# Power to Detect Changes in Occupancy and Abundance of Rare, Imperfectly Detected Species using Single-Season Hierarchical Models

Karl A. Lamothe, Adam S. van der Lee, Scott M. Reid, D. Andrew R. Drake

Fisheries and Oceans Canada
Ontario and Prairie Region
867 Lakeshore Road
Burlington, ON
L7S 1A1

2023

Canadian Manuscript Report of
Fisheries and Aquatic Sciences 3271

Fisheries and Oceans    Pêches et Océans
Canada    Canada

Canada

## Canadian Manuscript Report of Fisheries and Aquatic Sciences

Manuscript reports contain scientific and technical information that contributes to existing knowledge but which deals with national or regional problems.  Distribution is restricted to institutions or individuals located in particular regions of Canada.  However, no restriction is placed on subject matter, and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Manuscript reports may be cited as full publications.  The correct citation appears above the abstract of each report.  Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Manuscript reports are produced regionally but are numbered nationally.  Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-900 in this series were issued as Manuscript Reports (Biological Series) of the Biological Board of Canada, and subsequent to 1937 when the name of the Board was changed by Act of Parliament, as Manuscript Reports (Biological Series) of the Fisheries Research Board of Canada.  Numbers 1426 - 1550 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Manuscript Reports.  The current series name was changed with report number 1551.

## Rapport manuscrit canadien des sciences halieutiques et aquatiques

Les rapports manuscrits contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui traitent de problèmes nationaux ou régionaux.  La distribution en est limitée aux organismes et aux personnes de régions particulières du Canada.  II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports manuscrits peuvent être cités comme des publications à part entière.  Le titre exact figure au-dessus du résumé de chaque rapport.  Les rapports manuscrits sont résumés dans la base de données  *Résumés des sciences aquatiques et halieutiques.*

Les rapports manuscrits sont produits à l'échelon régional, mais numérotés à l'échelon national.  Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 900 de cette série ont été publiés à titre de Manuscrits (série biologique) de l'Office de biologie du Canada, et après le changement de la désignation de cet organisme par décret du Parlement, en 1937, ont été classés comme Manuscrits (série biologique) de l'Office des recherches sur les pêcheries du Canada.  Les numéros 901 à 1425 ont été publiés à titre de Rapports manuscrits de l'Office des recherches sur les pêcheries du Canada.  Les numéros 1426 à 1550 sont parus à titre de Rapports manuscrits du Service des pêches et de la mer, ministère des Pêches et de l'Environnement.  Le nom actuel de la série a été établi lors de la parution du numéro 1551.

2023

# POWER TO DETECT CHANGES IN OCCUPANCY AND ABUNDANCE OF RARE, IMPERFECTLY DETECTED SPECIES USING SINGLE-SEASON HIERARCHICAL MODELS

Karl A. Lamothe[1], Adam S. van der Lee[1], Scott M. Reid[2], and D. Andrew R. Drake[1]

[1]Fisheries and Oceans Canada
Ontario and Prairie Region
867 Lakeshore Road
Burlington, ON L7S 1A1

[2]Ministry of Northern Development, Mines, Natural Resources, and Forestry
Aquatic Research and Monitoring Section
Peterborough, ON K9J 8M5

Correct citation for this publication:

Lamothe, K.A., van der Lee, A.S., Reid, S.M., and Drake, D.A.R. 2023. Power to detect changes in occupancy and abundance of rare, imperfectly detected species using single-season hierarchical models. Can. Manuscr. Rep. Fish. Aquat. Sci. 3271: v + 27 p.

# TABLE OF CONTENTS

# ABSTRACT

Monitoring for species listed under the *Species at Risk Act* often does not begin until a suspected decline in abundance or distribution has occurred, leading to challenges for documenting trends. Here, simulations were performed to evaluate the power and precision of single-season occupancy and *N*-mixture models to detect proportional reductions in occupancy probability and abundance for imperfectly detected species in low abundance between two time periods. The results suggest that many sites and surveys are needed to achieve sufficient statistical power (i.e., 0.80) for detecting change when occupancy probability, detection probability, and abundance are low. For example, quantifying a 30% reduction in occupancy probability for a species with high detection probability (0.7) and moderate occupancy probability (0.5), 200 sites surveyed three times (600 samples) were needed to achieve a power of 0.80. For the same species with a detection probability of 0.30, the number of samples required increased to 1400. Even greater effort was needed to detect significant changes in abundance. Occupancy models generated estimates with greater accuracy and precision than *N*-mixture models for a given level of effort. Overall, the results suggest the need to maximize detection probability for rare species, which will reduce the effort needed to quantify trends with sufficient statistical power.

# RÉSUMÉ

Lamothe, K.A., van der Lee, A.S., Reid, S.M., and Drake, D.A.R. 2023. Power to detect changes in occupancy and abundance of rare, imperfectly detected species using single-season hierarchical models. Can. Manuscr. Rep. Fish. Aquat. Sci. 3271: v + 27 p.

L'abondance et la répartition des espèces inscrites en vertu de la *Loi sur les espèces en péril* doivent faire l'objet d'une surveillance. Dans le cas présent, des simulations ont été effectuées pour évaluer l'efficacité et la précision des modèles d'occupation à une seule saison et des modèles *N*-mixture pour détecter des réductions proportionnelles de la probabilité d'occupation et de l'abondance entre deux périodes chez les espèces mal détectées et en faible abondance. Les résultats laissent entendre que plusieurs sites et relevés sont nécessaires afin d'obtenir une efficacité statistique suffisante (0,8) pour détecter les changements lorsque la probabilité d'occupation, la probabilité de détection et l'abondance sont faibles. Par exemple, pour obtenir une efficacité de 0,8 lors de la quantification d'une réduction de 30 % de la probabilité d'occupation pour une espèce avec une probabilité de détection élevée (0,7) et une probabilité d'occupation modérée (0,5), il a fallu que 200 sites fassent l'objet d'un relevé trois fois (600 échantillons). Pour une espèce avec une faible probabilité de détection (0,3) et une probabilité d'occupation modérée (0,5), le nombre d'échantillons nécessaires est passé à 1 400. Il a fallu déployer encore plus d'efforts pour détecter les changements importants dans l'abondance. Les modèles d'occupation étaient plus exacts et précis que les modèles *N*-mixture pour un niveau d'effort donné. Dans l'ensemble, nos résultats suggèrent la nécessité de maximiser la probabilité de détection pour les espèces rares afin de réduire l'effort requis pour quantifier les tendances avec une efficacité statistique suffisante.

# INTRODUCTION

The conservation status of freshwater fishes in Canada is assessed with criteria that include indicators such as the number of mature individuals, the area of occupancy, and plausible threats to the species, with species designation under the *Species at Risk Act* (SARA) requiring consideration of trends over space and time (COSEWIC 2021). Additionally, trend information is needed for the evaluation of recovery actions, which are implemented once species are listed. Due to the impracticality of completing a census of most wild freshwater fish populations, documenting trends in species status can require the use of multi-year monitoring data combined with statistical models. Typically, mark-recapture is considered to be the best approach for estimating the abundance of freshwater fish populations, providing greater accuracy and precision of abundance estimates relative to depletion-based designs of unmarked individuals (Rosenberger and Dunham 2005; Zipkin et al. 2014). However, issues such as low recapture rate and tag retention have led to variable success when performing mark-recapture studies, particularly for freshwater fish species listed under SARA (Barnucz et al. 2021; Barnucz and Drake 2021). The use of repeat-survey designs for estimating occupancy and abundance of unmarked populations may, therefore, represent a suitable alternative to mark-recapture (MacKenzie et al. 2018; Kéry and Royle 2016).

Repeat-survey designs of unmarked individuals allow the joint modelling of species- or individual-based detection probabilities and estimates of occupancy or abundance when replicated spatially and temporally (MacKenzie et al. 2018; Kéry and Royle 2016). Imperfect detection is a well-recognized phenomenon affected by factors such as sampling gear effectiveness, habitat conditions, and local population size (Dextrase et al. 2014*a*; Dextrase et al. 2014*b*; Lamothe et al. 2019; Lamothe et al. 2023). False absences can lead to incorrect inferences about abiotic or biotic factors important for the persistence of the species (Tyre et al. 2003) and biased estimates of species abundance and distribution (MacKenzie et al. 2002; Kéry and Royle 2016). Statistical models that incorporate species detection probabilities or capture probabilities for individuals may provide more accurate interpretations of the status of SARA-listed fishes and the spatial boundaries of critical habitat.

The collection of repeat-survey data within sites can be achieved many ways, including temporal replicates, records collected by multiple observers, multiple independent collection methods, and spatial subsampling of a site (MacKenzie et al. 2018). Visual counts of unmarked individuals could occur at selected quadrats three to five times over a five-day period (e.g., MacKenzie et al. 2005), or they may occur within a single day. A common approach for sampling SARA-listed freshwater fishes is to perform repeat surveys within a site consecutively, where captured animals are removed from the site after each survey until successive surveys at the site are complete (i.e., depletion sampling), and subsequently moving on to sample the next site (e.g., Dextrase et al. 2014*a*; Lamothe and Drake 2022). By only visiting a site once, time spent travelling among sites is reduced and therefore the number of sites surveyed and associated spatial coverage will be less constrained by project resources. Moreover, repeatedly accessing sites over multiple days and directly (or indirectly) interacting with organisms may impose additional sampling-related harm to individuals, which is another consideration when sampling at risk species. In all cases, evaluating the consequences of using a depletion versus non-depletion survey design is an important step for planning monitoring efforts for imperiled species (MacKenzie and Royle 2005).

The lack of long-term monitoring data for most freshwater fishes is a substantial challenge when assessing conservation status. Investments in monitoring programs generally begin only after a suspected decline has occurred, or a conservation status has been assigned. This constraint can lead to low initial (i.e., baseline) site-specific abundances. As a result of the

difficulty of detecting rare or hard to sample species, and their usual low initial abundance, there is concern about the necessary effort needed to document changes in abundance or occupancy over time for many SARA-listed freshwater fishes. Power analyses can help guide the development of monitoring programs, or aid in the interpretation of monitoring program results, to ensure sufficient rigour when evaluating changes in species abundance and distribution, particularly when detection is imperfect (Guillera-Arroita and Lahoz-Monfort 2012). Prospective power analysis has been used to assess the effectiveness of sampling gear and methods to monitor the status of several Ontario fishes at risk (Reid and Dextrase 2017; Lamothe et al. 2023). However, statistical power is rarely incorporated for distribution- and population-related metrics that inform conservation status. By convention, a power of 0.80 is considered sufficient for most ecological studies (Peterman 1990; Cohen 2001). For the purposes of monitoring changes in the status of imperiled species, a power of 0.80 indicates a 20% probability of failing to identify a true change in abundance or distribution – a conclusion that can have severe consequences for the persistence of species. Similar to chosen significance levels (i.e., $\alpha =$ 0.05), the identified level of power deemed adequate for monitoring will reflect the level of risk tolerance for failing to meet SARA program objectives (e.g., Peterman 1990).

The objectives of this study were to determine how sampling effort (number of sample sites and repeat surveys) and survey design (depletion versus non-depletion sampling) affect statistical power to infer changes in abundance or occupancy probability for species at risk. To achieve these objectives, hierarchical single-season occupancy and $N$-mixture models were used with simulated species presence-absence and abundance data, with detection probabilities representative of imperiled freshwater fishes. Overall, this study provides information for the design of freshwater fish species monitoring programs, particularly as it relates to the use of single-season, single-species occupancy and $N$-mixture models for characterizing changes in occupancy or abundance over time.

## METHODS

Analyses were performed in four steps (Figure 1). First, presence-absence data for an imperfectly detected species were simulated to represent two independent surveys where the population had experienced a reduction in occurrence between surveys. With the simulated data, two single-season, single-species occupancy models (MacKenzie et al. 2002; MacKenzie et al. 2018) were built using the presence-absence data 1) before and 2) after proportional reductions to generate estimates of occupancy probability given that observations were imperfect (Step 2; Figure 1). Using the estimates of occupancy, the power to detect reductions in occupancy between the two surveys was calculated (Step 3; Figure 1). Finally, linear models were used to understand how the input parameters for occupancy models differentially influenced the final results (Step 4; Figure 1). These steps were repeated with simulated capture data (i.e., count data) reflecting an imperfectly detected species experiencing proportional reductions in abundance between two surveys, and using $N$-mixture models (Royle 2004; Kéry et al. 2005) to estimate abundance (Figure 1). The simulations were parameterized to reflect imperiled species with low abundance and low probability of detection. Below, a more detailed overview of each step is provided.
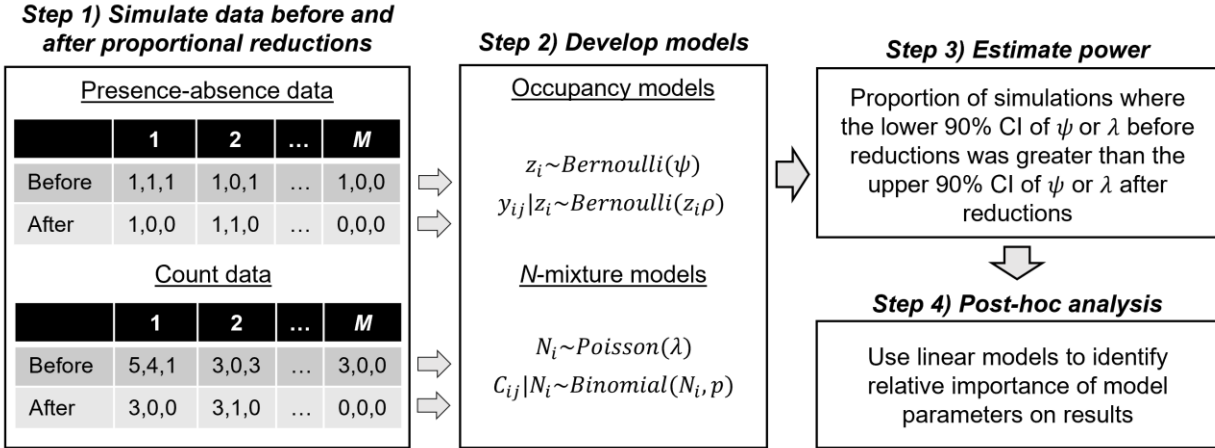
**Figure 1:** Step-by-step description of the analysis. $M$ = number of sites surveyed; $z_i$ = true occupancy state at site $i$; $\psi$ = probability of occupancy; $y_{ij}$ = observed species occurrence at site $i$ during survey $j$; $\rho$ = probability of detecting the species; $N_i$ = true abundance state at site $i$; $\lambda$ = mean expected abundance across all sites; $C_{ij}$ = observed species abundance at site $i$ during survey $j$; and, $p$ = probability of detecting an individual of the species.

## STEP 1) DATA SIMULATION

Presence-absence data were simulated for a single species experiencing proportional reductions ($R$) in occupancy probability ($\psi$) measured between two time periods. Sampling was simulated to occur across $M$ sites ($m$ = 10, …, 200) using $J$ repeat surveys ($j$ = 3, 5, 7) with defined $\psi$ and detection probability ($\rho$; Table 1). Proportional reductions in $\psi$ were performed while holding $\rho$ constant. Initial site-specific presence-absence data were drawn from a binomial distribution with a probability equal to $\psi$. Two survey designs ($g$) were implemented: depletion and non-depletion sampling. Depletion sampling describes the situation where an individual is sampled in survey $j$ and removed until all surveys at site $i$ are complete. Non-depletion sampling replaces individuals prior to performing subsequent surveys. In both scenarios, sites were assumed to be closed to immigration or emigration during the sampling period. Given the common challenge of data scarcity faced by conservation practitioners, we incorporated two sampling periods (as opposed to many years of data). Parameterization for the simulation was done to represent scenarios where the species showed low detection probability and variable levels of occupancy (Table 1). Proportional reduction ($R$) values were chosen to represent thresholds used by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) for evaluating conservation status (-30%, -50%, and -70%; Table 1). One thousand repetitions were performed for each combination of parameter values ($M,\ J,\ \psi,\ \rho,\ g$), totaling 810,000 repetitions.

**Table 1:** Parameters, abbreviations, and values included in the data simulations.

| Abbr. | Variable | Values |
|---|---|---|
| $J$ | Number of repeat surveys per site | 3, 5, 7 |
| $M$ | Number of sites | 10, 25, 50, 100, 200 |
| $\lambda$ | Mean site-specific species abundance | 5, 15, 25 |
| $p$ | Probability of detecting an individual of a species | 0.05, 0.15, 0.25 |
| $R$ | Proportional reduction in abundance or occupancy probability | -0.3, -0.5, -0.7 |
| $g$ | Survey design | Depletion or non-depletion |
| $\psi$ | Simulated probability of site-level species occupancy | 0.25, 0.50, 0.75 |
| $\rho$ | Probability of detecting a species | 0.3, 0.5, 0.7 |

Capture (count) data were simulated to represent an imperfectly detected species experiencing proportional reductions ($R$) in simulated site-specific abundance ($\lambda$) between two time periods. Initial $N$ values were drawn from a Poisson distribution with mean $\lambda$ for $M$ sites surveyed $J$ times, with observed abundances from survey $j$ at site $i$ influenced by the mean probability of detecting an individual of the species ($p$; Table 1). Similar to the occupancy simulations, a reduction in abundance of magnitude $R$ was applied while holding $p$ constant to produce a second data set of observed captures after experiencing $R$. One thousand repetitions were performed for each unique parametrization ($M, J, \lambda, p, g$), totaling 810,000 repetitions. Both the presence-absence and abundance simulations assume independence of sampling sites before and after applying proportional reductions (Guillera-Arroita and Lahoz-Monfort 2012). Note that $\rho$ can be approximately transformed from $p$, using: $\rho = 1 - (1 - p)^{\lambda}$.

## STEP 2) HIERARCHICAL MODELS

Hierarchical models are now common in the species monitoring literature as they allow the joint modelling of ecological and observational processes for unmarked individuals. Here, two of the most common hierarchical models, single-species occupancy models (MacKenzie et al. 2002; MacKenzie et al. 2018) and *N*-mixture models (Royle 2004; Kéry et al. 2005), were used to make independent estimates of mean occupancy $\hat{\psi}$ and mean site-level abundance $\hat{\lambda}$ before and after proportional reductions. Independent single-species occupancy models (MacKenzie et al. 2002) were developed on the simulated presence-absence data before and after (i.e., stacked) experiencing $R$ in $\psi$ across all combinations of $M, J, \psi, g,$ and $\rho$. For these models, the occupancy state ($z_i$) was modeled as $z_i \sim Bernoulli(\psi)$ and the observation process was modeled as $y_{ij} | z_i \sim Bernoulli(z_i \rho)$ where $y_{ij}$ is the observed survey data and $\rho$ is the probability of detecting the species (Figure 1).

Similarly, *N*-mixture models (Royle 2004, Kéry et al. 2005) were developed with the simulated data before and after $R$ for different combinations of $M, J, \lambda, g,$ and $p$. For these models, the latent abundance state was modeled as $N_i \sim Poisson(\lambda)$ and the observation process was modeled as $C_{ij} | N_i \sim Binomial(N_i, p)$ where $\lambda$ is the expected abundance (mean over all sites), $C_{ij}$ is the count data at site $i$ during survey $j$, and $p$ is the probability of detecting an individual of the target species (Figure 1; Kéry and Royle 2016). For depletion sampling,

where individual fish are removed between surveys, detection probability at a site for a given survey is conditional on the encounter history from the previous survey. The observation process follows a multinomial distribution and was modeled as $C_{ij}|N_i \sim Multinomial(N_i, \pi(p))$ where $\pi(p)$ is a function describing the detection probably across surveys given the encounter probability (Kéry and Royle 2016). For example, the detection probability for an individual captured in the first survey was $\pi_1 = p_1$, during the second survey was $\pi_2 = (1 - p_1)p_2$, and during the third survey was $\pi_3 = (1 - p_1)(1 - p_2)p_3$.

Occupancy and *N*-mixture models were developed using the 'unmarked' package (Fiske and Chandler 2011) within the R software (R Core Team 2020). Single-season occupancy models were built using the 'occu' function with the default arguments. *N*-mixture models using non-depletion and depletion survey designs were developed with the default arguments using the 'pcount' and 'gmultmix' functions, respectively. Violin plots were generated to visualize the initial and post-reduction estimates of $\hat{\psi}$ and $\hat{\lambda}$ across parameter combinations using the 'ggplot2' package (Wickham 2016).

## STEP 3) POWER ESTIMATES

Power to detect proportional changes ($R$) in occupancy probability ($\psi$) and site-level abundance ($\lambda$) was defined as the proportion of simulations where the lower 90% confidence interval (CI) of the estimated initial occupancy ($\hat{\psi}_0$) or abundance ($\hat{\lambda}_0$) was greater than the upper 90% CI of the estimated post-reduction occupancy probability ($\hat{\psi}_1$) or abundance ($\hat{\lambda}_1$; Figure 1). Overlap of confidence intervals does not always imply statistical significance; however, it does provide an estimate of confidence that a change has occurred. Tile plots were created to visualize power for each parameterization of the occupancy and *N*-mixture model simulations using the 'ggplot2' package (Wickham 2016) within the R software (R Core Team 2020). Power estimates were grouped to aid interpretation as bad (< 0.50), poor (0.50-0.79), good (0.80-0.89), great (0.90-0.99), or perfect (1.00). The coefficient of variation (CV) was calculated to compare the precision of $\hat{\psi}_0$, $\hat{\psi}_1$, $\hat{\lambda}_0$, and $\hat{\lambda}_1$ estimates across parameterizations.

## STEP 4) POST-HOC ANALYSIS

Linear models were used to understand the relative importance of the number sites ($M$), the number of surveys ($J$), detection probability ($\rho$ or $p$), and survey design ($g$) when estimating occupancy ($\hat{\psi}$) and abundance ($\hat{\lambda}$; Figure 1). The absolute difference in $\hat{\psi}_0$ and $\psi$ was first modelled as a function of $M$, $J$, $g$, and $\rho$. Similarly, the absolute difference in $\hat{\lambda}_0$ and $\lambda$ was modelled as a function of $M$, $J$, $g$, and $p$. Variables were converted to $z$-scores prior to modelling to allow for comparisons of effect sizes. Negative estimates of predictors suggest a positive effect on the accuracy of $\hat{\psi}_0$ and $\hat{\lambda}_0$.

## RESULTS

### OCCUPANCY MODELS

The accuracy (Figure 2; Figure 3) and precision (Figure 4) of occupancy models improved with increased occupancy probability ($\psi$), detection probability ($\rho$), number of survey sites ($M$), and the number of surveys per site ($J$). The number of survey sites, $M$, had the greatest influence on the accuracy of $\hat{\psi}_0$, followed by $\rho$, $J$, and $g$ (Table 2). A non-depletion survey design led to greater accuracy of $\hat{\psi}$ than a depletion design (Table 2; Figure 2; Figure 3). For example, the average $\hat{\psi}_0 \pm 1$ SD was 0.30 ± 0.16 and 0.25 ± 0.06 for depletion versus non-depletion survey

designs, respectively, when $\psi = 0.25$, $M = 100$, $J = 5$, $p = 0.3$, and $R = 0.7$. The biggest difference in mean $\hat{\psi}$ and $\psi$ was observed in simulations with $\rho = 0.3$, $\psi = 0.25$, and $J = 3$; differences neared 0 with increasing $\rho$, $\psi$, and $J$ (Table 3). Approximately 95% of the simulations captured the true $\psi$, or post-reduction $\psi$, within the 90% confidence intervals of $\hat{\psi}_0$ and $\hat{\psi}_1$, respectively, regardless of $\rho$ or $J$ (Table 4).
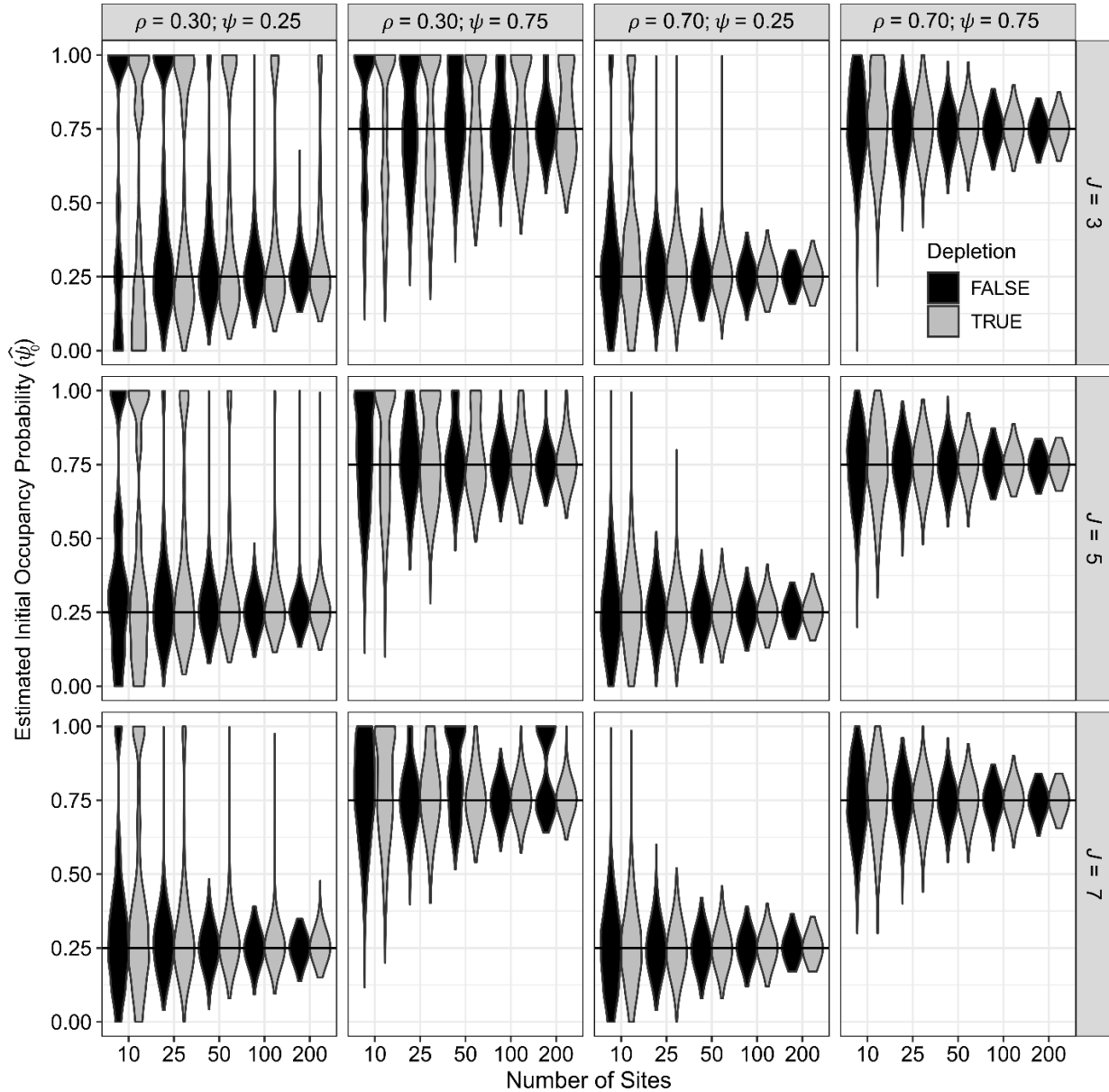


**Figure 2:** Estimated initial occupancy probability ($\hat{\psi}_0$) when sampling was performed using a depletion (grey) or non-depletion (black) design across 10, 25, 50, 100, and 200 sites and using $J = 3$, 5, or 7 repeat surveys. Plotted are simulations where 1) detection probability ($\rho$) = 0.30 and initial simulated occupancy probability ($\psi$) = 0.25, 2) $\rho = 0.30$ and $\psi = 0.75$, 3) $\rho = 0.70$ and $\psi = 0.25$, and 4) $\rho = 0.70$ and $\psi = 0.75$. Black lines represent simulated $\psi$. The violin plots are scaled to have the same maximum width.
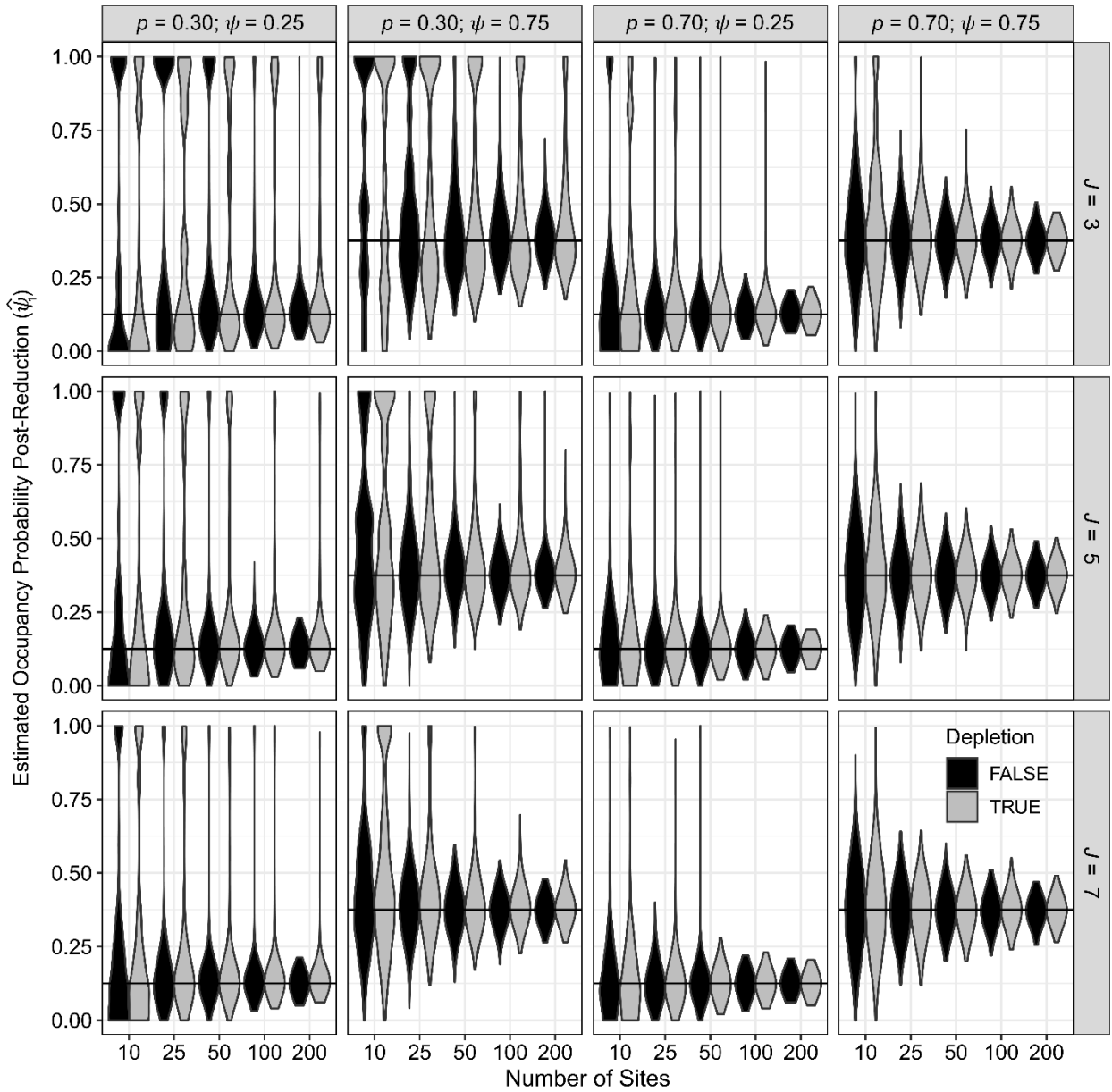
**Figure 3:** Estimated occupancy probability after a 50% reduction ($\hat{\psi}_1$) when sampling was performed using a depletion (grey) or non-depletion (black) design across 10, 25, 50, 100, and 200 sites and using $J$ = 3, 5, or 7 repeat surveys. Plotted are simulations where 1) detection probability ($\rho$) = 0.30 and initial simulated occupancy probability ($\psi$) = 0.25, 2) $\rho$ = 0.30 and $\psi$ = 0.75, 3) $\rho$ = 0.70 and $\psi$ = 0.25, and 4) $\rho$ = 0.70 and $\psi$ = 0.75. Black lines represent $\psi * (1 - R)$. The violin plots are scaled to have the same maximum width.

**Table 2:** Results of the modelled differences in the mean initial occupancy estimate ($\hat{\psi}_0$) and true occupancy ($\psi$), and the mean initial abundance estimate ($\hat{\lambda}_0$) and true abundance (). Negative effect sizes indicate an increase in accuracy.

| Model | Variable | Estimate | Std. Error | t-value | p-value |
|---|---|---|---|---|---|
| $\hat{\psi}_0 - \psi$ | Depletion design, $g$, 0: false, 1: true | 0.15 | 0.002 | 73.63 | <0.001 |
| $\hat{\psi}_0 - \psi$ | Detection probability, $\rho$ | -0.23 | 0.001 | -228.49 | <0.001 |
| $\hat{\psi}_0 - \psi$ | Number of sites, $M$ | -0.32 | 0.001 | -324.48 | <0.001 |
| $\hat{\psi}_0 - \psi$ | Number of surveys, $J$ | -0.17 | 0.001 | -168.24 | <0.001 |
| $\hat{\lambda}_0 - \lambda$ | Depletion design $g$, 0: false, 1: true | -0.25 | 0.002 | -125.26 | <0.001 |
| $\hat{\lambda}_0 - \lambda$ | Detection probability, $p$ | -0.36 | 0.001 | -371.83 | <0.001 |
| $\hat{\lambda}_0 - \lambda$ | Number of sites, $M$ | -0.23 | 0.001 | -236.11 | <0.001 |
| $\hat{\lambda}_0 - \lambda$ | Number of surveys, $J$ | -0.16 | 0.001 | -159.28 | <0.001 |

**Table 3:** Accuracy of occupancy probability estimates indicated by the mean difference (± SD) in estimated initial ($\hat{\psi}_0$) and true occupancy probability ($\psi$) and estimated post reduction occupancy ($\hat{\psi}_1$) and true post reduction occupancy ($\psi * (1 - R)$) when $J$ = 3, 5, 7, and $\rho$ = 0.3, 0.5, 0.7.

| Metric | J | $\rho$ | $\psi = 0.25$ | $\psi = 0.50$ | $\psi = 0.75$ |
|---|---|---|---|---|---|
| $\hat{\psi}_0$ | 3 | 0.3 | 0.16 ± 0.31 | 0.08 ± 0.24 | 0.03 ± 0.18 |
| $\hat{\psi}_0$ | 3 | 0.5 | 0.07 ± 0.21 | 0.03 ± 0.15 | 0.02 ± 0.12 |
| $\hat{\psi}_0$ | 3 | 0.7 | 0.02 ± 0.14 | 0.01 ± 0.11 | 0.01 ± 0.09 |
| $\hat{\psi}_0$ | 5 | 0.3 | 0.08 ± 0.22 | 0.04 ± 0.17 | 0.02 ± 0.13 |
| $\hat{\psi}_0$ | 5 | 0.5 | 0.02 ± 0.13 | 0.01 ± 0.11 | 0.01 ± 0.10 |
| $\hat{\psi}_0$ | 5 | 0.7 | 0.003 ± 0.09 | 0.003 ± 0.10 | 0.0002 ± 0.08 |
| $\hat{\psi}_0$ | 7 | 0.3 | 0.04 ± 0.16 | 0.02 ± 0.13 | 0.03 ± 0.12 |
| $\hat{\psi}_0$ | 7 | 0.5 | 0.01 ± 0.10 | 0.003 ± 0.10 | 0.005 ± 0.09 |
| $\hat{\psi}_0$ | 7 | 0.7 | -0.0003 ± 0.08 | -0.00006 ± 0.09 | -0.00003 ± 0.08 |
| $\hat{\psi}_1$ | 3 | 0.3 | 0.20 ± 0.36 | 0.15 ± 0.32 | 0.12 ± 0.28 |
| $\hat{\psi}_1$ | 3 | 0.5 | 0.10 ± 0.26 | 0.07 ± 0.22 | 0.05 ± 0.19 |
| $\hat{\psi}_1$ | 3 | 0.7 | 0.04 ± 0.17 | 0.02 ± 0.14 | 0.02 ± 0.12 |
| $\hat{\psi}_1$ | 5 | 0.3 | 0.11 ± 0.28 | 0.08 ± 0.23 | 0.06 ± 0.20 |
| $\hat{\psi}_1$ | 5 | 0.5 | 0.03 ± 0.17 | 0.02 ± 0.13 | 0.02 ± 0.12 |
| $\hat{\psi}_1$ | 5 | 0.7 | 0.01 ± 0.09 | 0.004 ± 0.09 | 0.003 ± 0.09 |
| $\hat{\psi}_1$ | 7 | 0.3 | 0.06 ± 0.22 | 0.04 ± 0.17 | 0.03 ± 0.15 |
| $\hat{\psi}_1$ | 7 | 0.5 | 0.01 ± 0.12 | 0.01 ± 0.10 | 0.005 ± 0.10 |
| $\hat{\psi}_1$ | 7 | 0.7 | 0.002 ± 0.07 | 0.002 ± 0.08 | 0.001 ± 0.09 |

**Table 4:** Accuracy of occupancy probability estimates indicated by the proportion of simulations where the upper and lower 90% confidence intervals of the estimated initial mean occupancy probability ($\hat{\psi}_0$) and post-reduction mean occupancy ($\hat{\psi}_1$) were less than or greater than the simulated initial occupancy probability ($\psi$) and the simulated occupancy probability post-reduction ($(\psi * (1 - R))$), respectively.

| Metric | $\psi$ | $\rho =$ | J = 3 <br> 0.3 | 3 <br> 0.5 | 3 <br> 0.7 | 5 <br> 0.3 | 5 <br> 0.5 | 5 <br> 0.7 | 7 <br> 0.3 | 7 <br> 0.5 | 7 <br> 0.7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LCI $\hat{\psi}_0>$ | 0.25 | | 0.054 | 0.056 | 0.059 | 0.057 | 0.060 | 0.059 | 0.054 | 0.063 | 0.054 |
| LCI $\hat{\psi}_0>$ | 0.50 | | 0.043 | 0.052 | 0.055 | 0.058 | 0.057 | 0.055 | 0.052 | 0.058 | 0.055 |
| LCI $\hat{\psi}_0>$ | 0.75 | | 0.027 | 0.055 | 0.059 | 0.052 | 0.051 | 0.050 | 0.049 | 0.056 | 0.052 |
| UCI $\hat{\psi}_0<$ | 0.25 | | 0.030 | 0.029 | 0.033 | 0.027 | 0.027 | 0.036 | 0.026 | 0.032 | 0.038 |
| UCI $\hat{\psi}_0<$ | 0.50 | | 0.044 | 0.040 | 0.041 | 0.035 | 0.039 | 0.052 | 0.037 | 0.047 | 0.054 |
| UCI $\hat{\psi}_0<$ | 0.75 | | 0.046 | 0.039 | 0.043 | 0.043 | 0.039 | 0.049 | 0.039 | 0.046 | 0.051 |
| LCI $\hat{\psi}_1>$ | 0.25*(1-R) | | 0.050 | 0.056 | 0.051 | 0.057 | 0.050 | 0.045 | 0.062 | 0.046 | 0.043 |
| LCI $\hat{\psi}_1>$ | 0.50*(1-R) | | 0.054 | 0.059 | 0.056 | 0.058 | 0.054 | 0.053 | 0.054 | 0.055 | 0.053 |
| LCI $\hat{\psi}_1>$ | 0.75*(1-R) | | 0.047 | 0.057 | 0.053 | 0.058 | 0.052 | 0.050 | 0.056 | 0.049 | 0.048 |
| UCI $\hat{\psi}_1<$ | 0.25*(1-R) | | 0.027 | 0.021 | 0.027 | 0.021 | 0.026 | 0.031 | 0.022 | 0.029 | 0.035 |
| UCI $\hat{\psi}_1<$ | 0.50*(1-R) | | 0.036 | 0.029 | 0.036 | 0.030 | 0.035 | 0.042 | 0.030 | 0.040 | 0.047 |
| UCI $\hat{\psi}_1<$ | 0.75*(1-R) | | 0.042 | 0.035 | 0.040 | 0.034 | 0.038 | 0.043 | 0.035 | 0.045 | 0.046 |

Precision of $\hat{\psi}$ was poor when $\psi$ was relatively low, regardless of $\rho$ or $J$ (Figure 4). Sampling without depletion led to greater levels of precision (Figure 4); for example, the CV of $\hat{\psi}_0$ was 0.52 and 0.22 when following a depletion and non-depletion design, respectively, where $\psi = 0.25$, $M = 100$, $J = 5$, and $\rho = 0.3$. However, the difference in precision between survey designs diminished with increased $\rho$ (Figure 4). For the same scenario where $\rho$ is increased to 0.75, the CV of $\hat{\psi}_0$ was 0.13 when following a depletion designed compared to 0.09 when following a non-depletion design.
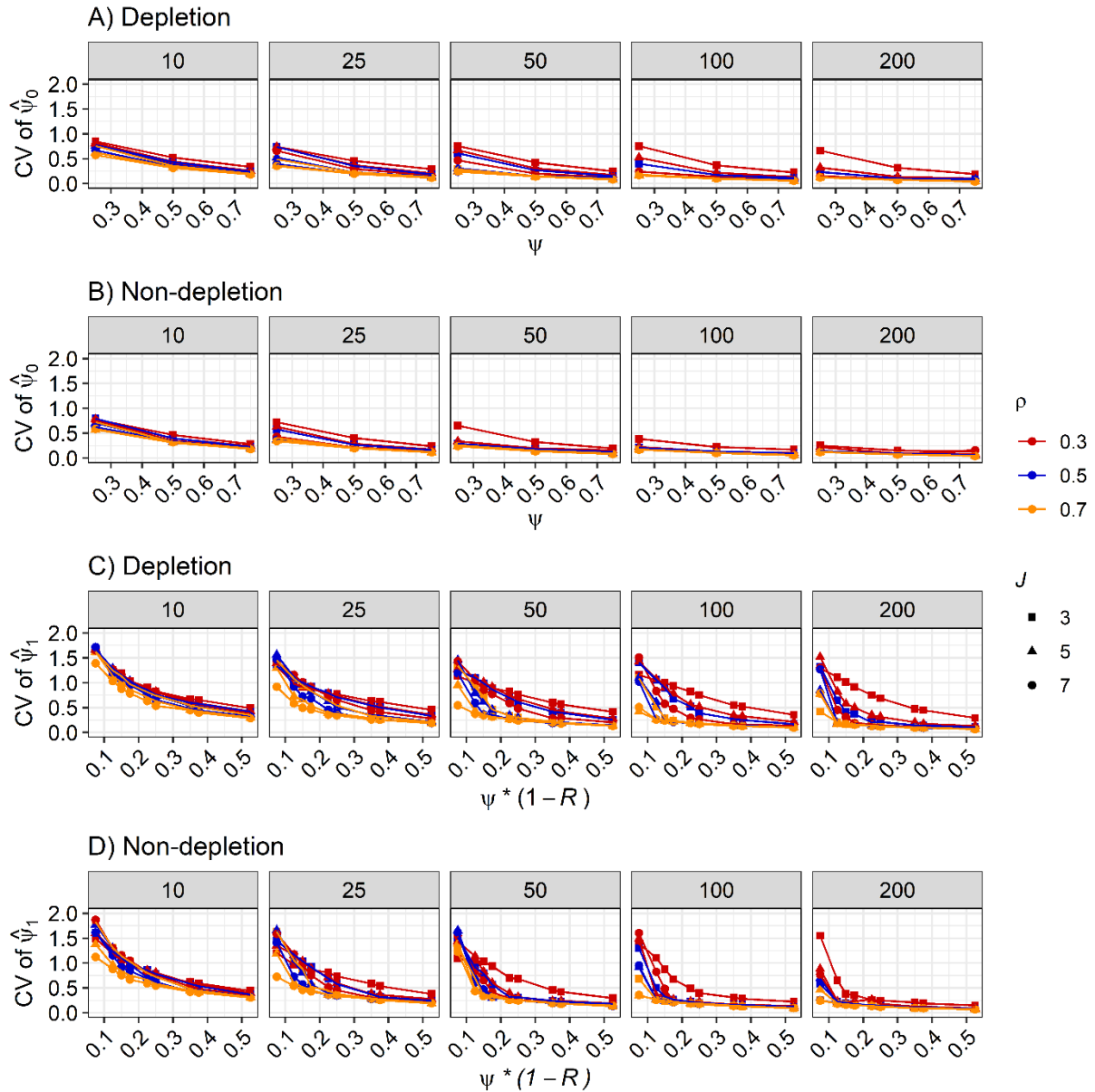
**Figure 4:** Coefficient of variation (CV) of the estimated occupancy probability ($\hat{\psi}$) before (A, B) and after (C, D) experiencing a proportional reduction ($R$) in simulated occupancy ($\psi$) across detection probabilities ($\rho$; colour), number of surveys ($J$; shape), and number of sites (panels; 10, 25, 50, 100, 200) following a depletion (A, C) or non-depletion (B, D) design.

The power to detect proportional reductions in occupancy probability ($\psi$) was low, particularly when detection probability ($\rho$), proportional reductions ($R$), and initial occupancy probability ($\psi_0$) were low (Figure 5). Ten sites were never sufficient to characterize changes in $\psi$, regardless of the survey design and species characteristics (i.e., $\psi_0$, $\rho$; Figure 5). Furthermore, 50 sites were insufficient (power < 0.80) for identifying a 30% change in $\psi$ (Figure 5). When species showed an $\psi$ = 0.50 or 0.75, 200 sites were sufficient to characterize a reduction of 70% if $\rho \geq 0.5$ and $J$ = 5 (Figure 5). Power to detect changes in occupancy was

improved by following a non-depletion survey design (Figure 5). For example, the power to detect a 70% change in occupancy was 0.904 when sampling was performed following a non-depletion design with $J = 3$, $M = 50$, $\psi = 0.5$, and $\rho = 0.7$, compared to a power of 0.563 when performing a depletion design with the same parameters. Power would be improved to 0.869 with the depletion design if the number of surveys $J$ was increased to 5 and the other parameters remained constant, representing an increased total effort of 100 surveys (i.e., 50 sites * 2 additional surveys).
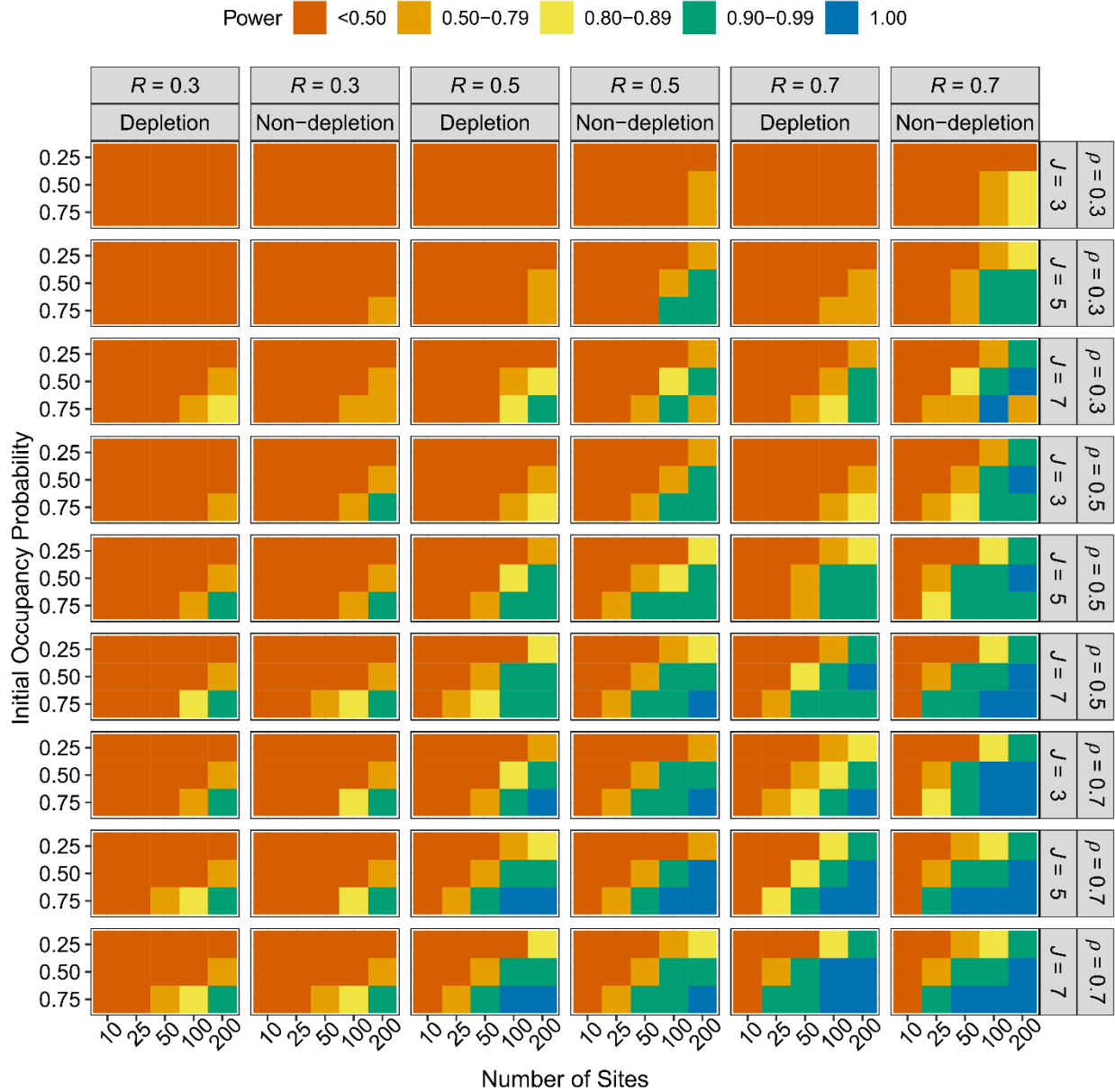
**Figure 5:** Power to detect proportional reductions ($R$ = 0.3, 0.5, 0.7) in occupancy probability when sampling is performed at 10, 25, 50, 100, or 200 survey sites using a differing number of surveys ($J$ = 3, 5, 7). Power was calculated for detecting a change in occupancy probability for species with different detection probabilities ($\rho$ = 0.3, 0.5, 0.7) and occupancy probabilities ($\psi$ = 0.25, 0.50, 0.7) when sampling was performed using a depletion or non-depletion survey design, respectively. The interpretation of power is red = bad (< 0.50), orange = poor (0.50-0.79), yellow = good (0.80-0.89), green = great (0.90-0.99), blue = perfect (1.00).

## N-MIXTURE MODELS

The ability of the *N*-mixture models to provide accurate estimates at low abundances was poor, but improved with increasing detection probability ($p$), number of survey sites ($M$), and number of surveys performed per site ($J$; Figure 6; Figure 7**Figure**). Estimates of mean abundance ($\hat{\lambda}$) followed a bimodal distribution when $p$ and $M$ were low (Figure 6; Figure 7), indicative of boundary estimates and poor model fit. Detection probability, $p$, had the greatest influence on the accuracy of $\hat{\lambda}_0$, followed by survey design ($g$), $M$, and $J$ (Table 2). Contrary to the $\hat{\psi}_0$ estimates, $\hat{\lambda}_0$ estimates improved with a depletion design (Table 2; Figure 6**Figure**). For example, the average $\hat{\lambda}_0 \pm 1$ SD was 5.85 ± 4.98 and 6.49 ± 6.40 for depletion versus non-depletion, respectively, when $\lambda = 5$, $M = 50$, $J = 3$, $p = 0.25$, and $R = 0.3$. On average, *N*-mixture models typically overestimated $\lambda$ and $\lambda * (1 - R)$, with variance of estimates improving with $J$ and $p$ (Table 5). The proportion of simulations that identified $\lambda$ and $\lambda * (1 - R)$ within the 90% confidence intervals before and after proportional reductions improved with $J$ and $p$ (Table 6).
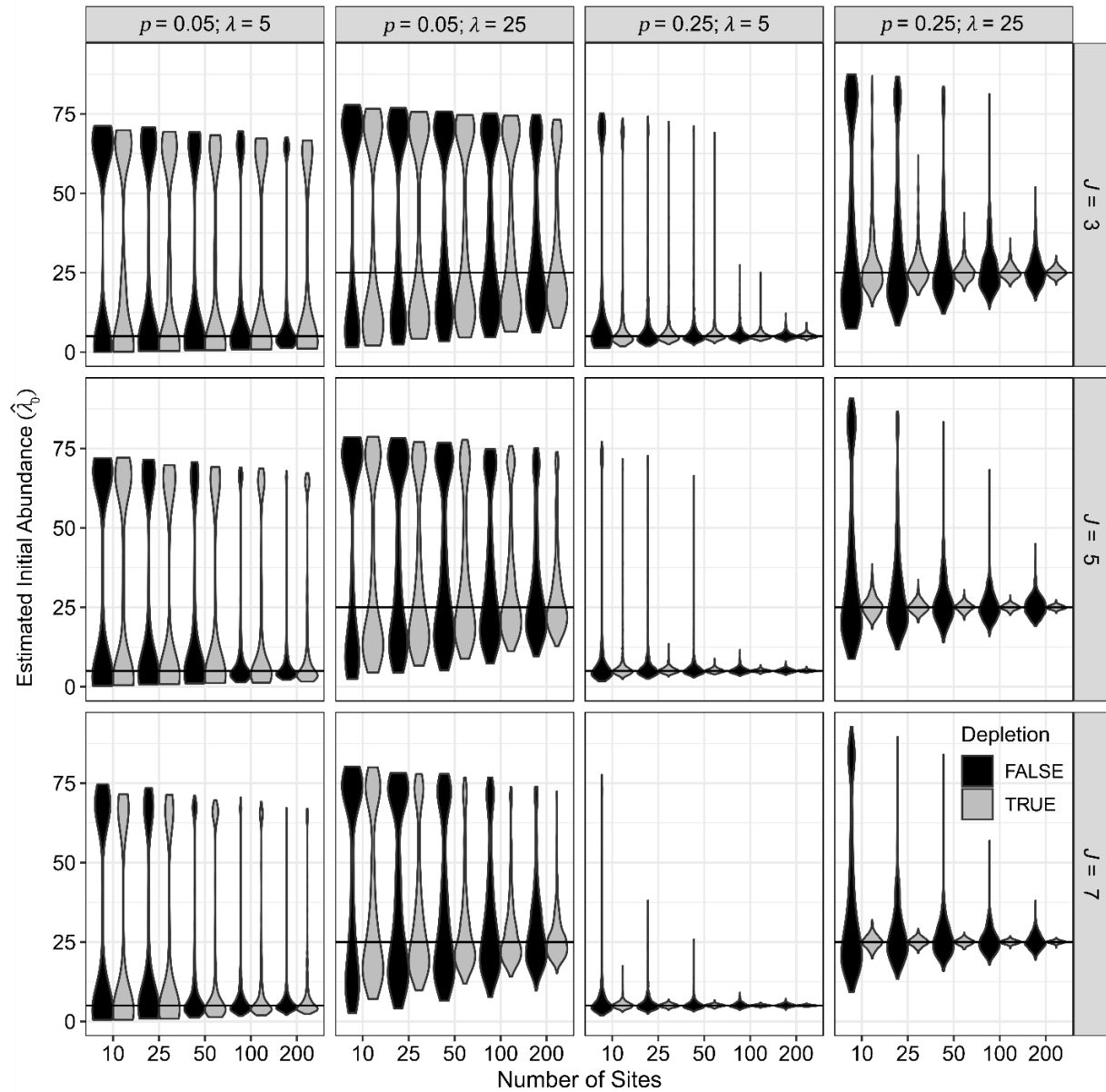
**Figure 6:** Initial estimated site-level abundance ($\hat{\lambda}_0$) when sampling was performed using a depletion (grey) or non-depletion (black) design across 10, 25, 50, 100, and 200 sites and using $J = 3$, 5, or 7 repeat surveys, where 1) detection probabilities ($p$) = 0.05 and simulated abundance ($\lambda$) = 5, 2) $p$ = 0.05 and $\lambda$ = 25, 3) $p$ = 0.25 and $\lambda$ = 5, and 4) $p$ = 0.25 and $\lambda$ = 25. Black lines represent simulated $\lambda$. The violin plots are scaled to have the same maximum width.
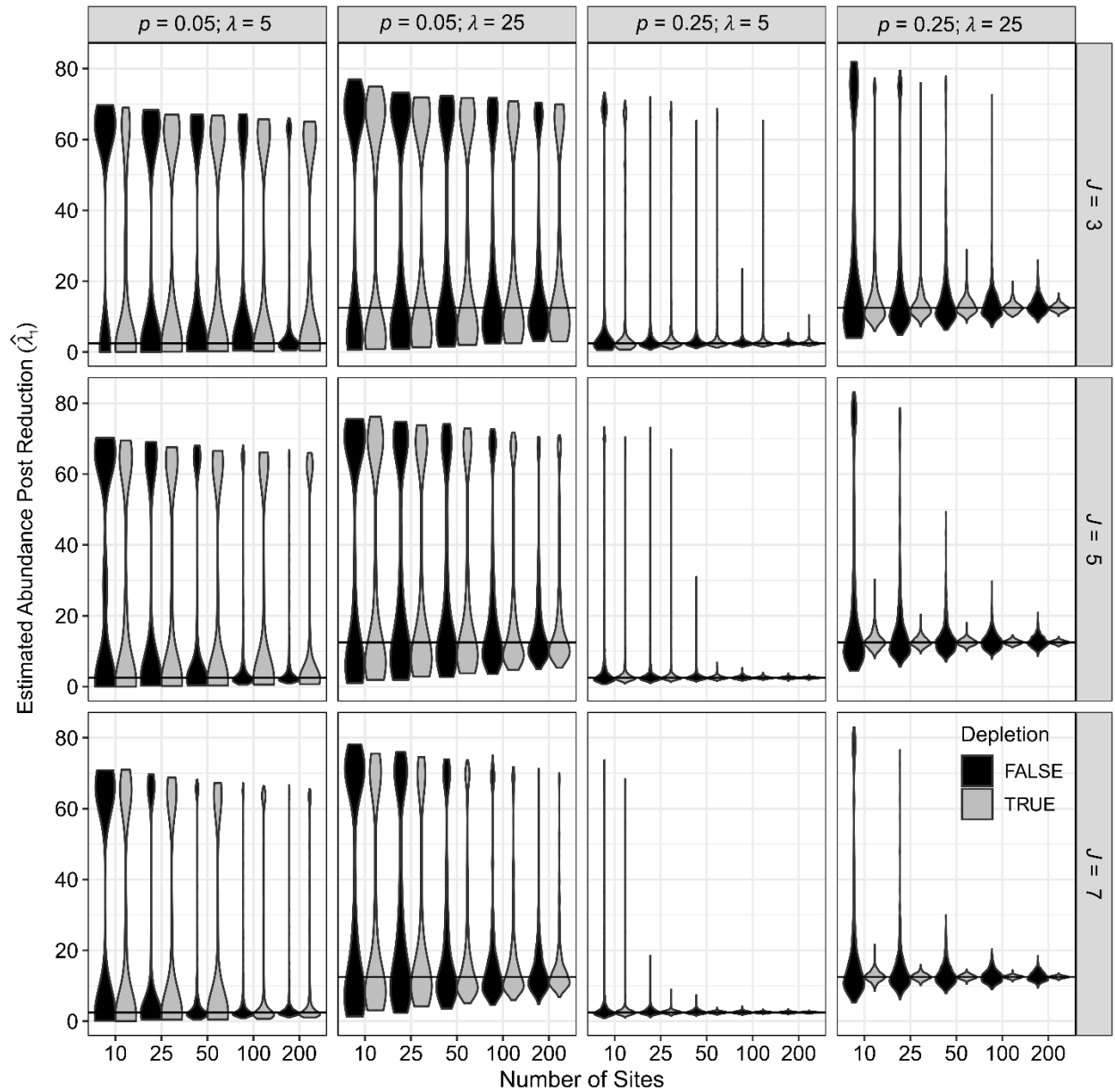
**Figure 7:** Initial post-reduction abundance ($\hat{\lambda}_1$) when sampling was performed using a depletion (grey) or non-depletion (black) design across 10, 25, 50, 100, and 200 sites and using $J$ = 3, 5, or 7 repeat surveys, where 1) detection probabilities ($p$) = 0.05 and simulated abundance ($\lambda$) = 5, 2) $p$ = 0.05 and $\lambda$ = 25, 3) $p$ = 0.25 and $\lambda$ = 5, and 4) $p$ = 0.25 and $\lambda$ = 25. Black lines represent post-reduction $\lambda$. The violin plots are scaled to have the same maximum width.

**Table 5:** Accuracy of abundance estimates indicated by the mean difference (± SD) in estimated initial abundance ($\hat{\lambda}_0$) and simulated abundance ($\lambda$) and estimated post reduction abundance ($\hat{\lambda}_1$) and true post reduction abundance (($\lambda * (1 - R)$)) when $J$ = 3, 5, or 7, and $p$ = 0.05, 0.15, or 0.25.

| Metric | J | p | $\lambda = 5$ | $\lambda = 15$ | $\lambda = 25$ |
|---|---|---|---|---|---|
| $\hat{\lambda}_0$ | 3 | 0.05 | 21.02 ± 28.94 | 18.14 ± 28.47 | 12.93 ± 27.13 |
| $\hat{\lambda}_0$ | 3 | 0.15 | 9.17 ± 21.25 | 7.87 ± 19.56 | 7.24 ± 19.25 |
| $\hat{\lambda}_0$ | 3 | 0.25 | 3.64 ± 13.94 | 3.91 ± 13.62 | 3.77 ± 13.94 |
| $\hat{\lambda}_0$ | 5 | 0.05 | 17.25 ± 27.25 | 14.08 ± 25.31 | 11.37 ± 23.80 |
| $\hat{\lambda}_0$ | 5 | 0.15 | 4.13 ± 14.26 | 4.28 ± 14.13 | 4.34 ± 14.50 |
| $\hat{\lambda}_0$ | 5 | 0.25 | 1.30 ± 7.77 | 1.88 ± 9.34 | 2.36 ± 10.60 |
| $\hat{\lambda}_0$ | 7 | 0.05 | 12.45 ± 24.13 | 10.35 ± 21.89 | 9.42 ± 21.20 |
| $\hat{\lambda}_0$ | 7 | 0.15 | 2.18 ± 10.27 | 2.73 ± 11.20 | 3.33 ± 12.46 |
| $\hat{\lambda}_0$ | 7 | 0.25 | 0.72 ± 5.44 | 1.34 ± 7.43 | 1.82 ± 9.22 |
| $\hat{\lambda}_1$ | 3 | 0.05 | 22.53 ± 28.97 | 20.84 ± 29.06 | 18.99 ± 28.70 |
| $\hat{\lambda}_1$ | 3 | 0.15 | 10.66 ± 22.66 | 8.73 ± 20.60 | 8.24 ± 20.08 |
| $\hat{\lambda}_1$ | 3 | 0.25 | 4.33 ± 15.08 | 3.77 ± 13.70 | 3.59 ± 13.17 |
| $\hat{\lambda}_1$ | 5 | 0.05 | 18.18 ± 27.42 | 16.38 ± 26.73 | 14.62 ± 25.64 |
| $\hat{\lambda}_1$ | 5 | 0.15 | 4.88 ± 15.92 | 3.96 ± 13.83 | 4.13 ± 13.82 |
| $\hat{\lambda}_1$ | 5 | 0.25 | 1.33 ± 8.26 | 1.40 ± 7.97 | 1.73 ± 8.88 |
| $\hat{\lambda}_1$ | 7 | 0.05 | 14.17 ± 25.40 | 12.26 ± 23.79 | 11.14 ± 22.65 |
| $\hat{\lambda}_1$ | 7 | 0.15 | 2.34 ± 11.04 | 2.31 ± 10.31 | 2.69 ± 11.07 |
| $\hat{\lambda}_1$ | 7 | 0.25 | 0.54 ± 4.94 | 0.85 ± 5.84 | 1.13 ± 7.06 |

| | | J = | 3 | 3 | 3 | 5 | 5 | 5 | 7 | 7 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Metric | $\lambda$ | p = | 0.05 | 0.15 | 0.25 | 0.05 | 0.15 | 0.25 | 0.05 | 0.15 | 0.25 |
| LCI $\hat{\lambda}_0>$ | 5 | | 0.439 | 0.245 | 0.132 | 0.382 | 0.145 | 0.067 | 0.308 | 0.096 | 0.069 |
| LCI $\hat{\lambda}_0>$ | 15 | | 0.459 | 0.273 | 0.168 | 0.395 | 0.183 | 0.105 | 0.334 | 0.132 | 0.099 |
| LCI $\hat{\lambda}_0>$ | 25 | | 0.448 | 0.288 | 0.181 | 0.402 | 0.199 | 0.134 | 0.354 | 0.165 | 0.120 |
| UCI $\hat{\lambda}_0<$ | 5 | | 0.333 | 0.131 | 0.065 | 0.203 | 0.058 | 0.038 | 0.156 | 0.042 | 0.034 |
| UCI $\hat{\lambda}_0<$ | 15 | | 0.359 | 0.185 | 0.104 | 0.286 | 0.112 | 0.075 | 0.230 | 0.087 | 0.068 |
| UCI $\hat{\lambda}_0<$ | 25 | | 0.389 | 0.209 | 0.137 | 0.309 | 0.132 | 0.095 | 0.246 | 0.107 | 0.086 |
| LCI $\hat{\lambda}_1>$ | 5*(1-$R$) | | 0.453 | 0.252 | 0.122 | 0.384 | 0.137 | 0.061 | 0.309 | 0.083 | 0.072 |
| LCI $\hat{\lambda}_1>$ | 15*(1-$R$) | | 0.451 | 0.256 | 0.144 | 0.383 | 0.153 | 0.079 | 0.318 | 0.108 | 0.074 |
| LCI $\hat{\lambda}_1>$ | 25*(1-$R$) | | 0.452 | 0.269 | 0.153 | 0.390 | 0.175 | 0.098 | 0.330 | 0.132 | 0.086 |
| UCI $\hat{\lambda}_1<$ | 5*(1-$R$) | | 0.278 | 0.096 | 0.047 | 0.181 | 0.046 | 0.027 | 0.127 | 0.033 | 0.028 |
| UCI $\hat{\lambda}_1<$ | 15*(1-$R$) | | 0.325 | 0.137 | 0.077 | 0.238 | 0.076 | 0.049 | 0.184 | 0.057 | 0.048 |
| UCI $\hat{\lambda}_1<$ | 25*(1-$R$) | | 0.355 | 0.169 | 0.099 | 0.263 | 0.095 | 0.067 | 0.207 | 0.077 | 0.060 |

Relative to the occupancy models (Figure 4), precision of the *N*-mixture models was poor. Generally, precision improved with a non-depletion design and when detection probability was relatively high ($p$ = 0.25; Figure 8). However, due to the poor fit of the models, particularly when $M$ ≤ 50 and $p$ = 0.05 (Figure 6; Figure 7), precision estimates may be misleading (Figure 8).
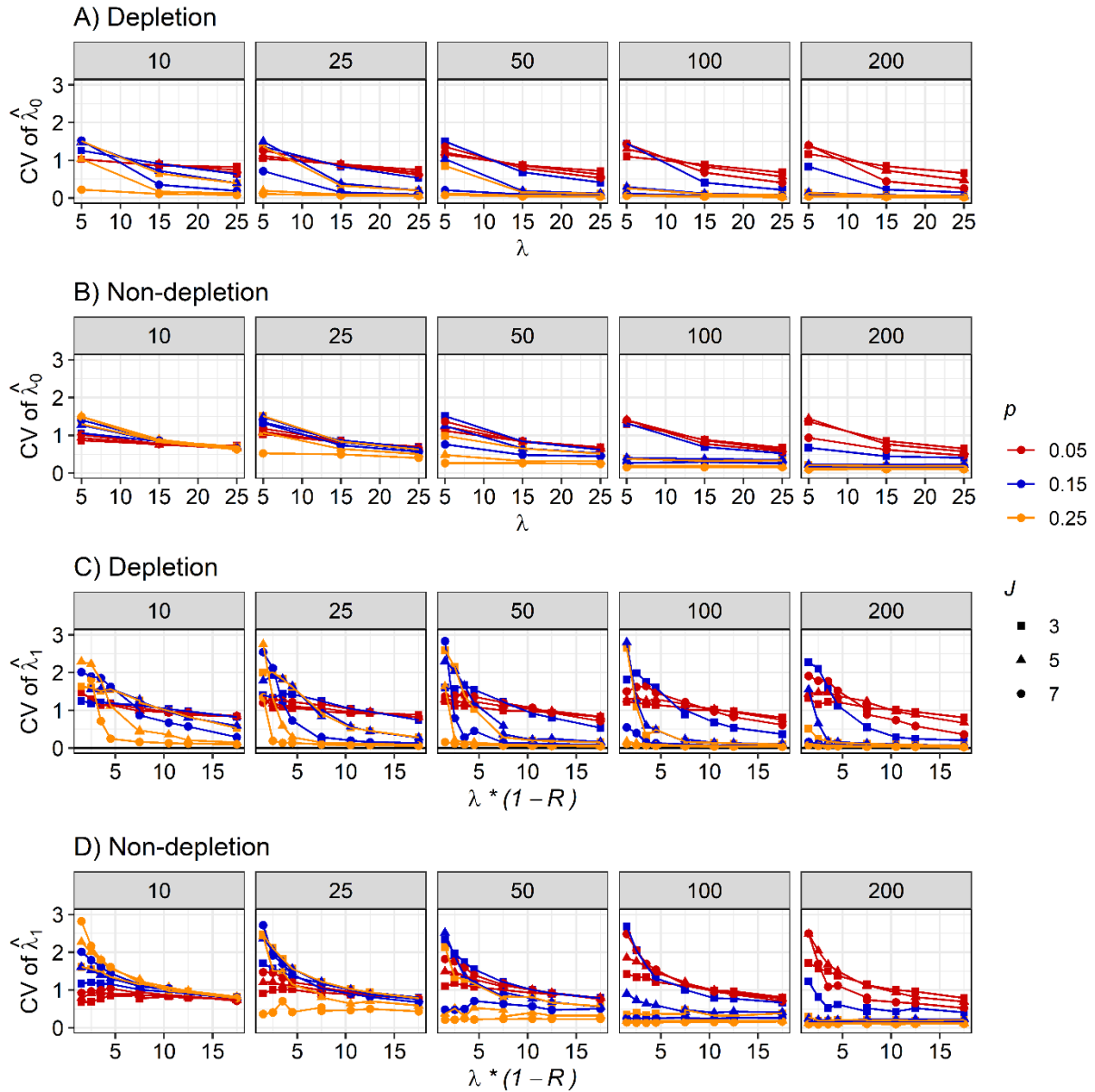
**Figure 8:** Coefficient of variation (CV) of the estimated abundance ($\hat{\lambda}$) before (A, B) and after (C, D) experiencing a proportional reduction ($R$) across detection probabilities ($p$; colour), number of surveys ($J$; shape), and number of sites (panels; 10, 25, 50, 100, 200) following a depletion (A, C) or non-depletion (B, D) design.

The power to detect changes in $\lambda$ improved with increasing $p$, $M$, $J$, $R$ and when following a depletion design (Figure 9). No sampling scenarios could detect a 30% reduction with 80% power when $\lambda$ = 5 or 15 (Figure 9). Using a depletion survey design with $J$ = 7 repeat surveys provided sufficient power (i.e., > 0.80) to detect a 70% change in $\lambda$ when individuals showed high detection probabilities ($p$ = 0.25), regardless of $\lambda$ or $M$ (Figure 7). In comparison, when using a non-depletion design with $J$ = 7, $R$ = 0.7, and $p$ = 0.25, power was poor (0.50-0.79) when $M$ = 10 and $\lambda$ = 5 or 15, and when $M$ = 25 and $\lambda$ = 5.
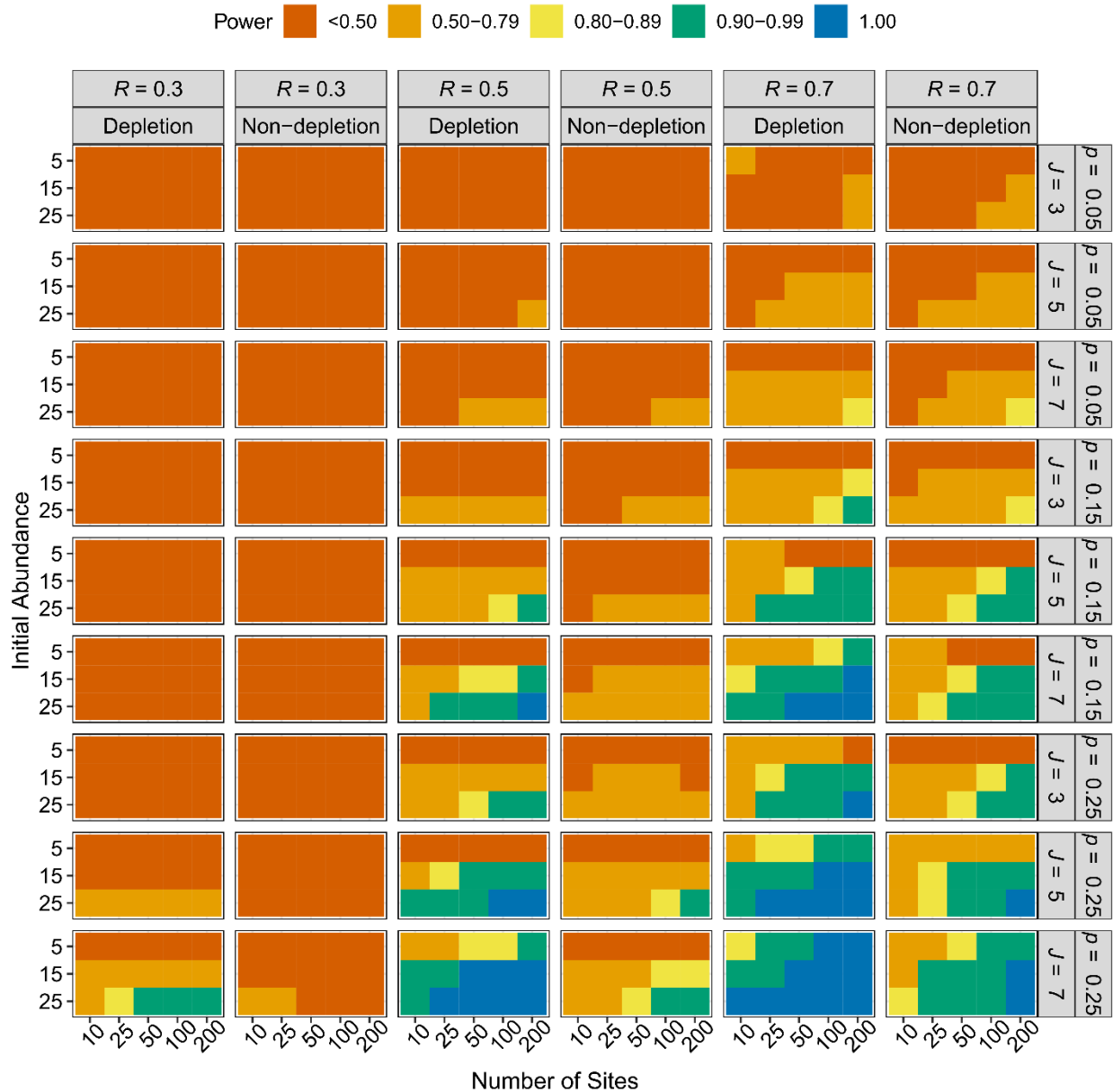
**Figure 9:** Power to detect proportional reductions ($R$ = 0.3, 0.5, 0.7) in site-level abundance when sampling was performed at differing numbers of sites ($M$ = 10, 25, 50, 100, 200) using a differing number of surveys ($J$ = 3, 5, 7). Power was calculated for detecting a change in abundance for species that demonstrate different detection probabilities (p = 0.05, 0.15, 0.25; columns) and initial abundances ($\lambda$ = 5, 15, 25) when sampling was performed using a depletion and non-depletion survey design, respectively. The interpretation of power is red = bad (< 0.50), orange = poor (0.50-0.79), yellow = good (0.80-0.89), green = great (0.90-0.99), blue = perfect (1.00).

## DISCUSSION

Monitoring is a critical aspect of species conservation in Canada as it provides trend information for species assessment by COSEWIC and, by extension, the SARA process. Benchmarks have

been established by the International Union for Conservation of Nature (IUCN) and implemented by COSEWIC for assessing the conservation status of a species based on spatial and temporal trends, commonly represented as proportional reductions in abundance, area of occupancy, extent of occurrence, and/or quality of habitat (SARA 2002; IUCN 2012; COSEWIC 2021). However, due to the issues of imperfect detection and low species abundance, describing changes in abundance or distribution with reasonable statistical power is challenging. The results of this study demonstrate that the power to detect changes in occupancy and abundance over time improve with increasing numbers of sites and surveys, detection probability, and estimated occupancy or abundance. These results align with previous studies that have identified a substantial amount of effort is required to detect SARA-listed freshwater fishes and to characterize change in occupancy over time (Reid and Hogg 2014; Reid and Dextrase 2017; Reid and Haxton 2017; Lamothe et al. 2023).

*N*-mixture models are more frequently being used to estimate species abundance but have only rarely been applied for SARA-listed freshwater fish species (Lamothe et al. 2023). The results of this study suggest that the use of independent single-season *N*-mixture models to detect changes in rare and imperfectly detected species abundance over time requires many sampled sites and repeat surveys and is limited to scenarios where species are undergoing significant declines (70% reductions or greater) and when mean site-level abundance values and detection probabilities are relatively high. The accuracy and precision of abundance estimates were often low, particularly when the number of sites and surveys and detection probabilities were low, leading to uncertainty in estimating true trends. The ability to detect changes in occupancy probability was greater than detecting changes in abundance, with estimates being more accurate and precise for detecting changes in occupancy given the same level of sampling effort; nevertheless, considerable effort was often required to detect trends in occupancy with sufficient power. Moreover, detecting a proportional change in occupancy equal in magnitude to a change in abundance may represent a more biologically significant impact to the species, depending on the nature of the decline.

When species or individuals of a species have a low detection probability, the number of sites and repeat surveys needed to achieve reasonable (i.e., ≥ 0.80) statistical power for detecting change between two surveys using independent single-season hierarchical models was typically large. For example, if following a depletion design, 100 sites need to be repeatedly surveyed five times in each time period (100 * 5 * 2 = 1000 surveys) to detect a 70% reduction in occupancy when species detection probability was relatively high ($\rho = 0.7$) and initial occupancy probability = 0.25. The required effort is even greater when making inferences on relatively small changes in occupancy or abundance between two time periods (i.e., 30% proportional reduction). The challenge of estimating occupancy probability and abundance at very low detection probabilities and sample sizes has been demonstrated previously (Guillera-Arroita et al. 2010; Lamothe et al. 2023). Estimates of occupancy tend to reach a boundary when the proportion of sites where the species was not detected is smaller than the proportion of zeros in the detection history raised to the power of *J* (i.e., number of repeat surveys; Guillera-Arroita et al. 2010). This finding suggests that boundary estimates are an issue when working on species with low detection probability and when sample sizes are small (Guillera-Arroita et al. 2010), as demonstrated in Figure 2. Similarly, biased abundance estimates for imperfectly detected individuals from *N*-mixture models can arise when the number of sites and surveys is low. For example, if two surveys are performed at 20 sites, the probability of detecting an individual would need to be greater than 0.70 to produce an unbiased estimate of abundance (Kéry and Royle 2016), much larger than the detection probability values included in the simulations used in this study.

Several empirical studies have quantified the probability of detection and occupancy of SARA-listed freshwater fishes in Ontario (e.g., Dextrase et al. 2014*a*; Dextrase et al. 2014*b*; Reid and Haxton 2017). For example, using backpack electrofishing sampling methods at 36 sites in southern Ontario rivers following a non-depletion survey design, the probability of detecting Channel Darter (*Percina copelandi*) and Northern Sunfish (*Lepomis peltastes*) after four repeat surveys was estimated as 0.69 (95% CI: 0.61-0.77) and 0.61 (95% CI: 0.47-0.75), respectively (Reid and Haxton 2017). The estimated occupancy probability for Channel Darter was 0.88 (95% CI: 0.75-0.97) and for Northern Sunfish was 0.38 (95% CI: 0.23-0.54; Reid and Haxton 2017). Power to detect future changes in occupancy probability was estimated using a closed-form estimator (Guillera-Arroita and Lahoz-Monfort 2012) and suggested that a 70% decline in future site occupancy could be detected for Channel Darter, but not Northern Sunfish (Reid and Haxton 2017). Based on the approach used to estimate power in our study, the power to detect a 30%, 50%, or 70% change in occupancy for Channel Darter given the previous estimates and level of effort is 0.60, 0.98, and 1.00, respectively. Similarly, the power to detect a 30%, 50%, or 70% change in occupancy for Northern Sunfish given the previous estimates and effort is 0.14, 0.30, and 0.60.

In practice, the sampling resources required to implement a species monitoring program for detecting change in rare, imperfectly detected species over time with statistical certainty is high when detection probability is low. The consequences of low abundance and detection probabilities emphasize the importance of using the most effective sampling procedures and gear. Improving detection probability is the best solution for reducing the effort necessary for quantifying proportional reductions in occupancy or abundance over time (McKann et al. 2013). Pilot studies can help to determine the best gear for sampling species to maximize detection probability, while also providing a baseline expectation for gear-specific detection probabilities and informing how sampling procedures can be improved. For example, previous studies have demonstrated that increasing the size of camera arrays and allowing a greater duration of camera trap sampling can improve detection probability (e.g., of terrestrial mammals, including relatively rare species such as Bobcat (*Lynx rufus*); O'Connor et al. 2017; and Redside Dace, a SARA-listed freshwater fish (Castañeda et al. 2020*a*). In other cases, approaches involving multiple types of sampling gear can be used to increase detection probability and/or avoid captures of non-target species (Harkins et al. 2019). For example, environmental DNA techniques are improving the ability of researchers to detect rare aquatic species (Schmelzle and Kinziger 2016; Boothroyd et al. 2016; Stickland and Roberts 2019) and, combined with traditional sampling, may improve detection probability. Ultimately, an understanding of the efficiency of surveys at detecting target species is a critically important consideration when designing a species monitoring program.

Alternative survey designs to the ones used in this study may provide a less resource-intensive approach to sampling rare species (MacKenzie et al. 2005; Specht et al. 2017). For example, rather than a traditional non-depletion sampling design, which was shown to have greater power for detecting change in occupancy than a depletion design, repeat samples can be taken conditional on a positive detection in the first survey (i.e., conditional sampling; Specht et al. 2017). If the species was not detected in the first survey, field crews move to the next site after measuring any relevant habitat covariates. When a detection occurs in the first survey, the number of surveys at that site is dependent on prior expectations of detection and occupancy probabilities. In addition to improving estimates of occupancy and detection probability for rare species, conditional sampling also allows field crews to monitor a greater number of sites, potentially improving knowledge of species distribution (Specht et al. 2017).

Power analysis is an important step when developing a monitoring program, but like any analysis, it has several underlying assumptions. In this study, we assumed independence

between sites within and across sampling periods for both the presence-absence and abundance simulations (as per Guillera-Arroita and Lahoz-Monfort 2012). Meeting this assumption requires a completely random survey design across sites, which is often difficult to implement for rare, imperfectly detected species. A random survey design for rare species can also cause zero-inflation in the data, or no detections at all, and subsequently results in lower estimated detection probability, further inflating the required number of sites for detecting change in species abundance and occupancy. Instead, stratified random sampling across potentially suitable habitat types may be more effective than a completely random design. A second assumption of this analysis was that proportional reductions in occupancy and abundance were applied uniformly across the entire sampling frame. Reductions in species abundance and distribution are more likely to occur in particular areas of the distribution, namely range edges or areas of perturbation, and therefore may be missed or amplified depending on the sampling frame. Third, the relationship between proportional reductions in occupancy and abundance is unknown, and likely not generalizable across species. For example, the occupancy probability of a schooling fish may remain relatively constant until the school becomes too small to offer protection from predators. Finally, the simulations in this study assumed a constant detection probability between surveys and through time. A variety of scenarios could lead to a violation of this assumption that may differ between depletion and non-depletion survey designs. For example, fishes may avoid sampling gear after being captured, leading to reduced detection probabilities in subsequent surveys following a non-depletion design. Alternatively, the disruption of habitat during an initial survey may lead to increases or decreases in the detection probability of species during subsequent surveys when performing depletion surveys. Previous simulation studies that tested the assumptions of *N*-mixture models demonstrated that survey-specific heterogeneity in detection probability can result in an overestimation of abundance and an underestimation of detection probability (Kéry and Royle 2016; Link et al. 2018). Given the assumptions of hierarchical models, study-specific simulations that encompass the anticipated survey design and estimated parameters are recommended prior to initiating monitoring.

Dynamic occupancy and *N*-mixture models may provide a more applicable study design when estimating trends in species status compared with the static models used in this study. Rather than applying static models for each individual year, dynamic models estimate the probability of colonization and local extirpation to facilitate an understanding of change through time (MacKenzie et al. 2003; Dail and Madsen 2011). Simulations to explore biases in model output for small sample sizes and power analyses for dynamic occupancy and *N*-mixture models have shown promising results (McKann et al. 2013; Ficetola et al. 2018; Banner et al. 2019). For example, a power of at least 0.80 can be achieved for detecting a 30% trend in abundance for a species with an initial mean abundance of 15 and a detection probability of 0.05 using seven repeat surveys at 30 sites (Ficetola et al. 2018), which is far less effort than that shown in this study for stacked single-season models. In addition to increased power, factors hypothesized to influence the colonization or extirpation processes can be incorporated into dynamic occupancy models to provide a more causal understanding of population and/or species dynamics (Wheeler et al. 2018; Castañeda et al. 2020*b*). For example, the colonization and local extirpation rates of three imperiled fish species, Flannelmouth Sucker (*Catostomus latipinnis*), Bluehead Sucker (*C. discobolus*), and Roundtail Chub (*Gila robusta*), were linked to peak annual flows in the riverine systems of Utah, United States (Budy et al. 2015). As peak discharge increased, local extirpation probability of these imperiled species decreased, and colonization probability increased (Budy et al. 2015).

As the number of SARA-listed species continues to increase, monitoring efforts are more frequently being considered from a multi-species perspective, aiming to document trends in the occurrence and abundance of as many imperiled species as possible. However, this approach

can reduce the number and frequency of sites sampled for any single species given the need to allocate effort across species and the potential need for deploying multiple sampling gear types per site. As a result, the power to detect statistically significant trends in occupancy or abundance for any individual species may be reduced. Therefore, future work to quantify the sampling time, resource requirements, and power to detect individual species within a multi-species monitoring program is warranted.

## ACKNOWLEDGEMENTS

## REFERENCES

Banner, K.M., Irvine, K.M., Rodhouse, T.J., Donner, D., and Litt, A.R. 2019. Statistical power of dynamic occupancy models to identify change: informing the North American Bat Monitoring Program. Ecol. Indic. **105**: 166-176. doi: 10.1016/j.ecolind.2019.05.047.

Barnucz, J., and Drake, D.A.R. 2021. Mark-recapture sampling for lake chubsucker (*Erimyzon sucetta*) and grass pickerel (*Esox americanus vermiculatus*) in L-Lake, Ontario, 2018. Can. Data Rep. Fish. Aquat. Sci. 1347: vii + 28 p.

Barnucz, J., Gáspárdy, R.C., Smith, K., and Drake, D.A.R. 2021. Fish community inventory and mark-recapture sampling of SARA-listed fishes in Point Pelee National Park, Ontario, 2019. Can. Data Rep. Fish. Aquat. Sci. 1323: vii + 72 p.

Boothroyd, M., Mandrak, N.E., Fox, M., and Wilson, C.C. 2016. Environmental DNA (eDNA) detection and habitat occupancy of threatened spotted gar (*Lepisosteus oculatus*). Aquat. Conserv.: Mar. Freshw. Ecosyst. **26**(6): 1107-1119. doi: 10.1002/aqc.2617.

Budy, P., Conner, M.M., Salant, N.L., and Macfarlane, W.W. 2015. An occupancy-based quantification of the highly imperiled status of desert fishes of the southwestern United States. Conserv. Biol. **29**(4): 1142-1152. doi: 10.1111/cobi.12513.

Castañeda, R.A., Weyl, O.L.F., and Mandrak, N.E. 2020a. Using occupancy models to assess the effectiveness of underwater cameras to detect rare stream fishes. Aquat. Conserv.: Mar. Freshw. Ecosyst. **30**(3): 565-576. doi: 10.1002/aqc.3254.

Castañeda, R.A., Mandrak, N.E., Barrow, S., and Weyl, O.L.F. 2020b. Occupancy dynamics of rare cyprinids after invasive fish eradication. Aquat. Conserv.: Mar. Freshw. Ecosyst. **30**(7): 1424-1436. doi: 10.1002/aqc.3364.

Cohen, B. H. 2001. Explaining psychological statistics (2nd ed.). John Wiley & Sons, Inc., New York, NY.

COSEWIC [Committee on the Status of Endangered Wildlife in Canada]. 2021. Table 2. COSEWIC quantitative criteria and guidelines for the status assessment of Wildlife Species. Accessed 2022-07-26. Available at: https://cosewic.ca/index.php/en-ca/assessment-process/wildlife-species-assessment-process-categories-guidelines/quantitative-criteria.html.

Dail, D., and Madsen, L. 2011. Models for estimating abundance from repeated counts of an open metapopulation. Biometrics **67**(2): 577-587. doi: 10.1111/j.1541-0420.2010.01465.x.

Dextrase, A.J., Mandrak, N.E., and Schaefer, J.A. 2014a. Modelling occupancy of an imperilled stream fish at multiple scales while accounting for imperfect detection: implications for conservation. Freshw. Biol. **59**(9): 1799-1815. doi: 10.1111/fwb.12384.

Dextrase, A.J., Mandrak, N.E., Barnucz, J., Bouvier, L., Gaspardy, R., and Reid, S.M. 2014b. Sampling effort required to detect fishes at risk in Ontario. Can. Manuscr. Rep. Fish. Aquat. Sci. 3024: v + 50 p.

Ficetola, G.F., Romano, A., Salvidio, S., and Sindaco, R. 2018. Optimizing monitoring schemes to detect trends in abundance over broad scales. Anim. Conserv. **21**(3): 221-231. doi: 10.1111/acv.12356.

Fiske, I., and R. Chandler. 2011. unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. J. Stat. Softw. **43**(10): 1-23. doi: 10.18637/jss.v043.i10.

Guillera-Arroita, G., and Lahoz-Monfort, J.J. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. Methods Ecol. Evol. **3**(5): 860-869. doi: 10.1111/j.2041-210X.2012.00225.x.

Guillera-Arroita, G., Ridout, M.S., and Morgan, B.J.T. 2010. Design of occupancy studies with imperfect detection. Methods Ecol. Evol. **1**(2): 131-139. doi: 10.1111/j.2041-210X.2010.00017.x.

Harkins, K.M., Keinath, D., and Ben-David, M. 2019. It's a trap: optimizing detection of rare small mammals. PLoS One **14**(3): e0213201. doi: 10.1371/journal.pone.0213201.

IUCN. 2012. IUCN Red List Categories and Criteria: Version 3.1 (2nd ed). Gland, Switzerland and Cambridge, UK. iv + 32 pp.

Kéry, M., and Royle, J.A. 2016. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Volume 1 prelude and static models. Elsevier Academic Press, Boston, USA.

Kéry, M., J. A. Royle, and H. Schmid. 2005. Modeling avian abundance from replicated counts using binomial mixture models. Ecol. Appl. **15**(4): 1450-1461. doi: 10.1890/04-1120.

Lamothe, K.A., Dextrase, A.J., and Drake, D.A.R. 2019. Characterizing species co-occurrence patterns of imperfectly detected fishes to inform species reintroduction efforts. Conserv. Biol. **33**(6): 1392-1403. doi: 10.1111/cobi.13320.

Lamothe, K.A., and Drake, D.A.R. 2022. Stage-specific abiotic and biotic associations of the imperilled silver shiner *Notropis photogenis* in an urban drainage. Ecol. Freshw. Fish **31**(3): 571-582. doi: 10.1111/eff.12652.

Lamothe, K.A., Reid, S.M., and Drake, D.A.R. 2023. Considerations around effort and power for an occupancy-based Redside Dace *Clinostomus elongatus* monitoring program across spatial and temporal scales. DFO Can. Sci. Advis. Sec. Res. Doc. 2023/034. iv + 37 p.

Link, W.A., Schofield, M.R., Barker, R.J., and Sauer, J.R. 2018. On the robustness of N-mixture models. Ecology **99**(7): 1547-1551. doi: 10.1002/ecy.2362.

McKann, P.C., Bray, B.R., and Thogmartin, W.E. 2013. Small sample bias in dynamic occupancy models. J. Wildl. Manag. **77**(1): 172-180. doi: 10.1002/jwmg.433.

MacKenzie, D.I., and Royle J.A.. 2005. Designing occupancy studies: general advice and allocating survey effort. J. Appl. Ecol. **42**(6): 1105-1114. doi: 10.1111/j.1365-2664.2005.01098.x.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., and Langtimm, C.A.. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology **83**(8): 2248-2255. doi: 10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G., and Franklin, A.B. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology **84**(8): 2200-2207. doi: 10.1890/02-3090.

MacKenzie, D.I., Nichols, J.D., Sutton, N., Kawanishi, K., and Bailey, L.L. 2005. Improving inferences in population studies of rare species that are detected imperfectly. Ecology **86**(5): 1101-1113. doi: 10.1890/04-1060.

MacKenzie, D. I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., and Hines, J.E. 2018. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence (2nd ed). Elsevier Academic Press, Burlington, MA.

O'Connor, K.M., Nathan, L.R., Liberati, M.R., Tingley, M.W., Vokoun, J.C., and Rittenhouse, T.A.G. 2017. Camera trap arrays improve detection probability of wildlife: investigating study design considerations using an empirical dataset. PLoS One **12**(4): e0175684. doi: 10.1371/journal.pone.0175684.

Peterman, R.M. 1990. Statistical power can improve fisheries research and management. Can. J. Fish. Aquat. Sci. **47**(1): 2-15. doi: 10.1139/f90-001.

R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reid, S.M., and Dextrase, A.J. 2017. Monitoring lake populations of Eastern Sand Darter (*Ammocrypta pellucida*): a comparison of two seines. J. Freshw. Ecol. **32**(1): 499-511. doi: 10.1080/02705060.2017.1344734.

Reid, S.M., and Haxton, T. 2017. Backpack electrofishing effort and imperfect detection: influence on riverine fish inventories and monitoring. J. Appl. Ichthyol. **33**(6): 1083-1094. doi: 0.1111/jai.13463.

Reid, S.M., and Hogg, S. 2014. An evaluation of multiple-pass seining to monitor Blackstripe Topminnow populations in the Sydenham River (Ontario, Canada). J. Appl. Ichthyol. **30**(5): 962-969. doi: 10.1111/jai.12447.

Rosenberger, A.E., and Dunham, J.B. 2005. Validation of abundance estimates from mark-recapture and removal techniques for Rainbow Trout captured by electrofishing in small streams. N. Am. J. Fish. Manag. **25**(4): 1395-1410. doi: 10.1577/M04-081.1.

Royle, J.A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. Biometrics **60**(1): 108-115. doi: 10.1111/j.0006-341X.2004.00142.x.

Schmelzle, M.C., and Kinziger, A.P. 2016. Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. Mol. Ecol. Resour. **16**(4): 895-908. doi: 10.1111/1755-0998.12501.

Specht, H.M., Reich, H.T., Iannarilli, F., Edwards, M.R., Stapleton, S.P., Weegman, M.D., Johnson, M.K., Yohannes, B.J., and Arnold, T.W. 2017. Occupancy surveys with

conditional replicates: An alternative sampling design for rare species. Methods Ecol Evol. **8**(12): 1725-1734. doi: 10.1111/2041-210X.12842.

Species at Risk Act [SARA]. 2002. Species at Risk Act. S.C. 2002, c. 29. Accessed 2022-07-26. Available at: https://laws.justice.gc.ca/eng/acts/S-15.3/FullText.html.

Strickland, G.J., and Roberts, J.H. 2019. Utility of eDNA and occupancy models for monitoring an endangered fish across diverse riverine habitats. Hydrobiologia **826**: 129-144. doi: 10.1007/s10750-018-3723-8.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., and Possingham, H.P. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecol. Appl. **13**(6): 1790-1801. doi: 10.1890/02-5078.

Wheeler, K., Wenger, S.J., Walsh, S.J., Martin, Z.P., Jelks, H.L., and Freeman, M.C. 2018. Stream fish colonization but not persistence varies regionally across a large North American river basin. Biol. Conserv. **223**: 1-10. doi: 10.1016/j.biocon.2018.04.023.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.

Zipkin, E.F., Sillett, T.S., Campbell Grant, E.H., Chandler, R.B., and Royle, J.A. 2014. Inferences about population dynamics from count data using multistate models: a comparison to capture-recapture approaches. Ecol. Evol. **4**(4): 417-426. doi: 10.1002/ece3.942.