*), mountain hawk-eagle (*Nisaetus nipalensis*), and eastern buzzard (*Buteo japonicus*).

For eastern buzzard, two surveys were conducted during the environmental impact assessment, yielding two different predicted collision numbers. Therefore, the two survey results were treated separately as Eastern Buzzard 1 and Eastern Buzzard 2.

e) Carcass survey design

Carcass surveys at both Sites A and B were conducted at regular intervals during the post-construction monitoring period in accordance with the Ministry of the Environment guidelines (2011, 2015).

For each site and species, only the overlap between the carcass survey period and the seasonal presence period (P_{rd}) was used to calculate the effective observation period (T_{eff}).

f) Search area and search coverage ratio

The search area was also defined in accordance with the Ministry of the Environment guidelines (2011, 2015). At both Sites A and B, the search area was defined as a circular area centered on the turbine base, with a radius equal to the maximum blade-tip height (hub height plus blade radius).

The search coverage ratio (E_{cov}) for Site A ranged from 64% to 91% based on aerial photograph interpretation and expert judgment (Kagawa, personal communication). The arithmetic mean value of 76.4% was used in the analysis. For Site B, search areas were reported for individual turbines in the post-construction monitoring report, and the arithmetic mean search coverage ratio of 82.3% was used.

g) Detection probability and persistence probability

Neither Site A nor Site B reported the detection probability (F_{nd}) or persistence probability (P_{per}) in their post-construction monitoring reports. Therefore, the detection probability (F_{nd}) was set to 1.0 based on Kitano & Shiraki (2013), who estimated this value from field experiments in which bird models were placed and subsequently searched for by observers.

The carcass persistence probability ($P_{per}[yr]$) was calculated following Kitano & Shiraki (2013), based on the carcass persistence curve derived from scavenger-removal trials. The probability, $R(d)$, that a carcass remained d days after death was expressed as

$$R(d) = \min(1, 1.0733 \exp(-0.013d)) \qquad (23)$$

Although the survey interval varied among years, within-year variation was small. Assuming that collision-related deaths occur at a constant daily rate and disregarding the probability of detecting a carcass that was overlooked in a previous survey, the mean survey interval I_{yr}(days) was calculated separately for each year yr. The annual carcass persistence probability $P_{\mathrm{per},yr}$ was then calculated as follows:

$$P_{per,yr} = \frac{1}{I_{yr}} \sum_{d=1}^{I_{yr}} R(d) \tag{24}$$

5) **Analysis of factors affecting observed carcass counts**

The confidence intervals estimated from carcass survey data were compared with the collision numbers predicted by the two CRM models. A prediction was regarded as consistent with the carcass survey data when the predicted collision number fell within the corresponding 95% confidence interval. Predictions outside the confidence interval were regarded as inconsistent with the observed data.

The GLM analyses were conducted using a tidy dataset (Supplementary Table S1), in which each record (i) represented a unique observation defined by study site, survey period, target species, and collision risk model. The relationship between observed carcass counts (S_i) and predicted collision numbers ($\lambda_{m,i}$) is influenced by both statistical variability and systematic discrepancies arising from the collision risk models themselves. To investigate these effects, generalized linear models (GLMs) were fitted as described below.

For the first analysis, a Poisson generalized linear model (GLM) was fitted to the tidy dataset (Supplementary Table S1), using the observed carcass count (S_i) as the response variable and the predicted collision number ($\lambda_{m,i}$) as the explanatory variable. To account for differences in effective survey effort, the offset term $\log W_i$ was included in the model, where $P_{rd,i}$ denotes the annual seasonal presence period of the target species at record i and the numerator represents the effective survey effort adjusted for detection probability, carcass persistence, and search coverage ratio.

First, a baseline model was fitted using the logarithm of the predicted collision number of record i, $\log(\lambda_{m,i})$, as the explanatory variable. To account for differences in effective survey effort, the log-transformed adjusted survey coverage was incorporated as an offset term:

$$\log \mu_i = \beta_0 + \beta_1 \log(\lambda_{m,i}) + \log(\frac{T_{eff,i,yr} \cdot F_{nd,i} \cdot P_{per,i,yr} \cdot E_{cov,i}}{P_{rd,i}}) \qquad (25)$$

Next, candidate models were fitted by including additional explanatory variables: the standardized number of surveys ($N_{um,i}^{std}$) and the species classification group ($G_{rp,i}$). These candidate models were compared using the Akaike Information Criterion (AIC), and the model with the lowest AIC was selected as the best model.

$$\log \mu_i = \beta_0 + \beta_1 \log(\lambda m_{,i}) + \beta_2 \cdot N_{um,i}^{std} + \beta_3 \cdot G_{rp,i} + \log(\frac{T_{eff,i,yr} \cdot F_{nd,i} \cdot P_{per,i,yr} \cdot E_{cov,i}}{P_{rd,i}}) \qquad (26)$$

where i denotes the i-th record in Supplementary Table S1, corresponding to a combination of study site, survey year, target species, and prediction model.

For the second analysis, we examined, on an annual basis, how the structure of the collision risk models (CRM) influenced the discrepancy between the predicted number of detected carcasses($\lambda_{m,i} \cdot W_i$) and the observed carcass counts (S_i).

To this end, we extended the assumption that predicted collision numbers directly explain observed carcass counts by introducing a scaling coefficient.

Given that the tidy dataset (Supplementary Table S1) contains predicted collision numbers from two collision risk models (CRM_{moe} and CRM_{yui}) for two wind farms, we further examined whether the scaling coefficient differed between collision risk models and between sites.

First, the adjusted survey coverage (W_i) was defined as

$$W_i = \frac{T_{eff,i,yr} \cdot F_{nd,i} \cdot P_{per,i,yr} \cdot E_{cov,i}}{P_{rd,i}} \tag{27}$$

Under the assumption that

predicted collision numbers directly explain observed carcass counts,

$$\mu_i = W_i \cdot \lambda_{m,i} \tag{28}$$

A scaling coefficient (k) was then introduced to quantify systematic discrepancies between predicted and observed values:

$$\mu_i = W_i \cdot k \cdot \lambda_{m,i} \tag{29}$$

Here, k is a scaling coefficient introduced to quantify systematic discrepancies between predicted and observed values.

When $k = 1$, the predicted values are, on average, consistent with the observed values. When $k > 1$, the predicted values tend to underestimate the observations, whereas when $k < 1$, they tend to overestimate them. μ_i denotes the expected value of the observed carcass count S_i. A Poisson GLM with a log-link function was used for estimation.

$$\log \mu_i = \log \lambda_{m,i} + \log W_i + \beta_0 \qquad (30)$$

By treating $\log(\lambda_{m,i})$ and $\log(W_i)$ as offset terms, the scaling coefficient was estimated as $\beta_0 = \log k$.

This formulation allows the proportional relationship between the predicted collision numbers and the observed carcass counts to be evaluated.

Next, to examine factors influencing variation in the scaling coefficient, a GLM was constructed with site (Sites A and B) and collision risk model (CRM_{moe} and CRM_{yui}) as fixed-effect explanatory variables.

$$\log \mu_i = \log \lambda_{m,i} + \log(W_i) + \beta_{site} + \beta_{CRM} + \beta_0 \qquad (31)$$

In this model, $\log(\lambda_{m,i})$ and $\log(W_i)$ were treated as known offset terms, while the effects of site (β_{site}), collision risk model (β_{CRM}), and the intercept (β_0) were estimated.

The intercept corresponds to the logarithm of the scaling coefficient under the reference condition (Site A with CRM_{yui}). Since predicted collision numbers from the CRM_{moe} were not available at Site B, the interaction between site and collision risk model could not be directly estimated. Therefore, the scaling coefficient for the hypothetical combination of Site B and the CRM_{moe} was derived from the estimated model parameters.

6) Validation of predicted collision numbers and required observation period.

When evaluating the validity of predicted collision numbers based on carcass surveys, the ability to statistically evaluate model validity depends on the duration of the observation period. In particular, when collisions occur at low frequencies, it is inherently difficult to assess the validity of predicted collision numbers based on short-term observations.

Assuming that the annual number of collisions follows a Poisson process, we theoretically examined the observation period required to statistically evaluate whether the expected observed carcass count $(p \cdot \lambda)$ differs significantly from the observed cumulative carcass count.

When the observation period is T years, the cumulative carcass count S_{cum} is assumed to follow a Poisson distribution with mean $p \cdot \lambda \cdot T$, the null hypothesis $H_o\colon \lambda = \lambda_{pred}$ was evaluated using a likelihood ratio test with a significance level of 0.05.

For each combination of the expected observed carcass count ($p \cdot \lambda_{pred}$) and observed cumulative carcass count S_{cum}, we explored the range of observation periods T for which H_0 could not be rejected. The minimum value of T was defined as the lower bound, and the maximum value as the upper bound. Considering the assumed operational lifetime of wind power facilities, the upper bound was set to 20 years.

The lower bound represents the shortest observation period for which the null hypothesis (H_0) is not rejected. Any shorter observation period would lead to rejection of H_0, indicating that the predicted collision rate is significantly lower than the observed collision rate. In contrast, the upper bound represents the longest observation period for which H_0 is not rejected. Any longer observation period would lead to rejection of H_0, indicating that the predicted collision rate is significantly higher than the observed collision rate. Thus, the interval between the lower and upper bounds defines the range of observation periods for which the null hypothesis cannot be rejected.

The null hypothesis (H_0) assumes that the true collision rate is equal to the CRM prediction. This hypothesis concerns the agreement between the predicted and true collision rates and does not imply statistical independence between predicted and observed carcass counts.

RESULTS

1) Validation of predicted collision numbers using cumulative carcass survey data.

Figure 1 and Table 3 compare the cumulative predicted collision numbers ($\sum \lambda_{\mathrm{moe,yr}}$ and $\sum\lambda_{\mathrm{yui,yr}}$) with the 95% confidence intervals estimated from cumulative carcass counts (S_{cum}). Comparing the interval estimation methods, the Garwood method produced wider confidence intervals than the other methods. The log-transformed Wald method provided relatively narrower confidence intervals when $S_{cum} > 0$. The likelihood-ratio (LR) method yielded confidence intervals slightly narrower than those of the Garwood method for $S_{cum} > 0$. When $S_{cum} = 0$, all three methods produced one-sided confidence intervals with a lower bound of zero (Figure 1).

The proportions of cases in which cumulative predicted collision numbers fell within the estimated 95% confidence intervals were 62.5% (5/8) for $\mathrm{CRM_{moe}}$ and 71.4% (10/14) for $\mathrm{CRM_{yui}}$, both lower than the nominal coverage probability of 95%.

When no carcasses were detected ($S_{cum} = 0$), all predictions from both $\mathrm{CRM_{moe}}$ (4/4) and $\mathrm{CRM_{yui}}$ (7/7) fell within the estimated confidence intervals, reflecting the zero lower confidence limit in these cases. In contrast, when one or more carcasses were detected ($S_{cum} > 0$), the proportions of predictions within the confidence intervals decreased to 25.0%

(1/4) for CRM_{moe} and 42.9% (3/7) for CRM_{yui}.

All cases in which the predicted collision numbers fell outside the estimated confidence intervals corresponded to underestimation of collision numbers inferred from cumulative carcass counts, and no cases of overestimation were observed (Figure 1).

2) Analysis of factors affecting observed carcass counts.

a) Model construction results.

Table 4 summarizes the model selection results. First, baseline models using only the log-transformed predicted collision number as the explanatory variable were evaluated separately for the CRM_{moe} and CRM_{yui}. For the CRM_{moe}, the baseline model showed significant overdispersion, with a dispersion ratio of 2.16 (P-value = 0.007). In contrast, the Yui baseline model had a dispersion ratio of 1.20 (P-value = 0.169), suggesting little evidence of substantial overdispersion.

Next, extended models were fitted by adding the standardized number of surveys($N_{um,i}^{std}$) and species classification group ($G_{rp,i}$) as explanatory variables. Based on model selection using Akaike Information Criterion (AIC), the best-supported model for the CRM_{moe} included both $N_{um,i}^{std}$ and $G_{\mathrm{rp},i}$ in addition to the baseline predictor, whereas the best-supported model for the

CRM_{yui} included only $G_{\mathrm{rp},i}$ in addition to the baseline predictor (Table 4). The dispersion ratios of the selected models were close to 1, indicating that overdispersion was not evident in these models.

Parameter estimates of the selected models are presented in Table 5. For CRM_{moe}, none of the explanatory variables showed statistically significant effects. In contrast, for CRM_{yui}, species group was a significant predictor, with the waterfowl group showing higher observed carcass counts than the reference group (other species) ($p = 0.012$).

b) Estimation of the scaling coefficient k

To examine whether the scaling coefficient differed between sites and collision risk models, a Poisson generalized linear model (GLM) including site (Sites A and B) and collision risk model (CRM_{moe} and CRM_{yui}) as explanatory variables was fitted. The results are presented in Table 6.

Under the baseline condition (Site A and CRM_{yui}), the estimated scaling coefficient was 5.44 (95% CI: 3.01–9.82), indicating that observed carcass counts were approximately 5.4 times the predicted collision numbers after adjustment for survey effort.

The scaling coefficient for Site B was estimated to be 1.43 times that for Site A (95% CI: 0.50–4.10), although this difference was not statistically significant (P-value = 0.511).

Similarly, the scaling coefficient for CRM_{moe} was estimated to be 1.31 times that for CRM_{yui} (95% CI: 0.57–3.03), but this difference was not statistically significant (P-value = 0.522).

Because no CRM_{moe} predictions were available for Site B, the interaction between site and collision risk model could not be evaluated. Therefore, only main effects were included in the scaling model.

3) Validation of predicted collision numbers and required observation period

Table 7 presents the lower and upper bounds of the observation period required to conclude that the expected number of detected collisions $(p \cdot \lambda)$ is not significantly different from the cumulative number of detected carcasses (S_{cum}), for each combination of $(p \cdot \lambda)$ and (S_{cum}),.

For example, when$(p \cdot \lambda) = 1$ and $(S_{cum}) = 2$, the predicted collision number would be judged to be underestimated after one year of observation, not significantly different from the observed count after 2–14 years, and overestimated after 15 years or more.

In contrast, when $(p \cdot \lambda)$ = 0.05, underestimation cannot be concluded even if no carcasses are detected (S_{cum}) = 0 over a 20-year observation period. If (S_{cum}) = 1, underestimation can be concluded when the cumulative carcass count is observed within the first nine years, whereas the null hypothesis cannot be rejected from 10 years onward within the 20-year period. Furthermore, if (S_{cum}) = 2, underestimation can be concluded throughout the 20-year observation period.

When $(p \cdot \lambda)$ is small, overestimation often cannot be concluded within the assumed 20-year operational lifetime of the facility, even when the cumulative number of detected carcasses (S_{cum}) is zero or small.
In contrast, when $(p \cdot \lambda)$ was relatively large, the lower bound of the required observation period tended to increase with the cumulative carcass count, indicating that a minimum observation period was required even when many carcasses were detected over a short period.

Furthermore, for some combinations of $(p \cdot \lambda)$ and cumulative number of detected carcasses (S_{cum}), the required observation period reached or exceeded 20 years, indicating that statistical validation may remain difficult within the assumed 20-year operational lifetime of the facility.

4. Discussion

1) Validation of predicted collision numbers based on carcass survey data

In this study, comparison with confidence intervals derived from carcass survey data showed that, in cases where discrepancies occurred between predicted and observed collision numbers, all CRM predictions falling outside the estimated confidence intervals were biased toward underestimating the number of collisions.

Our results suggest that uncertainty in the estimation of collision mortality, which forms the basis of such assessments, may increase as the predicted number of collisions becomes very small.

2) Analysis of factors influencing the number of detected carcasses

a) Model selection and overdispersion

The detection of overdispersion in the basic model indicates that variation in the number of detected carcasses could not be adequately explained by the predicted collision number ($\lambda_{m,i}$) alone. The disappearance of overdispersion after the addition of the standardized number of surveys ($N_{um_i^{std}}$) and taxonomic group ($G_{rp,i}$) suggests that these additional factors accounted for part of the variation that was unexplained by the basic model.

Furthermore, because the use of a negative binomial distribution did not improve AIC, the results do not appear to depend strongly on the assumed variance structure. Therefore, for the present dataset, an analytical framework based on a Poisson model was considered appropriate.

b) Interpretation of the scaling coefficient

In addition, we quantified the relationship between carcass survey results and predicted collision numbers using the scaling coefficient k. The GLM analysis estimated k to be approximately 5.4 under the reference condition (Site A and CRM_{yui}), indicating that the observed carcass counts were approximately 5.4 times the predicted collision numbers after accounting for adjusted survey coverage.

In contrast, neither the site effect nor the difference between the two collision risk models was statistically significant. The scaling coefficient for Site B was estimated to be approximately 1.43 times that for Site A, but the confidence interval was wide and the site effect was not statistically significant. Similarly, the scaling coefficient for CRM_{moe} was estimated to be approximately 1.31 times that for CRM_{yui}, but this difference was also not statistically significant.

This point estimate was smaller than the approximately threefold difference between the two models suggested by Yui (personal communication). However, because the 95% confidence interval (0.57–3.03) included a threefold difference, the relationship suggested by Yui could not be ruled out.

Figure 2 shows the relationship between the predicted number of detected carcasses ($\lambda_{m,i} \cdot W_i$) and the observed carcass count (S_i). Excluding cases in which no carcasses were detected, nearly all data points lie above the 1:1 line. This indicates that the predicted numbers of detected carcasses were lower than the observed carcass counts.

The scaling coefficient estimated in this study is specific to the study sites and survey conditions and should not be directly applied to other sites. Nevertheless, our results suggest that a systematic discrepancy may exist between predicted collision numbers and observed carcass counts.

c) Possibility of overestimation of the avoidance rate

To explore one possible source of this underestimation, we focused on the default avoidance rate (A_{vd} = 0.98; 0.95 for White-tailed Eagle), which is the only parameter that cannot be directly derived from field observations. We examined whether these avoidance rates were consistent with the observed carcass counts (S_{cum}). The objective of this analysis was not to propose new avoidance rates, but to assess the validity of the values currently adopted in the collision risk models (CRMs).

The results are shown in Table 8. Among the 14 validation cases, the confidence intervals for the avoidance rate fell below the default avoidance rate in four cases: Ducks, Sparrowhawk, and Northern Goshawk at Site A, and Honey-buzzard at Site B.
These results suggest that, for these species (including waterfowl within the Ducks group), the default avoidance rates may be higher than the values supported by the carcass survey data.

As shown in Table 8, the lower confidence limit of the avoidance rate estimated using the likelihood-ratio method tended to become negative as the predicted collision number (λ_m) decreased. This reflects the increasing uncertainty in avoidance-rate estimation when predicted collision numbers are small. In other words, an avoidance rate of zero, or even a negative value (i.e., attraction), cannot be statistically excluded.

In current environmental impact assessments in Japan, species-specific avoidance rates are used for some species, whereas the default value of 0.98 is assigned to most species. Because avoidance rates may vary among wind farms, they should be validated through post-construction monitoring or, as illustrated in Table 8, estimated from a sufficiently large number of post-construction observations while accounting for environmental variability. Therefore, the default avoidance rate of 0.98 should be regarded as an assumption requiring further validation. Such validation should include multi-year post-construction carcass surveys and, where feasible, flight observations to characterize bird flight activity around operating turbines and thereby support the evaluation of avoidance rates under actual operating conditions.

The avoidance rates adopted in the Japanese collision risk models (CRMs) evaluated in this study were originally based on the framework developed for the Band model (Scottish Natural Heritage 2010), which was subsequently updated by NatureScot (2025). Although the formulations used to calculate collision probability differ between the Band model and the Japanese CRM implementations, both estimate collision numbers assuming no avoidance before applying an avoidance rate. Thus, the conceptual definition of avoidance is common to these models, and avoidance rates are conceptually comparable between them. However, our results suggest that the numerical values of these avoidance rates should be evaluated using post-construction monitoring data under Japanese conditions.

d) Interpretation of the scaling coefficient in terms of the avoidance rate

Collision numbers predicted by the Japanese collision risk models are expressed as the product of multiple model components. Therefore, the scaling coefficient representing the discrepancy between predicted collision numbers and observed carcass counts can be interpreted as a correction factor applied to one or more of these components.

In this study, we focused on the avoidance rate because it is the only major parameter that cannot be directly estimated from field observations. We therefore examined how the observed discrepancy could be interpreted under the assumption that it was attributable to the avoidance rate.

The point estimates of the scaling coefficient derived from Table 6 ranged from 5.44 to 7.78 (5.44 for Site A with CRM_{yui}, $5.44 \times 1.43 = 7.78$ for Site B with CRM_{yui}, and $5.44 \times 1.31 = 7.13$ for Site A with CRM_{moe}).

If the discrepancy between predicted and observed carcass counts is assumed to be entirely attributable to the non-avoidance component of the CRM, the scaling coefficient can be reinterpreted in terms of an equivalent avoidance rate.

In both CRM_{moe} and CRM_{yui}, the predicted collision number is proportional to the probability of non-avoidance. If A_{old} denotes the avoidance rate originally assumed in the CRM and

A_{new} denotes the equivalent avoidance rate corresponding to the scaling coefficient k, the following relationship holds:

$$A_{\text{new}} = 1 - k(1 - A_{\text{old}}) = 1 - k + kA_{\text{old}}. \quad (32)$$

Thus, the default avoidance rate of 0.98 corresponds to approximately 0.84–0.89, whereas the value of 0.95 for White-tailed Eagle corresponds to approximately 0.61–0.73. These values are illustrative reinterpretations of the scaling coefficient under the assumption that the observed discrepancy is entirely attributable to the avoidance component.

However, the estimated scaling coefficient should not necessarily be interpreted as reflecting uncertainty in the avoidance rate alone. The discrepancy between predicted collision numbers and observed carcass counts may also arise from uncertainties in other model components, such as estimates of flight frequency and blade-strike probability. Consequently, the estimated scaling coefficient may reflect the cumulative effects of uncertainties across multiple model components rather than the effect of a single component (parameter). A more informative approach would be to estimate the avoidance rate independently using data that directly quantify bird–turbine interactions.

3) Feasibility of validating predicted collision numbers and the required monitoring period

The results presented in Table 7 highlight the limitations of using carcass survey data to validate predicted collision numbers, particularly when collisions are rare events. Without long-term monitoring, it is difficult to rule out the possibility that observed values reflect stochastic temporal variation. In particular, when the cumulative number of observed carcasses is zero or very small, it may not be possible to validate the predictions within the approximately 20-year expected operational lifetime of a wind power facility.

In any case, our results suggest that long-term monitoring of collision fatalities may be necessary to validate predicted collision numbers. A more practical approach, rather than validating individual projects separately, would be to pool data from multiple wind power projects. This would increase the cumulative expected number of detected carcasses (λ_{cum}) and consequently improve the statistical power for comparing predicted and observed cumulative carcass counts. Such an approach could provide a more robust basis for evaluating the performance of collision risk models across projects and environmental conditions.

Our findings demonstrate that evaluating predicted collision numbers solely on the basis of point estimates has important limitations. Therefore, uncertainty in predicted collision numbers should also be explicitly considered, particularly when point estimates fall below the confidence intervals estimated from post-construction carcass survey data. Kagawa et al. (2024) addressed this issue by repeatedly calculating predicted collision numbers for each observation day and constructing prediction intervals. Such interval-based approaches not only explicitly represent the uncertainty associated with collision risk predictions but also provide a more rational basis for comparison with post-construction carcass survey results. Similarly, when species-specific avoidance information is unavailable, NatureScot (2024) recommends sensitivity analysis using multiple assumed avoidance rates (95%, 98%, 99%, and 99.5%) rather than relying on a single avoidance-rate assumption.

4) Distinguishing Model Validation from Environmental Risk Assessment

The statistical difficulty of validating low predicted collision numbers should be considered in the context of the purpose of environmental impact assessment (EIA). Model validation and risk assessment are related but distinct processes with different objectives. Model validation aims to determine how accurately a model predicts observed outcomes and is important for evaluating and improving model performance.

In contrast, the primary objective of risk assessment in an EIA is not necessarily to demonstrate that predicted collision numbers precisely match observed values, but rather to determine whether the predicted risk is sufficiently low relative to an acceptable threshold. Therefore, the statistical difficulty of validating low predicted collision rates should not, in itself, be interpreted as evidence that the predicted risk is unacceptable or that the model is unsuitable for risk-based decision making. Conversely, placing excessive emphasis on predictive accuracy could lead to a situation in which even a model predicting a low level of

risk is regarded as problematic simply because its predictions cannot be statistically validated with high precision.
A more appropriate framework would distinguish between model validation, which addresses predictive performance, and risk assessment, which addresses whether the predicted environmental impact is acceptably low.

From the perspective of ecological risk assessment, the implications of uncertainty in predicted collision numbers extend beyond individual collision events to potential impacts on the affected population.
Unlike human health risk assessment, ecological risk assessment focuses not on impacts on individual organisms but on the persistence of populations (e.g., Matsuda 2008). One useful indicator of population-level impacts is Potential Biological Removal (PBR; Wade 1998), which can be used to assess whether cumulative collision mortality exceeds a biologically sustainable level (e.g., Sugimoto & Matsuda 2011; Shimada & Matsuda 2007; NEDO 2018).
From this perspective, it is important to distinguish between uncertainty in predicted collision numbers and the ultimate question of whether the resulting population-level risk is acceptable.

One of the major challenges in post-construction monitoring is the time and financial cost associated with carcass surveys. In Japan, carcass surveys are generally conducted by trained surveyors, although in some cases wind farm personnel conduct visual inspections during routine operation and maintenance activities (Yui, personal communication). Smallwood (2007) pointed out that carcass surveys conducted by personnel without expertise in bird identification may have limited reliability. In the United States, trained detection dogs have been widely used for carcass surveys and have substantially improved detection efficiency (Smallwood et al. 2020).

However, to our knowledge, the use of trained detection dogs for carcass surveys has not yet become widespread in Japan. Therefore, the introduction of new technologies, such as automated camera-based carcass detection systems, should be considered to expand cumulative survey coverage (Happ et al. 2021). Such technologies could facilitate the accumulation of large datasets across projects, thereby contributing not only to more efficient post-construction monitoring but also to more robust validation of collision risk models and more reliable risk-based environmental assessments.

ACKNOWLEDGMENTS

We would like to express sincere gratitude to Dr. Masatoshi Yui (Professor Emeritus, Iwate Prefectural University) for his generous support in providing post-construction monitoring data from wind power facilities.

The authors used ChatGPT (OpenAI) as a supplementary tool for language editing, manuscript organization, and discussion of statistical interpretations. All analyses, interpretations, and final decisions regarding the manuscript were made by the authors.

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of their affiliated organizations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Band W (2000) Windfarms and birds – calculating a theoretical collision risk assuming no avoiding action. Scottish Natural Heritage Guidance Note.

Cook ASCP, Salkanovic E, Masden E, Lee HE & Kiilerich AH (2025) A critical appraisal of 40 years of avian collision risk modelling: How have we got here and where do we go next? Environ Impact Assess Rev 110: 107717. https://doi.org/10.1016/j.eiar.2024.107717

Erickson WP, Johnson GD, Strickland MD, Young DP Jr, Sernka KJ & Good RE (2000) Avian and bat mortality associated with the initial phase of the Foote Creek Rim Windpower Project, Carbon County, Wyoming. West, Inc., Cheyenne, Wyoming. Available at https://www.osti.gov/servlets/purl/822422.

Ferrer, M., De Lucas, M., Janss, G. F., Casado, E., Munoz, A. R., Bechard, M. J., & Calabuig, C. P. (2012). Weak relationship between risk assessment studies and recorded mortality in wind farms. Journal of Applied Ecology, 49(1), 38-46.

Garwood F (1936) Fiducial limits for the Poisson distribution. Biometrika 28: 437–442.

Gauthreaux SA (1996) Suggested practices for monitoring bird populations, movements, and mortality in wind resource areas. In: Proceedings of the National Avian–Wind Power Planning Meeting II. Pp 88–110. National Wind Coordinating Committee, Washington, D.C.

Happ C, Sutor A, Hochradel K (2021) Methodology for the automated visual detection of bird and bat collision fatalities at onshore wind turbines. Journal of Imaging 7(12):272.

Kagawa H, Shimada Y, Morio T, Hashimoto T & Matsuda H (2024) Simulation of collision risk of migratory small birds with wind turbines. Poster presented at the Annual Meeting of the Ornithological Society of Japan (in Japanese).

Kitano M & Shiraki S (2013) Estimation of bird fatalities at wind farms with complex topography and vegetation in Hokkaido, Japan. Wildl Soc Bull 37: 41–48.

Madsen EA & Cook ASCP (2016) Avian collision risk models for wind energy impact assessments. Biol Conserv 194: 1–10.

Matsuda H (2008) Introduction to ecological risk science (in Japanese). Kyoritsu Shuppan, Tokyo.

Ministry of the Environment of Japan (2011) Guidelines for proper site selection of wind power facilities concerning birds and other wildlife. Environmental Impact Assessment Division, Tokyo (in Japanese; partially revised in 2015).

NatureScot (2024) Guidance on using an updated collision risk model to assess bird collision risk at onshore wind farms. NatureScot, Scotland.

NatureScot (2025) Wind farm impacts on birds – use of avoidance rates in the NatureScot wind farm collision risk model. NatureScot Guidance Document, updated September 2025.

Scottish Natural Heritage (2010) Use of Avoidance Rates in the SNH Wind Farm Collision Risk Model. SNH Avoidance Rate Information and Guidance Note. Scottish Natural Heritage, Inverness, UK.

Shimada Y (2021) Reducing bird collision risk per megawatt by introducing longer wind turbine blades. Ornithol Sci 20: 253–261.

Shimada Y & Matsuda H (2007) A bird collision risk management model for wind power projects. Jpn J Conserv Ecol 12: 126–142 (in Japanese with English abstract).

Smallwood KS (2007) Estimating wind turbine-caused bird mortality. J Wildl Manage 71: 2781–2791.

Smallwood KS, Bell DA & Standish S (2020) Dogs detect larger wind energy effects on bats and birds. J Wildl Manage 84: 1–13.

Sugimoto H & Matsuda H (2011) Collision risk of White-fronted Geese with wind turbines. Ornithol Sci 10: 61–71.

Tucker VA (1996) A mathematical model of bird collisions with wind turbine rotors. J Sol Energy Eng 118: 253–262.

Wade PR (1998) Calculating limits to the allowable human-caused mortality of cetaceans and pinnipeds. Mar Mamm Sci 14: 1–37.

Yui Y & Shimada Y (2013) Estimation of avian collision numbers with wind turbines using a spherical model. J Policy Stud 15: 11–26 (in Japanese with English abstract).

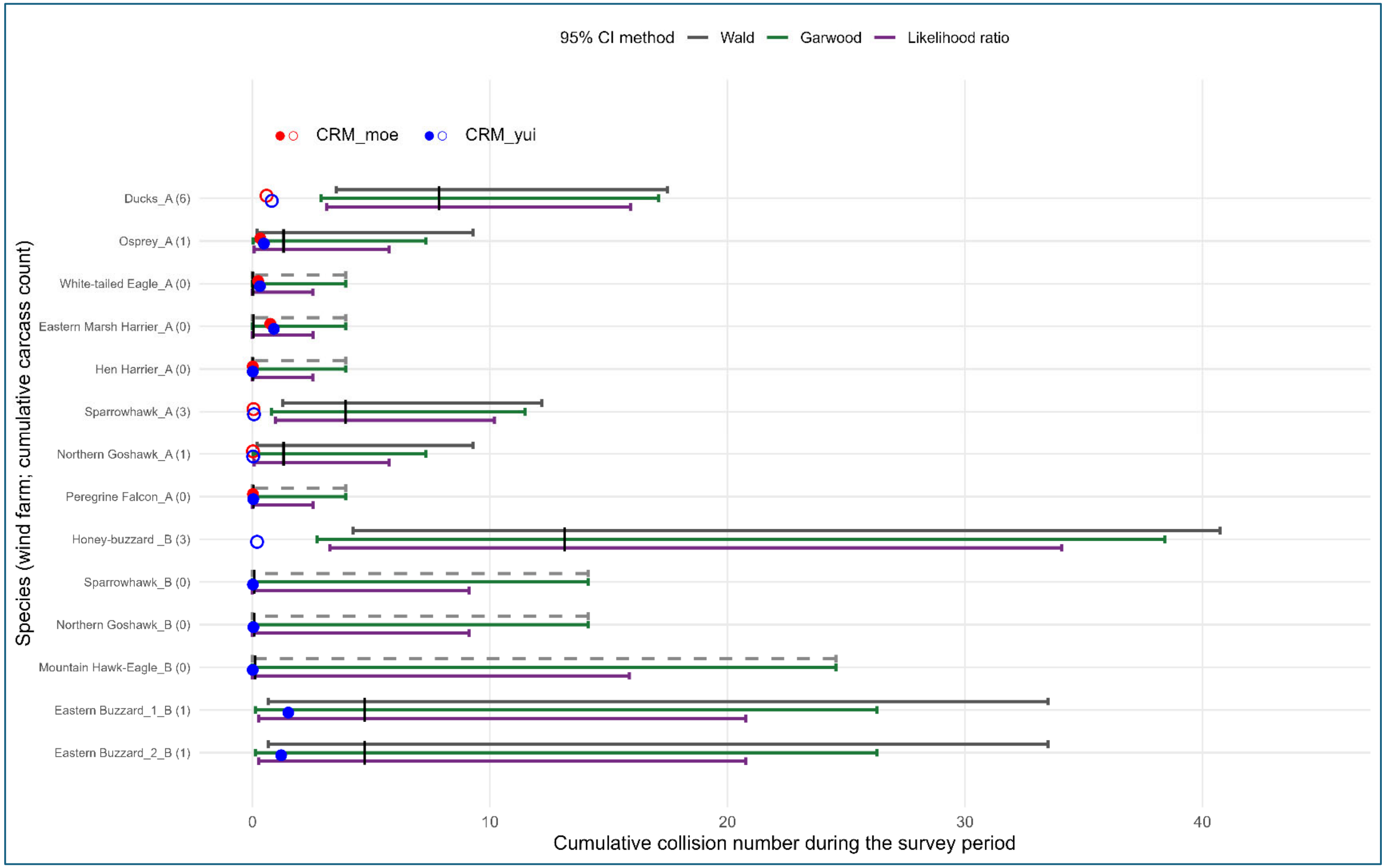

Figure 1. Comparison of cumulative predicted collision numbers ($\sum\lambda_{\mathrm{moe,yr}}$ and $\sum\lambda_{\mathrm{yui,yr}}$) and cumulative collision numbers estimated from carcass survey data ($\hat{\lambda}_{obs,cum}$) for 14 cases. Points indicate CRM predictions, and horizontal lines represent 95% confidence intervals derived from carcass survey data using the frequentist and likelihood-ratio methods. Filled symbols indicate cases where the prediction falls within the confidence interval, whereas open symbols indicate cases in which the prediction falls outside the interval.

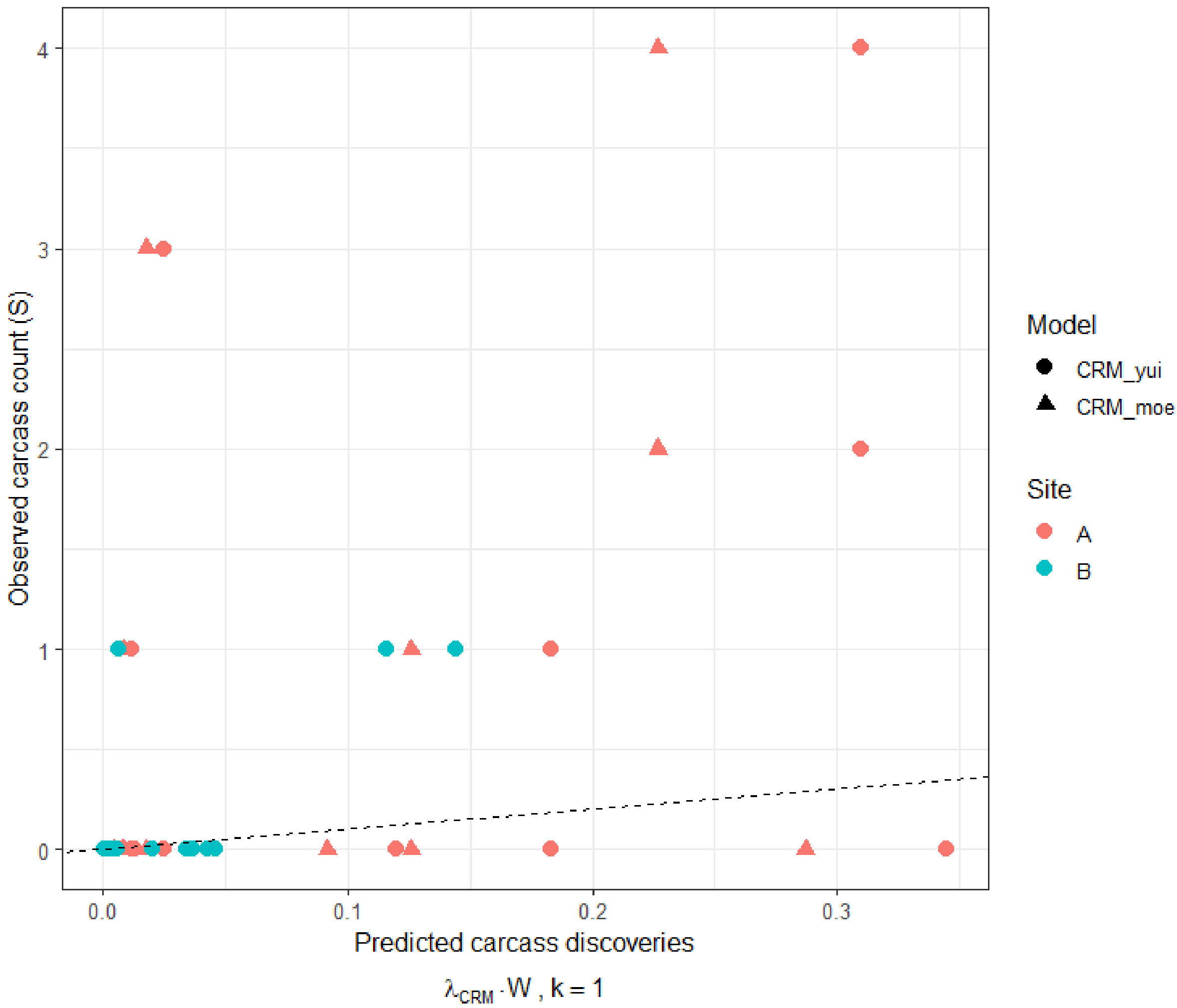

Figure 2. Relationship between predicted and observed carcass discoveries for all cases. Predicted carcass discoveries ($\lambda_{CRM} \cdot W$) were calculated from the CRM_{moe} and CRM_{yui} without scaling (k = 1). Symbols indicate CRM model type (circles: YUI; triangles: MOE), and colors indicate wind farm site (A and B). The dashed line represents the 1:1 relationship.

Table 1. Definitions of the parameters used in the collision risk models. The argument $[yr]$ is used when specifying a particular year.

Symbol	Definition
Observation-based quantities	
S_{cum}	Cumulative number of detected carcasses during the survey period
μ_{cum}	Expected cumulative number of detected carcasses under the Poisson distribution.
W_{cum}	Effective cumulative survey coverage (dimensionless), defined as $\frac{E_{tot}}{n \cdot P_{rd}}$, representing the proportion of the cumulative seasonal presence period covered by carcass surveys during the entire study period.
λ_{cum}	Cumulative number of collisions during the survey period under the Poisson model, defined as $\lambda_{cum} = \mu / W_{cum}$
$\hat{\lambda}_{cum} = \frac{S_{cum}}{W_{cum}}$	Observation-based estimate of cumulative collisions. Observation-based estimate of the cumulative number of collisions during the study period, obtained by correcting the cumulative number of detected carcasses (S_{cum}) using the cumulative survey coverage (W_{cum}).
Fundamental quantities	
θ	Expected daily number of collisions during the presence period of the target species.
P_{rd}	Mean annual presence period (days) of the target species at the site.
n	Number of survey years
Prediction-based quantities	
λ	Expected annual number of collisions during the seasonal presence period, defined as $\lambda = \theta \cdot P_{rd}$.

λ_{moe} Annual predicted collision number by the CRM_{moe}

λ_{yui} Annual predicted collision number by the CRM_{yui}

Detection effort and related indicators, duration of stay of target species.

F_{nd} Detection probability.

$P_{per,yr}$ Carcass persistence probability for year yr, estimated from carcass survey intervals using the formula of Kitano & Shiraki (2013), based on the mean survey interval $I_{yr} = T_{yr}/Num_{yr}$.

T_{yr} Number of days from the first carcass survey to the final carcass survey within a given year.

Num_{yr} Number of carcass surveys conducted in a given year.

Num_i^{std} Standardized number of carcass surveys for record i, calculated by centering $N_{um,i}$ at its overall mean and scaling by its standard deviation (z-score transformation); used as an explanatory variable in the GLM.

E_{cov} Search coverage ratio (the ratio of the effectively searched area to the target survey area). Although search coverage may vary among turbines, years, and vegetation conditions, a representative site-level value was used in this study.

E_{cum} Effort expended on carcass surveys. Defined as the product of the effective observation period $T_{eff,yr}$ and the discovery success probability p_{yr}. Where surveys are conducted over n years, the figures are aggregated.

$$E_{cum} = \sum_{yr=1}^{n} T_{eff,yr} \cdot p_{yr}$$

p_{yr} Probability of carcass detection in a given year. The product of the detection probability F_{nd} the persistence rate R_{yr} and the exploration rate E_{cov}. That is, $p_{yr} = F_{nd} \cdot R_{yr} \cdot E_{cov}$

$T_{eff,yr}$ The effective observation period represents how well the survey period covers the presence period of the target species. For each year, $T_{eff,yr}$ was defined as the smaller of T_{yr} and P_{rd}. Survey days exceeding the seasonal presence period were assumed not to contribute to the estimation for that species.

I_{yr} Mean survey interval (days) for year yr, defined as T_{yr}/Num_{yr}..

G_{rp} Species classification group used as an explanatory variable in GLM analysis.

Confidence interval estimation (1)

β_0 Intercept of the Poisson regression model with an offset term ($\beta_0 = \log \lambda_{cum}$)

$\hat{\beta}_0$ Estimated value of the intercept

$\mathrm{SE}(\hat{\beta}_0)$ Standard error of the estimated intercept

μ_L, μ_U Lower and upper confidence limits of μ estimated by the Garwood method.

$\hat{\lambda}_L, \hat{\lambda}_U$ Lower and upper confidence limits of λ derived from the Garwood method.

$L(\lambda)$ Likelihood function for λ under the Poisson model

$LR(\lambda)$ Likelihood ratio statistic ($LR(\lambda_{cum}) = -2(\log L(\hat{\lambda}_{cum}) - \log L(\lambda_{cum}))$

$\chi^2\,(v, q)$: The q-th quantile of the chi-squared distribution with v degrees of freedom

---

Confidence interval estimation (2)

F_d Daily frequency of entries into the rotor- swept zone

S_p: Slip coefficient.

P_c Contact probability during passage.

Collision risk model parameters

F_{req} Daily frequency of entry into the rotor-swept zone

P_{sg} Probability of passing through the rotor-swept zone without collision, conditional on entry

P_{cnt} Probability of collision during passage through the rotor-swept zone.

A_{vd} Avoidance rate (see (3))

A_{old} Originally assumed avoidance rate

A_{new} Equivalent avoidance rate derived from scaling factors.

O_{tm} Operating time ratio, defined as the total operating time of the wind turbine divided by the total time in a year.

Table 2. Summary of cumulative carcass counts, and survey-related quantities aggregated across years by site and species. S_{cum}, total number of carcasses detected during the study period; $\sum N_{um,yr}$, total number of surveys conducted; G_{rp}, species classification group; E_{cov}, search coverage ratio; P_{per}, carcass persistence rate estimated from survey intervals; $\sum T[yr]$, total observation period across years; $Years$, number of years included in the aggregation; $\sum P_{rd,yr}$, total bird presence days across the study period. Predicted collision numbers ($\lambda_{\mathrm{moe,yr}}$ and $\lambda_{\mathrm{yui,yr}}$) are expressed on an annual basis.

Site	Species	λ_{moe}	λ_{yui}	S_{cum}	$\sum N_{um,yr}$	G_{rp}	E_{cov}	P_{pet}	$\sum T_{yr}$	$Years$	$\sum P_{rd,yr}]$
A	Ducks	0.297	0.406	6	32	waterfowl	0.764	1	246	2	246
A	Osprey	0.165	0.240	1	72	raptor	0.764	1	550	2	550
A	White-tailed Eagle	0.120	0.157	0	46	raptor	0.764	1	364	2	364
A	Eastern Marsh Harrier	0.376	0.451	0	72	raptor	0.764	1	550	2	550
A	Hen Harrier	0.006	0.006	0	46	raptor	0.764	1	364	2	364
A	Sparrowhawk	0.023	0.032	3	96	raptor	0.764	1	730	2	730
A	Northern Goshawk	0.011	0.015	1	96	raptor	0.764	1	730	2	730
A	Peregrine Falcon	0.010	0.017	0	96	raptor	0.764	1	730	2	730
B	Oriental Honey-buzzard		0.039	3	32	raptor	0.823	1	275	5	975
B	Sparrowhawk		0.004	0	32	raptor	0.823	1	275	5	1050
B	Goshawk		0.008	0	32	raptor	0.823	1	275	5	1050
B	Mountain Hawk-Eagle		0.002	0	32	raptor	0.823	1	275	5	1825
B	Eastern Buzzard_1		0.302	1	32	raptor	0.823	1	275	5	1050
B	Eastern Buzzard_2		0.242	1	32	raptor	0.823	1	275	5	1050

Table 3. Summary of cumulative carcass counts (S_{cum}), cumulative predicted collision numbers ($\sum\lambda_{\mathrm{moe,yr}}$ and $\sum\lambda_{\mathrm{yui,yr}}$), and estimated cumulative collision numbers ($\hat{\lambda}_{cum}$), with 95% confidence intervals derived using the log-transformed Wald, Garwood, and likelihood-ratio methods. Asterisks indicate cases in which the predicted collision number fell below the lower limit of the corresponding 95% confidence interval.

Site	Species	S_{cum}	$\sum\lambda_{\mathrm{moe,yr}}$	$\sum\lambda_{\mathrm{yui,yr}}$)	$\hat{\lambda}_{cum}$	95% CI (Wald)		95% CI (Garwood)		95% CI (Likelihood ratio	
						lower	upper	lower	upper	lower	upper
A	Ducks	6	0.594	0.812	7.853	3.528*	17.481*	2.882*	17.094*	3.121*	15.914*
A	Osprey	1	0.330	0.480	1.309	0.184	9.292	0.033	7.293	0.075	5.763
A	White-tailed Eagle	0	0.240	0.314	0.000	0.000	4.828	0.000	3.921	0.000	2.538
A	Eastern Marsh Harrier	0	0.752	0.902	0.000	0.000	4.828	0.000	3.921	0.000	2.550
A	Hen Harrier	0	0.012	0.012	0.000	0.000	4.828	0.000	3.921	0.000	2.538
A	Sparrowhawk	3	0.046	0.064	3.927	1.266*	12.175*	0.810*	11.475*	0.977*	10.182*
A	Northern Goshawk	1	0.022	0.030	1.309	0.184*	9.292 *	0.033*	7.293 *	0.075*	5.763*
A	Peregrine Falcon	0	0.020	0.034	0.000	0.000	4.828	0.000	3.921	0.000	2.562
B	Honey-buzzard	3		0.194	13.141	4.238*	40.745*	2.710*	38.404*	3.268*	34.076*
B	Sparrowhawk	0		0.020	0.000	0.000	17.402	0.000	14.132	0.000	9.130
B	Goshawk_B	0		0.039	0.000	0.000	17.402	0.000	14.132	0.000	9.130
B	Mountain Hawk-Eagle	0		0.011	0.000	0.000	30.246	0.000	24.562	0.000	15.869
B	Eastern Buzzard_1	1		1.510	4.717	0.665	33.488	0.119	26.283	0.269	20.770
B	Eastern Buzzard_2	1		1.210	4.717	0.665	33.488	0.119	26.283	0.269	20.770

Table 4. Comparison of baseline and best-supported Poisson generalized linear models (GLMs) and their overdispersion diagnostics. The baseline model includes only the log-transformed predicted collision number as a predictor, whereas the best-supported model additionally includes survey-related variables. The dispersion ratio and associated P-values were used to assess overdispersion. Akaike's Information Criterion (AIC) values are shown for both Poisson and negative binomial (NB) models. Parameter estimates for the best-supported models are presented in Table 5.

CRM models	Dispersion ratio	P-values	AIC(Poisson)	AIC(NB)	Remarks
CRM_{moe} baseline model	2.16	0.007	42.9	------	----
CRM_{moe} best-supported model	1.11	0.343	33.0	35.0	The best-fit model is presented in Table 5.
CRM_{yui} baseline model	1.20	0.169	67.9	-----	----
CRM_{yui} best-supported model	1.23	0.142	62.5	64.5	The best-fit model is presented in Table 5.

Table 5. Parameter estimates of the selected Poisson generalized linear models (GLMs) for the CRM_{moe} and CRM_{yui}. The response variable is the observed carcass count for record i (S_i), with the log-transformed adjusted survey coverage, $\log W_i$, included as an offset, where W_i denotes the effective survey effort divided by the annual presence period for record i. Explanatory variables include the log-transformed predicted collision number, the standardized number of surveys (Num^{std}), and species classification group (G_{rp}; waterfowl vs. others). The variable was retained only in the selected CRM_{moe}.

Model	Variable	Estimate	SE	Z- value	P -value
CRM_{moe}	Intercept	-7.334	5.097	-1.439	0.150
	$\log(\lambda_{\mathrm{moe}})$	1.253	1.151	1.089	0.276
	Z_{sur}	6.096	4.628	1.317	0.188
	$G_{rp}(waterfowl)$	10.123	6.247	1.620	0.105
CRM_{yui}	Intercept	-0.121	0.631	-0.193	0.847
	$\log(\lambda_{\mathrm{yui}})$	0.129	0.193	0.673	0.501
	$G_{rp}(waterfowl)$	1.606	0.637	2.523	0.012

Table 6. Results of the scaling model relating to observed carcass counts to predicted collision numbers (λ_i). Coefficients were estimated using a Poisson generalized linear model (GLM). The intercept corresponds to the baseline scaling coefficient (k = exp(β_0) for Windfarm A under the CRM_{yui}, whereas the remaining coefficients represent multiplicative effects relative to this baseline.

Coefficients	Estimate	SE	Z- value	P-value	Multiplier exp (Estimate)	95% CI for exp (Estimate)
Intercept*	1.693	0.302	5.616	$1.95 \cdot 10^{-8}$ ***	5.44	3.01 - 9.82
Windfarm_B	0.355	0.539	0.658	0.511	1.43	0.50-4.10
CRM_{moe}	0.273	0.426	0.640	0.522	1.31	0.57 - 3.03

Table 7. Minimum (left) and maximum (right) numbers of observation years (n) required to determine whether the expected number of detected collisions ($p \cdot \lambda$) is significantly different from the cumulative number of detected carcasses (S_{cum}). “20+” or “>20” indicates bounds extending beyond the assumed facility service life of 20 years.

	Cumulative number of detected carcasses (S_{cum})								
$p \cdot \lambda$	0	1	2	3	4	5	6	7	10
0.01	---20+	>20---	>20---	>20---	>20---	>20---	>20---	>20---	>20---
0.05	---20+	10---20+	>20---	>20---	>20---	>20---	>20---	>20---	>20---
0.1	---20+	5---20+	13---20+	>20---	>20---	>20---	>20---	>20---	>20---
0.5	---11	1---20+	3---20+	5---20+	7---20+	9---20+	12---20+	14---20+	>20---
1	---5	1---10	2---14	3---17	4---20+	5---20+	6---20+	7---20+	11---20+
5	---1	1---2	1---2	1---3	1---4	1---4	2---5	2---5	3---7
10	---0	1---1	1---1	1---1	1---2	1---2	1---2	1---2	2---3

Table 8. Estimated avoidance rates derived from carcass survey data and collision numbers predicted by the CRM_{yui}. Lower and upper confidence limits of avoidance rates ($A_{vd,L}$ and $A_{vd,U}$) were calculated using the likelihood-ratio confidence limits (LR_L and LR_U) of the cumulative collision number estimated from carcass survey data ($\hat{\lambda}_{obs,cum}$). Predicted collision numbers assuming no avoidance were obtained by dividing λ_{yui} by ($1 -$ Aassumed A_{vd}), where Aassumed is the avoidance rate assumed in the CRM (0.98 for most species and 0.95 for White-tailed Eagle).

Species_Site_A_B	S_{cum}	λ_{yui}	λ_{yui} without avoidance	LR_L	LR_U	$A_{vd,L}$ with LR_L	$A_{vd,U}$ with LR_U	Assumed A_{vd}
Ducks_A	6	0.406	20.300	1.561	7.957	0.608	0.923*	0.98
Osprey_A	1	0.240	12.000	0.037	2.882	0.760	0.997	0.98
White-tailed Eagle_A	0	0.157	3.140	0.000	1.269	0.596	1.000	0.95
Eastern Marsh Harrier_A	0	0.451	22.550	0.000	1.275	0.943	1.000	0.98
Hen Harrier_A	0	0.006	0.300	0.000	1.269	-3.230	1.000	0.98
Sparrowhawk_A	3	0.032	1.600	0.488	5.091	-2.182	0.695*	0.98
Northern Goshawk_A	1	0.015	0.750	0.037	2.882	-2.842	0.950*	0.98
Peregrine Falcon_A	0	0.017	0.850	0.000	1.281	-0.507	1.000	0.98
Honey-buzzard _B	3	0.039	1.940	0.654	6.815	-2.513	0.663*	0.98
Sparrowhawk_B	0	0.004	0.195	0.000	1.826	-8.364	1.000	0.98
Goshawk_B	0	0.008	0.390	0.000	1.826	-3.682	1.000	0.98
Mountain Hawk-Eagle_B	0	0.002	0.105	0.000	3.174	-29.227	1.000	0.98
Eastern Buzzard_1_B	1	0.302	15.095	0.054	4.154	0.725	0.996	0.98
Eastern Buzzard_2_B	1	0.242	12.095	0.054	4.154	0.657	0.996	0.98

Supplementary Table S1 Tidy dataset used for the generalized linear model analyses

Species / Taxon	Site	λ_{moe}	λ_{yui}	S	N_{um}	G_{rp}	E_{cov}	T	Survey year	P_{rd}
Ducks	A	0.297	0.406	2	16	waterfowl	0.764	123	1	123
Ducks	A	0.297	0.406	4	16	waterfowl	0.764	123	2	123
Osprey	A	0.165	0.240	1	36	raptor	0.764	275	1	275
Osprey	A	0.165	0.240	0	36	raptor	0.764	275	2	275
White-tailed Eagle	A	0.120	0.157	0	23	raptor	0.764	182	1	182
White-tailed Eagle	A	0.120	0.157	0	23	raptor	0.764	182	2	182
Eastern Marsh Harrier	A	0.376	0.451	0	36	raptor	0.764	275	1	275
Eastern Marsh Harrier	A	0.376	0.451	0	36	raptor	0.764	275	2	275
Hen Harrier	A	0.006	0.006	0	23	raptor	0.764	182	1	182
Hen Harrier	A	0.006	0.006	0	23	raptor	0.764	182	2	182
Sparrowhawk	A	0.023	0.032	3	48	raptor	0.764	365	1	365
Sparrowhawk	A	0.023	0.032	0	48	raptor	0.764	365	2	365
Northern Goshawk	A	0.011	0.015	1	48	raptor	0.764	365	1	365
Northern Goshawk	A	0.011	0.015	0	48	raptor	0.764	365	2	365
Peregrine Falcon	A	0.010	0.017	0	48	raptor	0.764	365	1	365
Peregrine Falcon	A	0.010	0.017	0	48	raptor	0.764	365	2	365
Oriental Honey-buzzard	B		0.039	0	8	raptor	0.823	122	1	195
Oriental Honey-buzzard	B		0.039	1	6	raptor	0.823	39	2	195
Oriental Honey-buzzard	B		0.039	1	6	raptor	0.823	39	3	195
Oriental Honey-buzzard	B		0.039	1	6	raptor	0.823	39	4	195
Oriental Honey-buzzard	B		0.039	0	6	raptor	0.823	36	5	195
Sparrowhawk	B		0.004	0	8	raptor	0.823	122	1	210
Sparrowhawk	B		0.004	0	6	raptor	0.823	39	2	210
Sparrowhawk	B		0.004	0	6	raptor	0.823	39	3	210
Sparrowhawk	B		0.004	0	6	raptor	0.823	39	4	210
Sparrowhawk	B		0.004	0	6	raptor	0.823	36	5	210
Goshawk	B		0.008	0	8	raptor	0.823	122	1	210
Goshawk	B		0.008	0	6	raptor	0.823	39	2	210
Goshawk	B		0.008	0	6	raptor	0.823	39	3	210
Goshawk	B		0.008	0	6	raptor	0.823	39	4	210
Goshawk	B		0.008	0	6	raptor	0.823	36	5	210
Mountain Hawk-Eagle	B		0.002	0	8	raptor	0.823	122	1	365
Mountain Hawk-Eagle	B		0.002	0	6	raptor	0.823	39	2	365
Mountain Hawk-Eagle	B		0.002	0	6	raptor	0.823	39	3	365
Mountain Hawk-Eagle	B		0.002	0	6	raptor	0.823	39	4	365
Mountain Hawk-Eagle	B		0.002	0	6	raptor	0.823	36	5	365
Eastern Buzzard_1	B		0.302	1	8	raptor	0.823	122	1	210
Eastern Buzzard_1	B		0.302	0	6	raptor	0.823	39	2	210
Eastern Buzzard_1	B		0.302	0	6	raptor	0.823	39	3	210
Eastern Buzzard_1	B		0.302	0	6	raptor	0.823	39	4	210
Eastern Buzzard_1	B		0.302	0	6	raptor	0.823	36	5	210
Eastern Buzzard_2	B		0.242	1	8	raptor	0.823	122	1	210
Eastern Buzzard_2	B		0.242	0	6	raptor	0.823	39	2	210
Eastern Buzzard_2	B		0.242	0	6	raptor	0.823	39	3	210
Eastern Buzzard_2	B		0.242	0	6	raptor	0.823	39	4	210
Eastern Buzzard_2	B		0.242	0	6	raptor	0.823	36	5	210

N_{um} : Number of carcass surveys. E_{cov} : Search coverage ratio. T : Annual survey period (days). P_{rd} : Annual presence period of the target species (days). The effective observation period used in the GLM was calculated as $T_{eff} = \min(T, P_{rd})$.