

LETTER

Data-driven counterfactual evaluation of management outcomes to improve emergency conservation decisions

Thalassa McMurdo Hamilton^{1,2,3} | John G. Ewen² | Antony J. Beauchamp⁴ |
 Troy Makan⁵ | Marcus Rowcliffe^{2,3} | Stefano Canessa^{2,6}

¹Biodiversify Ltd, Newark, UK

²Institute of Zoology, Zoological Society of London, London, UK

³Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK

⁴TSO Biodiversity Threats, Northland Conservancy, Department of Conservation, Northland District Office, Whangarei, New Zealand

⁵Department of Conservation, Terrestrial Science Unit, Biodiversity Group, Rotorua, New Zealand

⁶Division of Conservation Biology, Institute for Ecology and Evolution, University of Bern, Bern, Switzerland

Correspondence

Thalassa McMurdo Hamilton,
 Biodiversify Ltd, Newark NG23 6JE, UK.
 Email:
Thalassa.McMurdoHamilton@zsl.org

Funding information

Natural Environment Research Council,
 Grant/Award Number: NE/L002485/1

Abstract

Monitoring is needed to assess conservation success and improve management, but naïve or simplistic interpretation of monitoring data can lead to poor decisions. We illustrate how to counter this risk by combining decision-support tools and quantitative counterfactual analysis. We analyzed 20 years of egg rescue for tara iti (*Sternula nereis davisae*) in Aotearoa New Zealand. Survival is lower for rescued eggs; however, only eggs perceived as imminently threatened by predators or weather are rescued, so concluding that rescue is ineffective would be biased. Equally, simply assuming all rescued eggs would have died if left in situ is likely to be simplistic. Instead, we used the monitoring data itself to estimate statistical support for a wide space of uncertain counterfactuals about decisions and fate of rescued eggs. Results suggest under past management, rescuing and leaving eggs would have led to approximately the same overall fledging rate, because of likely imperfect threat assessment and low survival of rescued eggs to fledging. Managers are currently working to improve both parameters. Our approach avoids both naïve interpretation of observed outcomes and simplistic assumptions that management is always justified, using the same data to obtain unbiased quantitative estimates of counterfactual support.

KEYWORDS

confusion matrix, decision making, decision tree, endangered species, hindsight bias, impact evaluation, nest management, uncertainty

1 | INTRODUCTION

Ecosystem restoration and species recovery often require decades of close-order management (Young et al., 2014). During that time, managers need to assess and adjust actions in response to improved knowledge and changing conditions. To know what is working and what must be changed, monitoring data are usually accumulated,

often at great expense and effort (Buxton et al. 2020). However, monitoring data may be harder to interpret than managers anticipate: many programs make incorrect inferences from those data (Christie et al., 2019) or fail to use them at all, perpetuating management that does not achieve its objectives and makes poor use of limited resources (Lindenmayer et al., 2013; Buxton et al., 2020).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. Conservation Letters published by Wiley Periodicals LLC.

Management also often evolves ad hoc, because expectations and resources change, and speedy intervention might initially be prioritized over thorough decision-making. For example, critically endangered species often require intensive mitigation of all possible threats like predation, human disturbance, or poor weather, combining multiple actions without perfect knowledge of the key threats (Jones et al., 2018). Managers might then be reluctant to change ongoing actions or to implement robust experimental approaches including controls, possibly because of risk aversion or status quo bias (Canessa et al., 2019; Runge, 2011). Moreover, conservation decisions rely heavily on expert judgement (Martínez-Abraín & Oro, 2013; Walsh et al., 2019), often with informal examination of outcomes from small samples.

To improve evaluations, conservation increasingly adopts counterfactuals to assess outcomes where controls are not available (Ferraro, 2009; Bull et al., 2021; Jellesmark et al., 2021). A counterfactual is an “if-then” hypothetical statement about what would have happened if a different action had been taken; for example, “if its habitat had not been protected, then the species would have declined further instead of recovering as it did.” Counterfactuals are a key tool to understand causality and can help reduce bias. However, if not built carefully, counterfactuals can enable “just-so” storytelling, where one implicitly justifies their original choice, assuming the counterfactual outcome would have been just as initially expected, without truly critically evaluating the hypothetical alternative (Roese & Olson, 1996). This can happen if the focus is too narrow (too few counterfactual scenarios that do not represent true uncertainty) or if the evaluation relies on subjective judgement without trying to reduce bias. These challenges are especially relevant in conservation, where data are scarce and messy, feedback slow, and causality difficult to infer, so multiple counterfactuals remain rare (Bull et al., 2021). Some level of subjective expert judgement is inevitably required (Baylis et al., 2016; Bull et al., 2021; Coetzee & Gaston, 2021), for example, when assessing whether a given action has prevented extinction (Hoffmann et al., 2015).

In summary, counterfactuals can significantly improve conservation decisions, but ideally, they should explore a range of scenarios and be assessed based on empirical evidence. In this study, we illustrate how to carry out such a broad, evidence-based counterfactual evaluation using an extreme example of a critically endangered bird, tara iti (New Zealand fairy tern *Sternula nereis davisae*). Rescue of threatened tara iti eggs is a key conservation action, but in the absence of control, evaluation needs a counterfactual approach across a wide range of scenarios. We combine a decision tree and a classification matrix to quantitatively evaluate past management and to identify

possible improvements across a wide range of uncertain counterfactual scenarios for tara iti.

2 | METHODS

2.1 | Study system and decision context

Tara iti is the rarest breeding tern in Aotearoa New Zealand, listed as “Nationally Critical” (Robertson et al., 2021), with a population of approximately 35–40 adult birds in 2021 (Department of Conservation, unpublished data; McMurdo Hamilton et al., 2021). It currently breeds at four sites close to estuarine and oceanic waters, on open sandy areas with little vegetation and patches of shell. Tara iti lay one to two eggs in a “scrape” on the sand, primarily on low-lying shell patches (Treadgold, 2000). During the breeding season, tara iti are vulnerable to predation of eggs, chicks, or adults by invasive mammalian predators and avian predators, nest disturbance or destruction by humans, and nest failure due to chilling, sanding, or tidal flooding of eggs, often in combination with poor weather (Treadgold, 2000).

Since 1997, all tara iti eggs have been monitored closely and, whenever threats are perceived, eggs are rescued to prevent nest failure. Whilst fledging rates improved during 1991–2002, it is unknown whether this was attributable to egg management (Ferreira et al., 2005). Tara iti management and monitoring data collection are coordinated by Aotearoa—New Zealand’s Department of Conservation (DOC) with support from DOC rangers and scientific staff, community volunteers, and Auckland Zoo.

Eggs are subject to two types of management: habitat management and egg rescue. Habitat management aims to maximize safe nesting locations and limit human disturbance with fences, ranger activities, creating shell patches, as well as lethally controlling native and non-native predators. Occasionally, if a nest is at risk from flooding events, sandbag walls and trenches are used in situ to stem the tide (Ferreira et al., 2005; Hansen, 2006). All eggs in this study were managed by such habitat improvement and predator control, which we do not assess here.

Direct egg rescue, the focus of this study, aims to maximize reproductive success when a threat to a nest—such as a nearby predator or flooding tide—is detected, despite habitat management and predator control. Managers considering such direct management have four alternative actions available: (1) leave eggs in situ, (2) shift nests with fertile eggs away from tidal inundation, (3) transfer imperiled eggs directly to foster pairs, and (4) bring eggs into temporary captivity (at Auckland Zoo) until the threat has passed and return them to either natal parents or foster pairs (details in Table 1). The assumption is that, when an

TABLE 1 Description of alternative egg rescue actions available to managers when a threat has been detected that is insurmountable by other forms of management

| Action | Description |
|---|---|
| None | No management (nest left as it is) |
| Donate | Eggs removed and either directly transferred to foster parents or via Auckland Zoo first (for ≤ 1 day) |
| Artificial incubation with donation | Eggs removed and incubated at Auckland Zoo (for > 1 day); natal parents abandon. Eggs donated to foster parents later |
| Artificial incubation with natal return | Eggs removed and incubated, in situ or at Auckland Zoo (> 1 day); dummy eggs put in natal nest. Eggs later returned to parents |
| Shift | Nest is gradually shifted away from incoming tide or sand over a period of a few hours |

egg is perceived under imminent threat, implementing one of actions 2–4 increases its chance of fledging compared to leaving it in situ. This approach and its suite of management actions has remained relatively unchanged since 1997. Hereafter, for brevity we refer to “rescued eggs” and “eggs left in situ” to indicate, respectively, eggs that were or were not manipulated in response to an immediate risk (action 1 above).

2.2 | Decision analysis

We investigated the effectiveness of egg rescue by summarizing the survival of eggs rescued and left in situ, with a focus on those lost to threats that the program controls for (storms, tides, and predators). We visualized the decision to rescue or leave eggs, and the sequence of events that follows, using a decision tree (Figure 1; Behn, 1982). On each branch, uncertainty nodes represent possible egg or chick fates after a decision: survive (egg hatch or chick fledge), die due to threats preventable by rescue (predator, tide, or storm), and die due to other/unknown causes. The artificial incubation branch includes an additional uncertainty node, because some eggs are dead on arrival to captivity or die in captivity, and a subsequent decision node to choose between returning surviving eggs to natal parents or donating them to foster parents. For all branches, the ultimate outcome is whether an egg survives to fledging stage.

To estimate the overall fledging probability along each branch, we estimated probabilities for each uncertainty node. We analyzed DOC’s breeding data from 1997 to 2017 ($n = 505$ eggs, of which $n = 299$ were fertile at candling), and categorized egg rescue types (Table 1), outcomes, and causes of failure (Table 2). We estimated multinomial probabilities in a Bayesian framework using JAGS (Plummer, 2003) through the “jagsUI” package in R (Kellner, 2015). We obtained 10,000 posterior samples for each probability using an uninformative Dirichlet prior (code in Supporting Information). We entered the estimated probabilities in the corresponding tree nodes and

calculated each branch’s probability of success (fledging) by multiplying the estimated probabilities. For example, the probability of an egg fledging after being donated to another pair was calculated as $p(\text{success}|\text{donate}) = (p(\text{survive zoo}))(p(\text{hatch}|\text{donate}))(p(\text{fledge}|\text{donate}))$. To propagate uncertainty, we repeated calculations for each set of posterior samples, obtaining a full probability distribution for fledging probability for each branch. However, we knew these results are likely biased because eggs are rescued in response to a perceived threat, not randomly, so we interpreted them further using counterfactuals.

2.3 | Counterfactuals to evaluate and improve outcomes

Tara iti eggs are rescued or left in situ depending on whether an immediate threat is perceived. There are two uncertain underlying parameters: the true accuracy of classification a , that is, the proportion of eggs whose fate at fledging was accurately predicted by managers, and the counterfactual rate of egg failure f , that is, the proportion of eggs that would have failed had they been left in situ. We estimated these two parameters using a confusion matrix (Equation 1):

| | Predicted failure (managed) | Predicted success (unmanaged) | Total |
|-------------------|--------------------------------|----------------------------------|-------|
| True failure TP | | FN | N_t |
| True success FP | FN | TN | N_s |
| Total | N_m | N_u | N |

(1)

The matrix contains three observed outcomes that can be predicted from parameters a and f : TN , true negatives (eggs left in situ that survived), calculated as $N(1 - f)a$; FN , false negatives (eggs left in situ that failed), calculated as $Nf(1 - a)$; and N_m , the total number of rescued eggs, calculated as $N[fa + (1 - f)(1 - a)]$. The total number of eggs N is thus equal to $N = TN + FN + N_m$. We implemented the equations above in JAGS to obtain posterior distributions

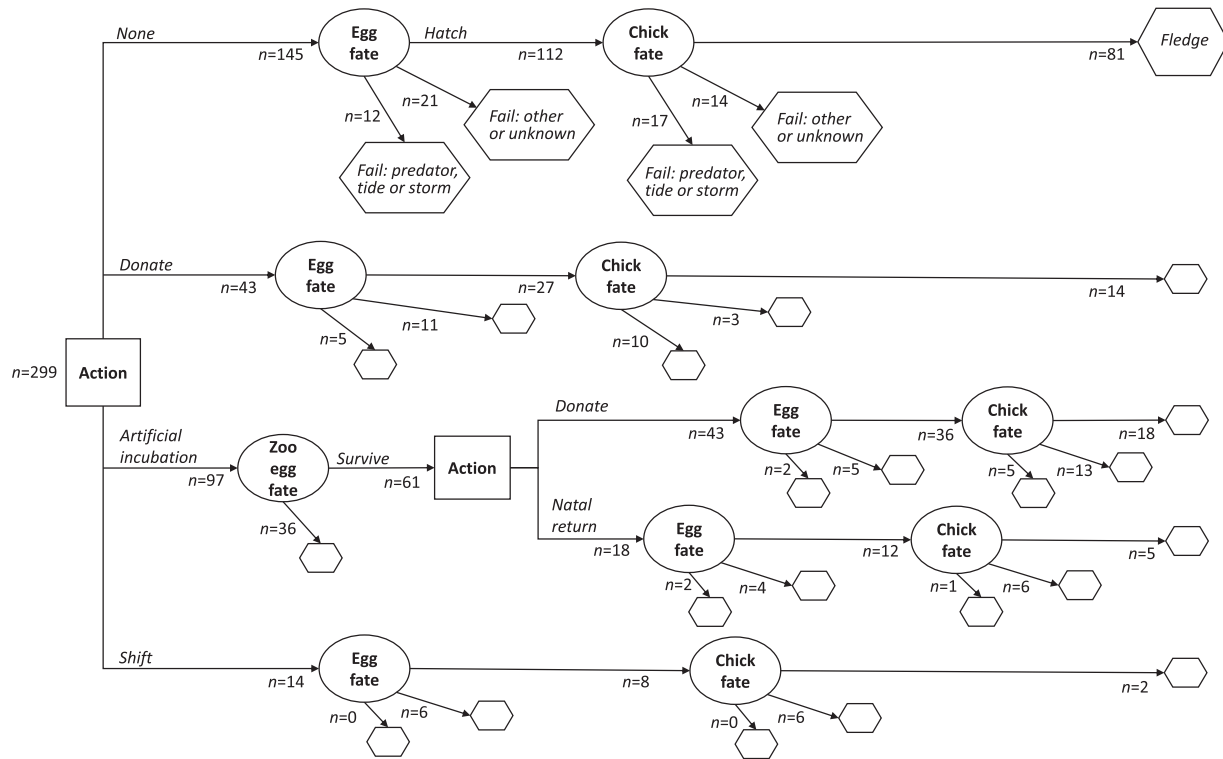


FIGURE 1 Decision tree showing the options for tara iti egg management after a threat has been detected. Each branch represents an egg management alternative and the temporal sequence of events from left to right. Squares represent decision points, ovals represent uncertainty nodes, and hexagons represent outcome in terms of hatch or fledge probability. Outcomes described on first branch apply to all branches but are not repeated across other branches to ease visualization. The numbers on each branch indicate the corresponding number of eggs in the dataset.

TABLE 2 Description of apparent causes of egg or chick failure, as recorded upon occurrence and categorized using the notes in the historical tara iti management dataset

| Outcome | Apparent cause | Description |
|----------------|--|--|
| Abandon | Embryo death, tide/storm, one parent, human disturbance, unknown | Parents have abandoned the eggs, cause inferred. Embryo death is discovered at time of postmortem. One parent indicates one of the pair left or died, leaving the other who also subsequently abandoned. |
| No hatch (egg) | Embryo death, tide/storm, damage, unknown | Like above, but field notes do not make it clear whether the pair abandoned, did not incubate enough, or field staff removed egg. |
| Removed (egg) | Infertile, embryo death, tide/storm, unknown | Eggs were removed either because the eggs were known infertile, parents had stopped incubating, or eggs had gone over date they could feasibly hatch, releasing the pair to lay again. |
| Predated | Invasive mammalian, aerial, or unknown predator, human removal | Mammalian predators are cats and rats, aerial predators are native Australasian harrier and native gull species, red-billed gull, and Southern black-backed gull, or recorded as unknown. |
| Die hatch | Tide/storm, unknown | Chick attempted to hatch but died. |
| Chick death | Health, human, tide/storm, unknown | Chick hatched but then died due to any suspected health reason, human trampling, or disturbance, or because they were starved/chilled in storms, or for unknown reasons. |

of a and f (Appendix S2). Because a and f are not identifiable, we expected a multimodal posterior distribution. Therefore, we repeated the analysis for three sets of pri-

ors for a and f : (1) fully uninformative priors for both, that is, $U(0,1)$, (2) semi-informative priors assuming managers would be better than random at detecting threats, that is,

$U(0,0.5)$, and (3) semi-informative priors assuming managers are better than random, that is, $U(0.5,1)$. In Appendix S1, we provide an alternative, optimization-based solution.

We then calculated the benefit of rescue as Equation (2):

$$\Delta_{mgt} = (TN + N_{mp}) - N(1 - f), \quad (2)$$

that is, the difference between (1) the expected number of successful eggs for each combination of parameters, as the sum of eggs left in situ that were correctly classified and succeeded (TN) and the total number of rescued eggs (N_m) multiplied by the mean success of rescued eggs estimated from the data (fledging probability p), and (2) the expected number of successful eggs if all were left in situ given the failure rate f . If $\Delta_{mgt} > 0$, rescuing eggs is better than leaving them in situ. Mapping Δ_{mgt} for all combinations of a and f , we effectively explored the entire possible counterfactual space, rather than building and assessing a single counterfactual. We used this space to assess the sensitivity of Δ_{mgt} to possible improvements in a (more accurately classifying eggs at risk of failure) and s (increasing the success rate of rescued eggs).

3 | RESULTS

3.1 | Effectiveness of egg rescue

Of 299 fertile eggs, 154 (51.5%) were rescued in response to a perceived threat; of these, 43 (27.9%) were donated, 97 (63%) artificially incubated, and 14 (9.1%) shifted. Of 145 eggs left in situ (48.5% of fertile eggs), 12 (8.3%) failed at egg stage due to threats preventable by rescue (predators, 2.8%; tides and storms, 5.5%). Eventually 81 chicks fledged from 145 eggs left in situ (56%), while 39 hatched from 154 rescued eggs (25.3%), including 14 from donated eggs (32.6% of 43), 23 from artificially incubated eggs (23.7% of 97), and two from shifted eggs (14.3% of 14). Figure 1 summarizes all survival data.

Estimating fledging from raw monitoring data, eggs left in situ had the highest fledging probability (mean $p = 0.55$, 95% confidence interval: 0.47–0.63; Figure 2). Overall, the mean fledging probability for rescued eggs was $p = 0.25$ (0.19–0.33). Donating rescued eggs to a foster pair had the highest fledging success ($p = 0.32$, 0.19–0.46), followed by artificial incubation and donation ($p = 0.26$, 0.16–0.36), artificial incubation and natal return ($p = 0.17$, 0.06–0.31), and shifting ($p = 0.14$, 0.02–0.35). Therefore, ignoring the lack of control in the data, leaving eggs in situ would be the best decision in 99.6% of decision tree solutions.

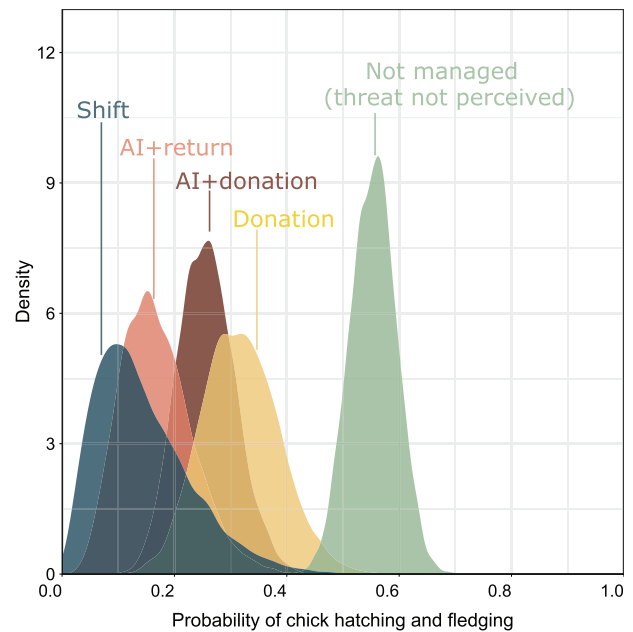


FIGURE 2 Probability distributions of successful fledging under different egg management alternatives, each calculated from 10,000 simulations of the decision tree. This is a naïve analysis of the biased dataset: as all nests that are judged to be under threat are subsequently managed, unmanaged eggs (green curve) do not represent a true control group. They only indicate the probability of fledging from nests that were not rescued in response to imminent threats by tides, weather events, or predators.

3.2 | Counterfactuals to evaluate and improve outcomes

When looking at the counterfactual space with uninformative priors, posterior distributions were clearly bimodal, with two combinations of a and f receiving the greatest support (Figures 3 and 4; mean $a = 0.53$, 95% credible interval: 0.37, 0.69; $f = 0.47$, 95% CI: 0.31, 0.63; $\Delta_{mgt} = -37.7$, 95% CI: $-87.4, 11.2$). When the prior for a was constrained to $U(0, 0.5)$, the posterior estimates were $a = 0.44$ (0.35, 0.50), $f = 0.38$ (0.30, 0.48), and $\Delta_{mgt} = -64.3$ ($-92.6, -34.8$). When the prior was constrained to $U(0.5, 1)$, the posteriors were $a = 0.62$ (0.53, 0.70), $f = 0.57$ (0.50, 0.65), and $\Delta_{mgt} = -9.8$ ($-32.2, 16.7$). If we assume managers are better than random at detecting threats, rescuing eggs under these values is likely to have led to fewer fledglings over the 20-year study period, with only a small chance of a net benefit (Figure 4). These results were confirmed when using the optimization-based solution described in Appendix S1.

Assuming managers are better than random, and that the counterfactual failure risk remains at the most supported value $f = 0.56$, rescue would have positive benefits if eggs were classified correctly over two thirds of the time ($a > 0.69$; Figure 3). If accuracy and risk remained

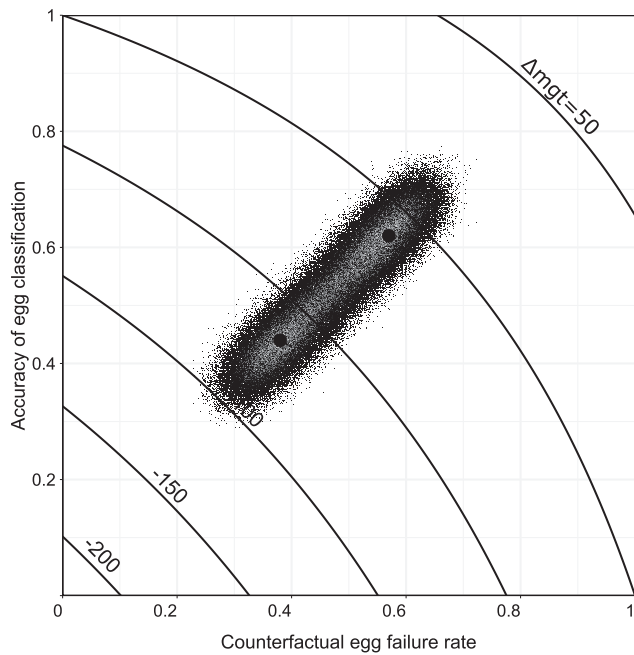


FIGURE 3 Counterfactual space. Each point represents a combination of the hypothetical accuracy of classifying eggs (a) and the failure rate of unmanaged eggs (f). The small points indicate 10,000 posterior samples for a and f from each of three Markov chains (lighter areas indicate more support). The two black dots indicate the mean posterior estimates for a and f when the priors are constrained to $U(0, 0.5)$ (bottom left dot) and $U(0, 1)$ (top right dot). The contour lines indicate the benefit of management for a given combination: for example, the $\Delta_{\text{mgt}} = -100$ line indicates values of a and f for which managing would have resulted in 100 fewer eggs than not managing; on the $\Delta_{\text{mgt}} = 0$ line, managing or not managing is indifferent (indifference threshold).

unchanged, managers would need to increase the fledging probability for rescued eggs to $p > 0.32$ for rescue to at least break even. This would represent an increase by almost 30% over the current mean $p = 0.25$, but is about the same as the current most effective action (donation).

4 | DISCUSSION

Tara iti egg rescue is typical of conservation interventions that evolve in the absence of true experimental controls, but still require good evaluation of effectiveness. Our results highlight the pitfalls of not evaluating monitoring data or doing so naively. Without a counterfactual analysis, one might conclude that egg rescue performs worse than leaving eggs in situ. Conversely, assuming that all eggs perceived as under imminent threat, and thus rescued, would have been lost if left in situ would be a simplistic counterfactual that simply confirms the benefit of interventions. Overall, our analysis indicated that current egg

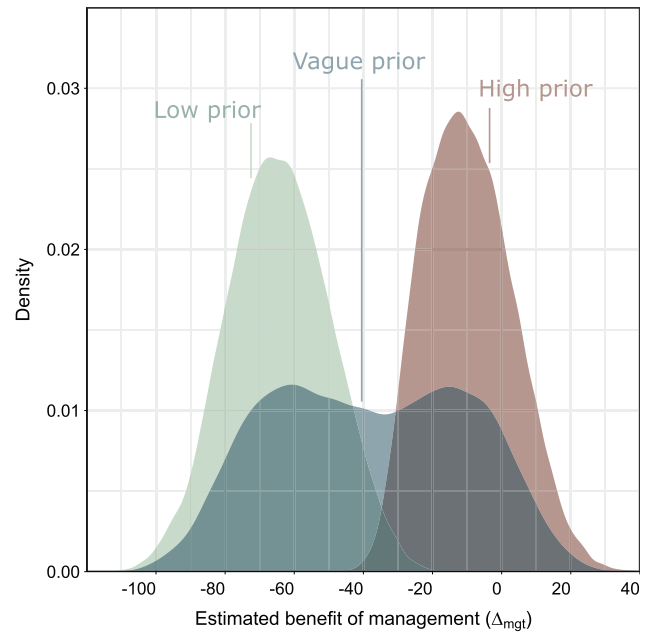


FIGURE 4 Posterior distributions of the estimated benefit of egg rescue (Δ_{mgt}), expressed as the difference between the number of fledglings observed and calculated for the estimated values of a and f . For example, the $\Delta_{\text{mgt}} = -100$ line indicates values of a and f for which managing would have resulted in 100 fewer eggs than not managing; for $\Delta_{\text{mgt}} = 0$, managing or not managing is indifferent. Curves represent posteriors for a and f priors of $U(0, 1)$ (“vague”), $U(0, 0.5)$ (“low”), and $U(0.5, 1)$ (“high”).

rescue might not benefit the population, but improving threat assessment and/or survival of rescued eggs would swing this balance.

To minimize subjective interpretation, we presented counterfactuals with uncertainty, which helped address risk aversion (Tulloch et al., 2015), and explored the entire counterfactual space, which helped consider improvements. Managers might try to improve the accuracy of decisions (the classification parameter f in our analysis), for example, by closer monitoring or greater use of technology. We assumed accuracy is equal across eggs and actions, but threats like spring tides are already easier to assess than, for example, predator sightings, so the latter might allow greater improvements. However, the specific perceived threat was not recorded for all eggs, so we could not estimate different classification rates. If managers cannot improve accuracy, they might change the threshold of perceived threat to trigger rescue. This would, however, incur a trade-off between false positive and false negative rates: with a higher threshold, fewer eggs might be rescued unnecessarily, but more eggs left in situ might be lost to avoidable threats. Managers might also offset classification errors by substantially increasing survival for rescued eggs. For tara iti, artificial incubation had low effectiveness, countering its general perception as a “safe” option

(Assersohn et al., 2021), but similar programs suggest there may be room for improvement (Dolman et al., 2015; Thompson et al., 2015; Collins et al., 2016). Yet another approach might be to implement controls to better learn about the key parameters; however, this was not possible here because of risk aversion for this critically endangered species.

While counterfactuals focus on alternatives available at the time of past decisions, they can also stimulate creative thinking for the future. For tara iti, alternatives could not always be chosen at will: for example, donation was the only alternative close to providing a net benefit by rescue, but it was constrained by availability of foster nests at any given time. The tara iti recovery team is already exploring alternative options beyond the current set, increasing the intensity of in situ habitat management and duration of predator control, and exploring options for captive breeding and release (McMurdo Hamilton et al., 2021). Other programmes have successfully kept tern eggs in situ by using nest platforms (Palestis, 2014) or shell patches (Booker & Moxom, 2019) instead of shifting. This could be combined with ongoing efforts to reduce the overall need for egg rescue, by improving habitat quality and keep nest sites safe from predators and tides (Raynor et al., 2012).

Even abundant data can be difficult to interpret and turn into decisions because of psychological barriers: for example, incorrect assessments of management effectiveness can favor status quo bias (Schwartz, 2020). In our case, rescued eggs had at least some success, which might facilitate a simplistic counterfactual that rescue was not perfect but still better than nothing. Our data-based assessment reduced this “storytelling” risk but required time and effort. First, it was made possible by careful collection of monitoring data over decades of management. Second, we found even our relatively simple counterfactual analysis cognitively challenging at times, with intricate hypotheticals and trade-offs. Tools like decision trees can assist this “slow thinking,” clarifying decisions and highlighting biases. Finally, decision analysis can be more effective if it is acceptable to recognize when actions are not effective and change course (Catalano et al., 2019). Conservation can be a long process that requires ongoing decisions as conditions change and programmes evolve. This process requires collecting monitoring data, analysing it, and interpreting results using the appropriate decision-support tools, to improve outcomes and minimize biases.

ACKNOWLEDGMENTS

We thank Katrina Hansen and David Wilson for managing the operational aspects of the tara-iti program during part of the period covered in this assessment. We acknowledge the contribution of past rangers and community volunteers, and tangata whenua to tara iti egg rescue and

conservation. This work was supported by the Department of Conservation and The Shorebirds Trust. TMH was supported by a Natural Environment Research Council PhD Studentship (NE/L002485/1). We thank two anonymous reviewers whose suggestions helped improve a previous version of this manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are all available in the article and its Supporting Information.

REFERENCES

- Assersohn, K., Marshall, A. F., Morland, F., Brekke, P., & Hemmings, N. (2021). Why do eggs fail? Causes of hatching failure in threatened populations and consequences for conservation. *Animal Conservation*, 24, 540–551. <https://doi.org/10.1111/acv.12674>
- Baling, M. (2008). Conservation of the fairy tern (*Sternula nereis* spp.) via subspecies level management. *Endangered Species*, 25, 86–93.
- Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A., & Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9, 58–64.
- Behn, R. D. (1982). *Quick analysis for busy decision makers*. Basic Books.
- Booker, H., & Moxom, D. (2019). Recovering the little tern colony at Chesil Beach, Dorset. *British Wildlife*, 31, 96–103.
- Bull, J. W., Strange, N., Smith, R. J., & Gordon, A. (2021). Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems. *Conservation Biology*, 35, 510–521.
- Butchart, S. H. M., Stattersfield, A. J., & Collar, N. J. (2006). How many bird extinctions have we prevented? *Oryx*, 40, 266–278.
- Buxton, R. T., Avery-Gomm, S., Lin, H. Y., Smith, P. A., Cooke, S. J., & Bennett, J. R. (2020). Half of resources in threatened species conservation plans are allocated to research and monitoring. *Nature Communications*, 11, 1–8.
- Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J. J., Southwell, D. M., Armstrong, D. P., Chadès, I., Lacy, R. C., & Converse, S. J. (2016). Adaptive management for improving species conservation across the captive-wild spectrum. *Biological Conservation*, 199, 123–131.
- Canessa, S., Taylor, G., Clarke, R. H., Ingwersen, D., Vandersteen, J., & Ewen, J. G. (2019). Risk aversion and uncertainty create a conundrum for planning recovery of a critically endangered species. *Conservation Science and Practice*, 2, e138.
- Catalano, A. S., Lyons-White, J., Mills, M. M., & Knight, A. T. (2019). Learning from published project failures in conservation. *Biological Conservation*, 238, 108223.
- Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., & Sutherland, W. J. (2019). Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, 56, 2742–2754.
- Coetzee, B. W. T., & Gaston, K. J. (2021). An appeal for more rigorous use of counterfactual thinking in biological conservation. *Conservation Science and Practice*, 3, e409.
- Collins, S. A., Sanders, F. J., & Jodice, P. G. R. (2016). Assessing conservation tools for an at-risk shorebird: Feasibility of

- headstarting for American Oystercatchers *Haematopus palliatus*. *Bird Conservation International*, 26, 451–465.
- Dolman, P. M., Collar, N. J., Scotland, K. M., & Burnside, R. J. (2015). Ark or park: The need to predict relative effectiveness of ex situ and in situ conservation before attempting captive breeding. *Journal of Applied Ecology*, 52, 841–850.
- Ferreira, S. M., Hansen, K. M., Parrish, G. R., Pierce, R. J., Pulham, G. A., & Taylor, S. (2005). Conservation of the endangered New Zealand fairy tern. *Biological Conservation*, 125, 345–354.
- Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, 122, 75–84.
- Hansen, K. (2006). *New Zealand fairy tern (Sterna nereis davisiae) recovery plan, 2005–15*. Science & Technical Pub., Department of Conservation.
- Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H., Carpenter, K. E., Chanson, J., Collen, B., Cox, N. A., Darwall, W. R., Dulvy, N. K., Harrison, L. R., Katariya, V., Pollock, C. M., Quader, S., Richman, N. I., Rodrigues, A. S., Tognelli, M. F., ... Stuart, S. N. (2010). The impact of conservation on the status of the world's vertebrates. *Science*, 330, 1503–1509.
- Hoffmann, M., Duckworth, J., Holmes, K., Mallon, D. P., Rodrigues, A. S., & Stuart, S. N. (2015). The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology*, 29, 1303–1313.
- Jellesmark, S., Ausden, M., Blackburn, T. M., Gregory, R. D., Hoffmann, M., Massimino, D., McRae, L., & Visconti, P. (2021). A counterfactual approach to measure the impact of wet grassland conservation on UK breeding bird populations. *Conservation Biology*, 35, 1575–1585.
- Jones, C. G., Cole, N., Canessa, S., Chauvenet, A., Fogell, D., & Ewen, J. G. (2018). Conserving island species: Journey to recovery. In J. Copley, S. Black, J. Groombridge, & C. Jones (Eds.), *Species conservation: Lessons from islands* (pp. 255–290). Cambridge University Press.
- Kellner, K., Meredith, M., & Kellner, M. K. (2019). Package 'jagsUI'. A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses. *R Package Version*, 1(1).
- Lindenmayer, D. B., Piggott, M. P., & Wintle, B. A. (2013). Counting the books while the library burns: Why conservation monitoring programs need a plan for action. *Frontiers in Ecology and the Environment*, 11, 549–555.
- Martin, T. G., Burgman, M. A., Fidler, F., Kuhnert, P. M., Low-Choy, S., McBride, M., & Mengersen, K. (2012). Eliciting expert knowledge in conservation science. *Conservation Biology*, 26, 29–38.
- McMurdo Hamilton, T., Canessa, S., Clark, K., Gleeson, P., Mackenzie, F., Makan, T., Moses-Te Kani, G., Oliver, S., Parker, K. A., & Ewen, J. G. (2021). Applying a values-based decision process to facilitate comanagement of threatened species in Aotearoa New Zealand. *Conservation Biology*, 35, 1162–1173.
- Palestis, B. G. (2014). The role of behavior in tern conservation. *Current Zoology*, 60, 500–514.
- Plummer, M. (2003). *JAGS: Just Another Gibbs Sampler*. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. <https://www.ci.tuwien.ac.at/Conferences/DSC-2003/>
- Raynor, E. J., Pierce, A. R., Leumas, C. M., & Rohwer, F. C. (2012). Breeding habitat requirements and colony formation by royal terns (*Thalasseus maximus*) and sandwich terns (*T. sandvicensis*) on barrier islands in the Gulf of Mexico. *Auk*, 129, 763–772.
- Roese, N. J., & Olson, J. M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology*, 32, 197–227.
- Robertson, H. A., Baird, K. A., Elliott, G. P., Hitchmough, R. A., McArthur, N. J., Makan, T. D., Miskelly, C. M., O'Donnell, C. F. J., Sagar, P. M., Scofield, R. P., Taylor, G. A., & Michel, P. (2021). Conservation status of New Zealand birds, 2016. New Zealand Threat Classification Series. <https://www.doc.govt.nz/globalassets/documents/science-and-technical/nztcs36entire.pdf>
- Runge, M. C. (2011). An introduction to adaptive management for threatened and endangered species. *Journal of Fish and Wildlife Management*, 2, 220–233.
- Schwartz, M. W. (2020). Conservation lessons from taboos and trolley problems. *Conservation Biology*, 35, 974–982.
- Thompson, T. R., Apa, A. D., Reese, K. P., & Tadvick, K. M. (2015). Captive rearing sage-grouse for augmentation of surrogate wild broods: Evidence for success. *The Journal of Wildlife Management*, 79, 998–1013.
- Treadgold, S. J. (2000). *Behavioural ecology of the endangered New Zealand Fairy Tern (Tara-iti) Sterna nereis davisiae: Implications for management*. Massey University. <https://mro.massey.ac.nz/handle/10179/7614>
- Tulloch, A. I. T., Maloney, R. F., Joseph, L. N., Bennett, J. R., Di Fonzo, M. M. I., Probert, W. J. M., O'Connor, S. M., Densem, J. P., & Possingham, H. P. (2015). Effect of risk aversion on prioritizing conservation projects. *Conservation Biology*, 29, 513–524.
- Walsh, J. C., Dicks, L. v., Raymond, C. M., & Sutherland, W. J. (2019). A typology of barriers and enablers of scientific evidence use in conservation practice. *Journal of Environmental Management*, 250, 109481.
- Young, R. P., Hudson, M. A., Terry, A. M. R., Jones, C. G., Lewis, R. E., Tatayah, V., Zuël, N., & Butchart, S. H. M. (2014). Accounting for conservation: Using the IUCN Red List Index to evaluate the impact of a conservation organization. *Biological Conservation*, 180, 84–96.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: McMurdo Hamilton, T., Ewen, J. G., Beauchamp, A. J., Makan, T., Rowcliffe, J. M., & Canessa, S. (2022). Data-driven counterfactual evaluation of management outcomes to improve emergency conservation decisions. *Conservation Letters*, e12925. <https://doi.org/10.1111/conl.12925>