



A counterfactual approach to measure the impact of wet grassland conservation on U.K. breeding bird populations

Sean Jellesmark,^{1,2*} Malcolm Ausden,⁴ Tim M. Blackburn,^{1,2} Richard D. Gregory,^{4,1} Mike Hoffmann,² Dario Massimino,⁵ Louise McRae,² and Piero Visconti³ †

¹Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, U.K.

²Institute of Zoology, Zoological Society of London, London, U.K.

³International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, Laxenburg, A-2361, Austria

⁴RSPB Centre for Conservation Science, The Lodge, Sandy, Bedfordshire, SG19 2DL, U.K.

⁵British Trust for Ornithology, The Nunnery, Thetford, IP24 2PU, U.K.

Abstract: Wet grassland populations of wading birds in the United Kingdom have declined severely since 1990. To help mitigate these declines, the Royal Society for the Protection of Birds has restored and managed lowland wet grassland nature reserves to benefit these and other species. However, the impact of these reserves on bird population trends has not been evaluated experimentally due to a lack of control populations. We compared population trends from 1994 to 2018 among 5 bird species of conservation concern that breed on these nature reserves with counterfactual trends created from matched breeding bird survey observations. We compared reserve trends with 3 different counterfactuals based on different scenarios of how reserve populations could have developed in the absence of conservation. Effects of conservation interventions were positive for all 4 targeted wading bird species: Lapwing (*Vanellus vanellus*), Redshank (*Tringa totanus*), Curlew (*Numenius arquata*), and Snipe (*Gallinago gallinago*). There was no positive effect of conservation interventions on reserves for the passerine, Yellow Wagtail (*Motacilla flava*). Our approach using monitoring data to produce valid counterfactual controls is a broadly applicable method allowing large-scale evaluation of conservation impact.

Keywords: causal inference, conservation effectiveness, impact evaluation, wetland conservation, wetland birds

Un Enfoque Hipotético para Medir el Impacto de la Conservación de Pastizales Húmedos sobre Poblaciones Reproductoras de Aves en el Reino Unido

Resumen: Las poblaciones de aves zancudas en los pastizales húmedos del Reino Unido han declinado gravemente desde 1990. Para ayudar con la mitigación de estas declinaciones, la Real Sociedad para la Protección de las Aves ha restaurado y manejado las reservas naturales en pastizales húmedos de tierras bajas para beneficiar a estas y otras especies. Sin embargo, el impacto de estas reservas sobre las tendencias poblacionales de las aves no ha sido evaluado experimentalmente debido a la falta de poblaciones control. Comparamos las tendencias poblacionales entre 1994 y 2018 de cinco especies de aves de importancia para la conservación, que se reproducen dentro de estas reservas naturales, mediante tendencias hipotéticas creadas a partir de los censos de observación de

*Address for correspondence to Piero Visconti, email: sean.jellesmark.18@ucl.ac.uk

Article impact statement: Post hoc quasi-experimental analysis shows that wet grassland conservation has positively affected targeted breeding bird populations.

†Current address: Biodiversity, Ecology and Conservation group, Biodiversity and Natural Resources Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria.

Paper submitted April 7, 2020; revised manuscript accepted January 4, 2021.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

aves reproductoras emparejadas. Comparamos las tendencias de las reservas con tres casos hipotéticos diferentes basados en diferentes escenarios de cómo las poblaciones de la reserva podrían haberse desarrollado en ausencia de la conservación. Los efectos de las intervenciones de conservación fueron positivos para las cuatro especies focales de aves zancudas: *Vanellus vanellus*, *Tringa totanus*, *Numenius arquata* y *Gallinago gallinago*. No hubo un efecto positivo de las intervenciones de conservación para la especie paserina *Motacilla flava*. Nuestra estrategia utilizando datos de monitoreos para producir controles hipotéticos válidos es un método ampliamente aplicable que permite la evaluación del impacto de la conservación a gran escala

Palabras Clave: aves de humedal, conservación de humedales, efectividad de la conservación, evaluación de impacto, inferencia causal

摘要: 自1990年以来,英国湿草地涉禽的数量急剧下降。为了缓解这些鸟类种群丧失,英国皇家鸟类保护协会通过恢复和管理低地湿草地自然保护区来帮助保护这些鸟类及其它物种。然而,由于缺乏对照种群,保护区对鸟类种群动态的影响还没有得到实验评估。本研究将在这些保护区繁殖的五种受保护鸟类1994-2018年的种群动态与配对的鸟类繁殖调查获得的反事实动态进行了比较。我们比较了保护区鸟类种群动态与三种反事实动态,这些反事实动态是基于保护区种群在没有保护的情况下如何发展的不同情景假设。结果表明,研究关注的四种涉禽:凤头麦鸡 (*Vanellus vanellus*)、红脚鹬 (*Tringa totanus*)、白腰杓鹬 (*Numenius arquata*) 和扇尾沙锥 (*Gallinago gallinago*) 的保护措施均取得了积极成效,而对保护区中雀形目鸟类黄鹌鹑 (*Motacilla flava*) 的保护干预没有积极作用。我们的方法利用监测数据产生了可靠的反事实对照,是一种可用于大规模评估保护影响的广适性方法。【翻译:胡怡思;审校:聂永刚】

关键词: 影响评估,湿地鸟类,湿地保护,因果推论,保护有效性

Introduction

Halting the decline of global biodiversity is currently one of humanity's greatest environmental challenges. Within animal populations, declines have predominantly been attributed to changes in land use, invasive species, exploitation of species and habitats, pollution, and climate change (IPBES, 2019). Different actors, including policy makers, nongovernmental organizations, and conservation practitioners, are addressing this global decline through a range of conservation actions, chiefly habitat and species management interventions in- and outside protected areas. From 2010 to 2018, £817 million was spent on average each year to promote and protect biodiversity in the United Kingdom alone (Department for Environment, Food & Rural Affairs, 2019). Despite these efforts, biodiversity indicators in the United Kingdom continue to decline (Hayhow et al., 2019). An important question in understanding the impact of conservation interventions on target populations is the extent to which those interventions mitigate or reverse population declines (Hoffmann et al., 2010, 2015). However, limited resources often mean that evaluation efforts do not extend beyond simple measures of association. Population trends are often monitored in protected areas, but appropriate control trends are not. Thus, whether population changes in target species are caused by the management measures or represent changes that would have occurred in the absence of that management remains untested.

To assess the impact of conservation, it is necessary to understand what would have happened in the absence of conservation, that is, the counterfactual conservation outcome (e.g., Baylis et al., 2016; Ferraro, 2009; Ferraro & Pattanayak, 2006). The exact form of the counterfac-

tual can never be known for certain. Ideally, a robust study design, such as a randomized controlled trial (RCT) (random assignment of treatment and control groups), could be used to infer the causal effect of a treatment by approximating the counterfactual outcome. However, RCT designs are rarely used in conservation. This is because randomization is often infeasible. For example, there can be legislative obligations to manage protected sites in ways considered beneficial to conservation, which makes it difficult to include unmanaged controls. In addition, the scale of conservation interventions and sampling units may be too large to allow for sufficient replication (Margoluis et al., 2009; Baylis et al., 2016; Wiik et al., 2019).

Conservation practitioners resort to other evaluation designs because of the financial, practical, and logistical challenges of the RCT design. These include after (A) methods (e.g., increasing or decreasing posttreatment population size), before-after (BA) methods (e.g., pre-treatment population changes are compared with post-treatment population changes), and control-impact (CI) methods (e.g., comparing population densities inside reserves with population densities outside reserves). Such approaches are important in determining the extent to which conservation objectives are being achieved and are a prerequisite for adaptive management. However, if potential biases are not properly addressed, these approaches cannot be used to determine cause and effect with a high level of confidence. The after study design describes the posttreatment rate of change and direction but does not provide insight into whether the change would have differed without the treatment. The before-after study design assumes that temporal variability and confounding factors before and after the intervention are

comparable, and control–impact assumes time-for-space substitution and comparability between groups. The validity of such inferences is therefore compromised if a population would have developed similarly regardless of conservation (e.g. A), if the effect of confounding variables is not homogenous across time (e.g., BA), and if local variation is systematically different between impact and control groups (e.g., CI) (e.g., Ferraro & Pressey, 2015; De Palma et al., 2018; Adams et al., 2019). To improve the credibility of an inference, the BA and the CI study design can be combined, forming the before–after–control–impact (BACI) study design (e.g., comparing pretreatment and posttreatment densities in a treated and a control group while accounting for the pretreatment density difference between treated and control group). Using simulated ecological data, the BACI design can be used to estimate the true effect size better than RCTs (1.3–1.8 times more likely to estimate $\pm 30\%$ of the true effect and direction), CIs (3.2–4.6 times more likely), and A study designs (7.1–10.1 times more likely) (Christie et al., 2019). However, this study design has many of the same limitations as the RCT and is further limited if appropriate controls cannot be identified *ex ante* (e.g., appropriate controls cannot be selected prior to measuring the outcome of interest if confounders are unknown or poorly understood).

To produce reliable conservation effect estimates, matching techniques are increasingly being used in conservation science (Sills et al., 2017; Schleicher et al., 2019; Sonter et al., 2019). The intent of matching is to create treatment and control groups with similar covariates by creating subsets of treatment and control samples so that comparisons are carried out with groups that have similar characteristics (e.g. comparing the outcome of a treated group to the outcome of a control group where both groups are from the same habitat type, elevation, and country). The postmatching control group then represents the counterfactual outcome of the treated group, and the effect of a given treatment can be inferred as the difference between outcomes. For example, Ferraro et al. (2007) tested the effectiveness of U.S. Endangered Species Act listing and funding on species recovery based on matching of a set of observable covariates to account for bias in the listing and funding process. They found listing is effective only when accompanied by adequate funding. Geldmann et al. (2019) assessed whether protected areas (PAs) reduce anthropogenic pressure. They used 10 variables linked to PA selection to match PAs to similar unprotected areas and found that, on average, PAs do not reduce human pressure. Nevertheless, although the theoretical potential of these methods has been highlighted, examples of their application remain scarce (Ferraro & Pattanayak, 2006; Margoluis et al., 2009; Joppa & Pfaff, 2010).

We adopted a matching approach to explore the impact of specific conservation interventions on a particular habitat of conservation concern in Europe: lowland

wet grassland (Franks et al., 2018). Conversion to other habitat types, changes in grazing regimes, drainage, and agricultural intensification have adversely affected these grasslands (Wilson et al., 2004). In particular, wetland bird species that use this habitat to breed and overwinter, such as wading birds (Charadriiformes), have exhibited severe breeding population declines as a result of these habitat changes (Wilson et al., 2005; Boatman et al., 2007; Colhoun et al., 2017). For example, Lapwing (*Vanellus vanellus*) populations, once abundant in the countryside of the United Kingdom, declined by 42% from 1995 to 2017 (Harris et al., 2019). To help mitigate these declines, the Royal Society for the Protection of Birds (RSPB) has allocated resources to purchasing, restoring, and managing reserves in lowland wet grassland habitats to benefit breeding wading birds in the United Kingdom. Conservation interventions, such as raising and manipulating water levels, beneficial stock grazing regimes, control and exclusion of generalist predators, and mechanical vegetation control, are implemented on these reserves (Ausden et al., 2019). Conservation efforts of this type are associated with increasing wading bird populations (Ausden & Hirons, 2002; Malpas et al., 2013; Smart et al., 2014). However, a central problem is whether the conservation actions result in positive benefits to the target populations: is the population performance better than would have occurred in the absence of these interventions? We tested this by comparing breeding trends on the reserves with matched counterfactual trends that represent how the trends may have developed in the absence of reserve-based conservation interventions. This is, to our knowledge, the first time post hoc evaluation of conservation interventions using quasi-experimental after-control-impact ACI analyses has been carried out for conservation interventions in the United Kingdom. We used trends after intervention thus after and matching reserve trends to counterfactual controls, therefore control–impact.

Methods

Data

We used bird counts from RSPB lowland wet grassland reserves and from the U.K. Breeding Bird Survey (Harris et al., 2019) for the period 1994–2018. The RSPB manages over 200 reserves across the United Kingdom; 47 of these contain lowland wet grassland (Appendix S11). Most of these reserves are in England (35); the rest are in Scotland (7), Wales (3), and Northern Ireland (2). We chose lowland wet grassland (i.e., periodically flooded grasslands below approximately 250 m elevation [Jefferson & Grice, 1998]) because this is a habitat in which considerable resources have been invested in habitat restoration and creation in recent decades. The area of lowland wet grassland on individual reserves varies from

18 to 1,300 ha (mean site area = 95 ha [SD 144]). Some reserves consist of two or more noncontiguous blocks of lowland wet grassland, which we refer to as *sites*. We used a total of 101 sites in the 47 reserves. We treated new acquisitions of land as separate sites. The RSPB reserves are managed in accordance with the biological requirements of priority species selected for that reserve. The number of breeding pairs of priority bird species are counted three times annually between April and June at each site with standard methods described in Gilbert et al. (1998) (Appendix S12).

The focal wetland species were Garganey (*Anas querquedula*), Shoveler (*A. clypeata*), Black-tailed Godwit (*Limosa limosa*), Lapwing, Curlew (*Numenius arquata*), Snipe (*Gallinago gallinago*), Redshank (*Tringa totanus*), and Yellow Wagtail (*Motacilla flava*). Analysis concentrated on the latter five abundant species. These species were chosen for practical reasons. First, populations breed on reserves; second, they are currently RSPB priority species and have been monitored both on reserves and in the wider countryside (see below); and third, and most importantly, conservation interventions are designed to closely match their biological breeding requirements, making the number of breeding birds a natural response to the conservation type we evaluated (Appendix S4).

In the case of Snipe and Yellow Wagtail, a large proportion of their breeding reserve population (59% and 90%, respectively, at the start of the period analyzed) occurred at a single reserve, the Ouse Washes in Norfolk and Cambridgeshire. This site is atypical because breeding birds are sometimes disrupted by flooding during the breeding season; the site is designed to temporarily store floodwater. This flooding is outside the control of the reserve management and explains population declines for Black-tailed Godwit (Ratcliffe et al., 2005). We therefore carried out analyses with and without the Ouse Washes for Snipe and Yellow Wagtail.

We obtained matching data to compute counterfactual population trends from the U.K. Breeding Bird Survey (BBS), managed by the British Trust for Ornithology. This scheme was started in 1994 and monitors changes in the national breeding trends of more than a hundred common and widespread bird species (Gregory et al., 2000; Harris et al., 2019). Surveying is performed in 1 × 1 km grids, each consisting of 10 transects. The type of habitat is recorded in a separate visit prior to 2 annual bird counts between April and June (Appendix S12). We used the habitat data recorded in the BBS and elevation data from the OS terrain 50 data set and the USGS EROS Archive - Digital Elevation (SRTM) 1 Arc-Second Global to calculate mean elevations.

We selected observations from lowland wet grassland sites and target species to create 1 reserve sample (i.e., treated sites) and matched the BBS data exactly on co-

variates affecting reserve selection and breeding trend (Table 1) to create the counterfactual sample (i.e. the control sites) for each species. We call this our benchmark counterfactual, as opposed to 2 other variants introduced to test sensitivity of the results (see below and Table 1). The counterfactuals were created by selecting observations from BBS grids containing certain habitats (Table 1) because we believe these are the best approximations of how reserve land would have developed without reserve conservation. We did not set a minimum proportion of the selected habitats or the exact mix of habitats that a grid had to contain to be included in the counterfactual sample. In the BBS, birds are counted in transect of 200 m and habitat is determined similarly. This also means that both bird numbers and habitat distinction come with some uncertainty regarding exactly where habitat changes and birds are observed. To account for this uncertainty, we operated on 1-km grid level.

We used a directed acyclic graph (DAG) to present our hypothesis for how wetland conservation affects breeding trends and to select matching covariates (Fig. 1) (e.g., Stuart, 2010; Pearl & Mackenzie, 2018; Hernan & Robins, 2020). Lowland wet grassland conservation is a cause of change in habitat quality (habitat, hydrology, food availability, and predator pressure) (Smart & Coutts, 2004; Verhulst et al., 2007; Eglington et al., 2008; Acreman et al., 2010; Ausden & Bolton, 2012; Smart et al., 2014), which then causes a change in the breeding trend. Habitat quality is improved by converting or forming the habitat from other habitat types to grassland, by changing the hydrological conditions using water control structures and land forming, by maintaining a habitable sward through grazing by domestic livestock and mowing; by mechanically removing shrubs and trees to remove perches for avian predators; and by reducing the impact of predation by controlling or excluding generalist predators.

We excluded counts from the matched control sample if they originated from grids spatially overlapping with the chosen reserves (see “Stable Unit Treatment Value Assumption” in Rubin [1980]). Transect counts were summarized for each grid, excluding transect counts with >10 individuals as birds on passage because it is not likely the study species breed in such high densities (Field & Gregory, 1999). The maximum annual grid count for each species was used, and grids surveyed only once over the entire period were excluded. Furthermore, to avoid uncertain trend estimates, we excluded all BBS species that were observed in <30 grids annually (Newson et al. 2009). Pre-analysis data manipulation and graphics were done with the tidyverse packages (Wickham et al., 2019) and DAGs with the dagitty package (Textor et al., 2016). All analysis, visualization, and manipulation were implemented using R version 3.5.1 (R Core team, 2019).

Table 1. Variables used in the creation of the benchmark, liberal, and stringent counterfactuals in an examination of the impact of wetland reserves on breeding wading bird numbers.

Benchmark	Liberal	Stringent
Elevation <250 m	From 1994–2018	Elevation <250m
From 1994 to 2018	From the United	From 1994 to 2018
From the United Kingdom	Kingdom	From the United Kingdom
Contains the target species	Contains target species	Contains the target species
Grids* containing seminatural grassland or marsh (chalk downland, grass moor, grass moor mixed with heather, machair other dry grassland, water meadow or grazing marsh, reed swamp, other open marsh or saltmarsh)		Grids* containing seminatural grassland types more similar to wet grassland (dry grasslands, water meadows or grazing marsh, reed swamp or other open marsh)

*Habitat data from the BTO/JNCC/RSPB Breeding Bird Survey recorded by volunteer surveyors using a standardized habitat recording form that allows each survey transect 2 primary habitat types. For more information see Appendix S12.

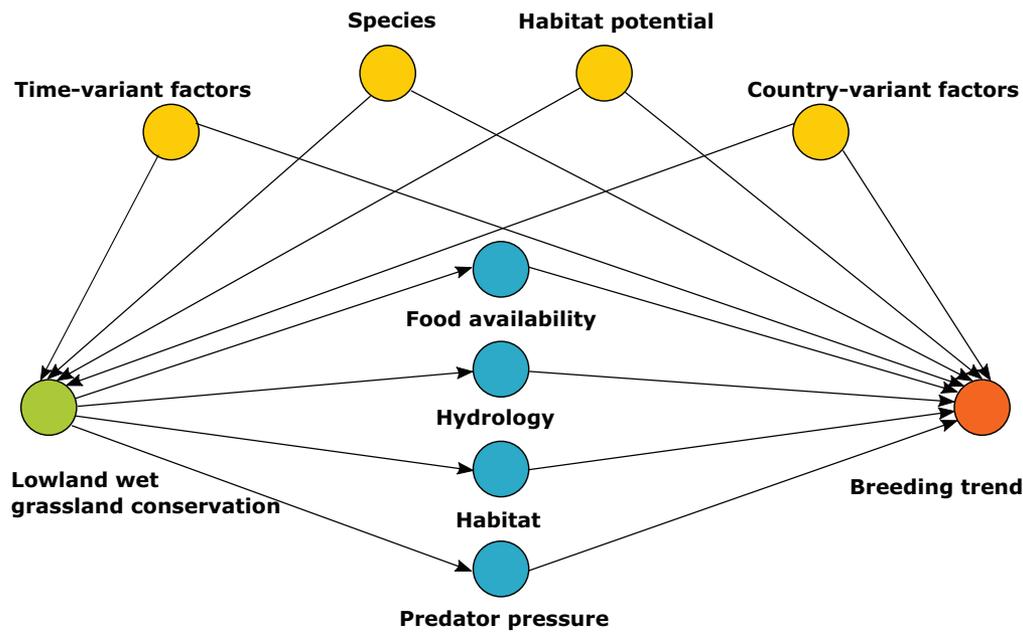


Figure 1. Hypothesized effect of management of habitat, hydrology, food availability, and predators in lowland wet grasslands on breeding bird trends (yellow circles, confounding factors that may affect reserve selection and the breeding trends).

Data analyses

We used imputed counts to calculate the species totals used to create reserve and counterfactual trend indices. Imputed means that if a given site (BBS grid or reserve site) in a given year has been monitored, then the observed count is used; otherwise, the missing count is estimated (Appendix S14). Missing population counts were estimated separately for each species x reserve or counterfactual combination with a loglinear model with Poisson error terms. Each count was modeled as a function of site and year effects (Eq. 1) with the rtrim package (Bogart et al., 2020). The SE was adjusted for overdispersion and temporal autocorrelation (Bogart et al., 2020; Pannekoek et al., 2018).

$$\ln Y_{ij} = \alpha_i + \beta_j, \tag{1}$$

where Y_{ij} is the estimated count for site i at time j , α_i is the average log-count of site i , and β_j is the average log-count deviation at time j across all sites.

We used indices to reflect relative changes in breeding pairs through time. The indices were calculated by dividing each annual total imputed count by a reference value that was set as the total count in the first time point (year 1994). Each set of indices was then tested against its counterfactual to examine whether the two sets of indices were different based on a Welch 2-sample t test. If any difference could be statistically substantiated ($p < 0.05$), the effect size was assessed as the mean trend of the counterfactual indices subtracted from the corresponding annual reserve indices.

A concern with quasi-experimental inferences is whether the correct variables have been included in the matching process (Stuart, 2010). We therefore

Table 2. Number of lowland wet grassland reserve sites and breeding bird survey* (BBS) grids used to create the reserve and counterfactual indices for each study species.

<i>Species</i>	<i>No. of reserve sites</i>	<i>Benchmark counterfactual (number of BBS grids)</i>	<i>Liberal counterfactual (number of BBS grids)</i>	<i>Stringent counterfactual (number of BBS grids)</i>
Curlew	23	371	2477	267
Lapwing	97	487	3223	380
Redshank	87	140	589	108
Snipe	56	227	1212	147
Yellow Wagtail	29	105	1019	82
Total number of sites/grid	101	1377	6507	1071

*Royal Society for the Protection of Birds, British Trust for Ornithology, and Joint Nature Conservation Committee Breeding Bird Survey.

created two alternative counterfactuals, imposing different matching requirements (Table 1). We created a liberal counterfactual that imposes only exact species as a covariate restriction. The liberal counterfactual relaxes the criterion to define like for like in control populations but has the potential advantage of increasing the number of control populations. This counterfactual assumes that, on average, the reserve populations would have developed like any other population in the United Kingdom. We also created a ‘stringent’ counterfactual that matches on exact species observations and has a subset of the habitat types used in the benchmark that is closer to the lowland wet grasslands in RSPB reserves. That is, matching grids were lowland (mean elevation below 250 m) and contained transects of either dry grassland; water meadows or grazing marsh; reed swamp; or open marshland. The stringent counterfactual thus assumes that, for each species, the average reserve trend would have developed like that of an average primarily lowland wet habitat regardless of conservation action. The increase in similarity requirements of matching populations comes at the cost of further limiting their numbers, thus potentially reducing the statistical power of the analyses. However, it might better describe the effect of conservation by reducing confounding effects. We assessed whether the results were robust to the counterfactual used by comparing the *t*-test results from both the liberal and stringent counterfactual (each one tested separately against the reserve indices) with the *t*-test results of the benchmark counterfactual (benchmark indices tested against reserve indices). We also examined the relationship between site age and changes in breeding counts and whether reserve trends were sensitive to exclusion of sites with large breeding counts (Appendix S8).

Results

Shoveler, Garganey, and Black-tailed Godwit were not sufficiently represented in the BBS data to create valid benchmark counterfactuals but showed either stable or

increasing trends on reserves (Appendix S5). The distribution of the remaining target species across lowland wet grassland reserve sites varied considerably. Lapwing and Curlew were present on most reserve sites and BBS grids, and Yellow Wagtail and Redshank were consistently rarer than other species, regardless of the counterfactual approach used (Table 2 and Appendix S2). The BBS grids used for the benchmark counterfactuals consisted primarily of farmland (45.5%), wet grassland transects (seminatural grassland types used in the stringent counterfactual in Table 1) (19.9%), and other seminatural grassland transects (remaining seminatural grassland types) (12.7%), whereas the liberal counterfactuals consisted primarily of farmland (67.3%) and other habitat types (24.7%). The stringent counterfactuals consisted primarily of wet grassland transects (27.6%) and farmland (47.4%) (Appendix S3). The largest relative increase in breeding pairs occurred in the first 10 years of reserve creation (Appendix S9).

The breeding indices for Snipe and Yellow Wagtail across all lowland wet grassland reserves could not be statistically distinguished from their benchmark counterfactuals (Snipe: $t = 1.9$, $df = 40$, $p = 0.07$; Yellow Wagtail: $t = -0.3$, $df = 39$, $p = 0.79$). However, when the Ouse Washes was excluded from the reserve data set (because its spring flooding is known to negate the effect of wetland management), the Snipe indices became more positive than its benchmark counterfactual (Fig. 2 & Appendix S6) ($t = 4$, $df = 47$, $p = 0.0002$). The indices for Yellow Wagtail were unchanged by this exclusion (Appendix S6).

Indices of Lapwing ($t = 7.6$, $df = 40$, $p < 0.0001$), Redshank ($t = 9.4$, $df = 45$, $p < 0.0001$), and Curlew ($t = 5.3$, $df = 35$, $p < 0.0001$) were all more positive on reserves. The mean annual trend difference represented an improvement of around 2.4% for Lapwing, 4.5% for Redshank, 1.5% for Snipe (Ouse Washes excluded), and 1.4% for Curlew. Thus, from 1994 to 2018 on lowland wet grassland reserves, Snipe populations increased by 36%, whereas the benchmark counterfactual remained stable around 1, suggesting that conservation interventions on these reserves were responsible for that

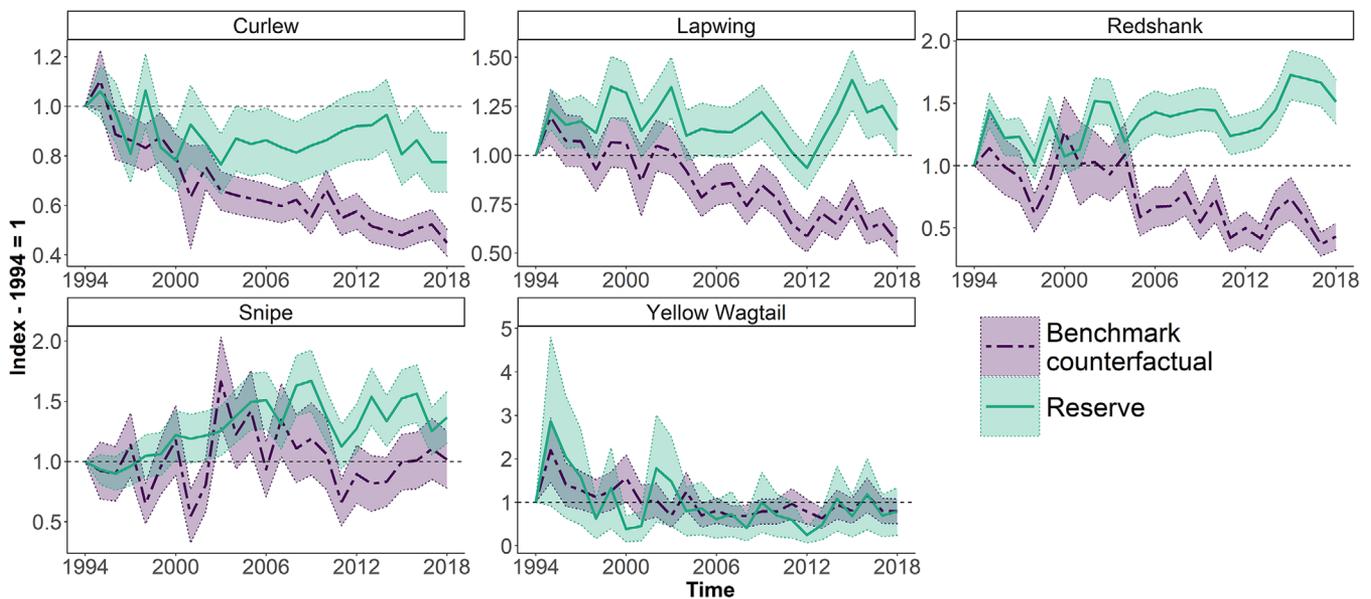


Figure 2. Breeding trends from 1994 to 2018 for the 5 target bird species inside reserves (solid line) and the benchmark counterfactual trends (dashed line) (shading, SE). The Ouse washes reserve was excluded for Snipe. Indices are calculated using imputed counts from loglinear models (see Methods).

increase. Curlew populations decreased by 23% compared with a 55% decline on the benchmark counterfactual, implying a 33% index improvement caused by conservation interventions on reserves. From 1994 to 2018, Lapwing populations increased by 13%, but the benchmark counterfactual suggested they would have decreased by 44% without conservation interventions, resulting in a 57% index improvement by conservation. Redshank populations on reserves increased by 51%, whereas the benchmark counterfactual decreased by 57% without conservation, attributing a relative index improvement of 108% to conservation interventions.

Regardless of which counterfactual we compared with, the reserve indices were more positive for the 4 wading bird species and similar for Yellow Wagtail (Fig. 3). The 3 counterfactuals were very similar for Lapwing, Redshank, and Yellow Wagtail but more dissimilar for Curlew and Snipe. The difference between the Curlew reserve indices and its liberal counterfactual became less pronounced ($t = 2.4$, $df = 39$, $p = 0.02$) (Fig. 3) than when the reserve indices were compared with the benchmark scenario, whereas the reserve indices differed more from their stringent counterfactuals for both Curlew ($t = 5.1$, $df = 32$, $p < 0.0001$) and Snipe ($t = 10.2$, $df = 48$, $p < 0.0001$).

Discussion

We used a quasi-experimental approach to demonstrate how long-term population monitoring data can be used to evaluate the impact of conservation. We found that

lowland wet grassland conservation has benefitted Lapwing, Redshank, and Curlew populations and, if an atypical site is excluded, that it also benefitted Snipe. We found no reserve effect for Yellow Wagtail and were not able to compare breeding populations of three other species (Black-tailed Godwit, Garganey, and Shoveler) because they were too rare outside of nature reserves, although they showed either stable or increasing trends on reserves. Based on the benchmark counterfactual trends, Snipe (Ouse Washes excluded), Lapwing, and Redshank populations all increased on reserves, but would have decreased or remained stable without this conservation, whereas Curlew populations decreased much less on reserves than they would otherwise have done. For the 4 wading bird species, the reserve indices were higher than their counterfactuals regardless of which counterfactual they were compared with; positive effects of reserve conservation were strong in all cases. However, different counterfactuals can produce different results, here illustrated by the different counterfactual trends in each species (Fig. 3). The effect of reserve conservation became less pronounced for Curlew under the liberal counterfactual, suggesting that this species may be faring slightly better in habitats other than wet grassland. Nevertheless, the differences in the three counterfactual trends for Curlew were small (Fig. 3). Overall, our findings concur with others (Ausden et al., 2019; Verhulst et al., 2007) in substantiating the positive effects of conservation actions on target breeding wetland bird populations.

The target wading bird species in our study should theoretically benefit from lowland wet grassland

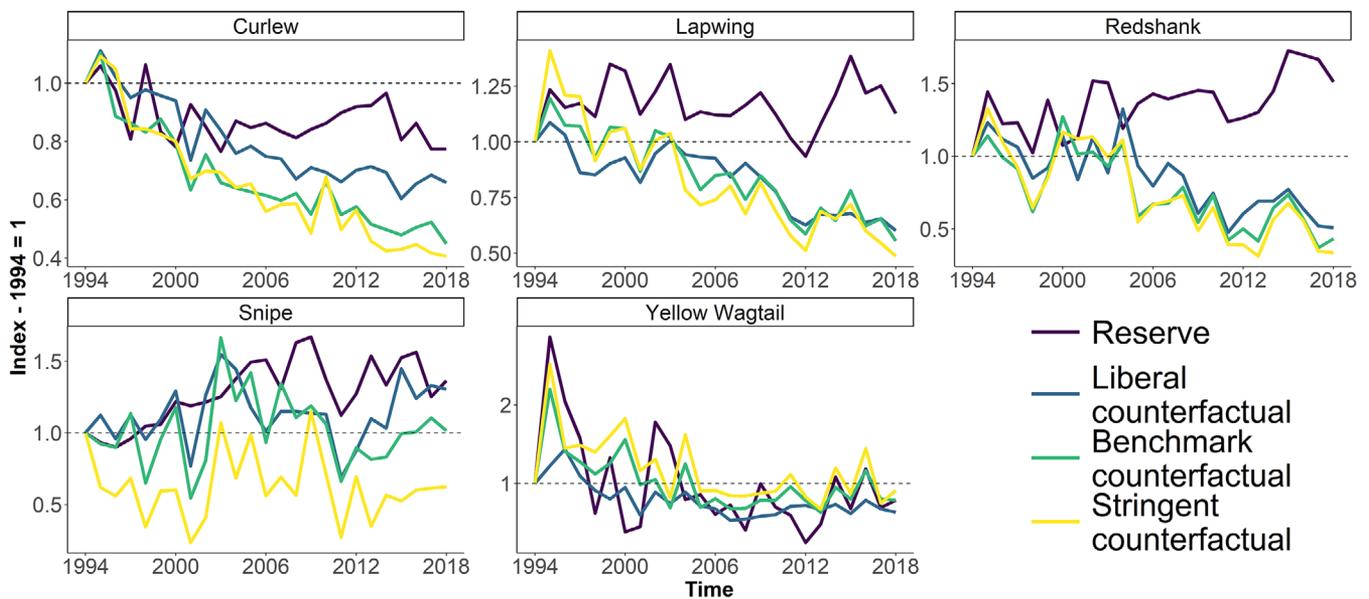


Figure 3. Reserve and counterfactual breeding trends from 1994 to 2018 based on the liberal, benchmark, and the stringent matching settings. The Ouse washes reserve was excluded for Snipe. Indices are calculated using imputed counts from loglinear models (see Methods). To ease visualization, SEs are excluded (see Appendix S7 for figure with SEs).

conservation, but not necessarily in equal measure. European grassland-breeding wading birds display species-specific responses to different types of grassland conservation (Franks et al., 2018). Wetland conservation management incorporates a range of different intervention types – from the conversion of, for example, ex-arable land to grassland, to changes in hydrology and grazing and mowing regimes. The degree to which each intervention type provides suitable conditions for the different study species may therefore differ. For example, Ausden et al. (2019) suggest that limiting livestock grazing in spring, which aims to reduce trampling of wading birds' nests, could also reduce habitat quality for Yellow Wagtail because they often feed in close association with domestic livestock. While Yellow Wagtail breed in wetland habitats, it has not been a priority species until recently and has not been actively targeted by management. This species is also the only long-distance migrant among the study species, and changes on its wintering grounds in Africa and migration paths may also affect its breeding population (Wood, 1992; Newton, 2006), thereby rendering conservation efforts in the breeding range less effective or redundant.

There are also multiple reserve specific conditions we did not account for. For example, because of improved breeding conditions, new sites recruit breeding pairs faster than older reserve sites (Appendix S9). Further research is needed to explore why reserve effects differ across study species (e.g. the declining reserve trend for Curlew in contrast to the increasing reserve trends for Redshank, Lapwing, and Snipe) and in particular how

population responses relate to site-specific interventions, reserve age and size, and finer-scale abiotic and habitat covariates.

We created separate reserve and counterfactual indices for each species based on the total annual number of breeding pairs. Because of the method used, a large decline on one reserve and stable or slightly increasing breeding numbers in all other reserves could still produce a decreasing trend if the total number of breeding birds declined overall. This can potentially mask the individual reserves' conservation success, as illustrated when excluding the Ouse Washes from the analysis of Snipe populations. However, our results were largely robust to exclusion of sites with large proportions of breeding numbers (Appendix S8).

The method we used provides several benefits over other evaluation methods for conservation impacts. It allows the use of population monitoring data sets to emulate a robust ex post study design. The interpretation of the results is intuitive (diverging lines in Fig. 2 mean that the observed scenario differs from its counterfactual), and results are easily communicated to an audience without statistical knowledge. Although our method is marginally more complex than study designs such as the "After," it does not require more resources. European monitoring data, such as the BBS data, are often freely available.

This method also allows a more detailed analysis of impacts than other study designs. For example, using the "After" method, which examines the reserve trend after the establishment of the reserve exclusively, Redshank

and Snipe would be the only species with a clear increasing trend. Assessing whether reserve conservation works exclusively based on whether a population trend is increasing implicitly assumes that the population would remain stable in the absence of conservation, which is far from the reality of ongoing population declines outside reserves (Harris et al., 2019). If the assessment had been done using a classical land-use control-impact study design, where the number of birds in each reserve would have been counted at one point in time, we would be able to compare densities but not trends. Our method (after-control-impact) *ex post* compares trends and depicts the dynamic development of populations through time, whereas control-impact studies provide only a temporal snapshot. The dynamic element is advantageous because it allows identification of divergent mechanisms through time and shows visually how adding new reserves affects the overall reserve trend.

Matching is increasingly being used in combination with regression techniques to assess the effect of conservation initiatives (Terraube et al., 2020). However, matching alone does not necessarily improve effect inferences and, because of reductions in sample size, may not have the same power to detect effects as regression techniques (Brazauskas & Logan, 2016). The RSPB reserve and BBS data sets we used covered long periods (>20 years) and included breeding bird counts derived from robust study designs. Such data sets are not common, and a quasi-experimental evaluation design like ours will not necessarily be applicable or appropriate elsewhere (see Walker et al., [2018] for alternative impact evaluation using BBS monitoring data). Furthermore, for matching to be appropriate, it requires a clear theory of how the treatment changes the outcome (Fig. 1 and “Data”) and careful selection of matching variables and methods accordingly (Schleicher et al., 2019). Using exact matching, we were able to retain sufficiently large sample sizes to run the loglinear models for 5 out of 8 species. Other quasi-experimental designs with fewer data or higher covariate complexity (higher number of covariates or continuous covariates) will either be impractical or require other matching methods (Iacus et al., 2019).

Reserves and BBS grids are surveyed using different survey protocols. Some of these differences could potentially lead to larger uncertainty and year-on-year variance; however, we do not believe this is the case. Each grid or site is surveyed with consistent effort each year, which means that a potential bias is also consistent and accounted for by using indices. Additionally, the counterfactuals created from the BBS are generally based on a relatively large number of annual observations. For further discussion see Appendix S13.

One way to create credible counterfactuals is through well-monitored control areas. This should reduce the likelihood of a mis-specified control group and enhance the credibility of the inference, but in order to make this

possible, monitoring of control sites must be a priority, with a further emphasis on consistent survey method. This may be difficult for the reasons described in the Introduction. Our results nonetheless suggest that dedicated conservation efforts have benefited target lowland wet grassland bird species and that monitoring programs can be used to evaluate the impact of conservation interventions by creating credible counterfactuals through matching approaches.

Acknowledgments

We are grateful to the volunteers who have contributed to the BTO, JNCC, and RSPB Breeding Bird Survey and to the BTO and RSPB for providing the data. S.J. thanks G. Albaladejo, S. Dove, and B. Reijenga for support and feedback in the early stages of this research. This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 766417. This communication reflects only the authors’ view, and the Research Executive Agency of the European Union is not responsible for any use that may be made of the information it contains.

Authors’ Contributions

P.V. conceived the initial project idea; S.J. designed the study with inputs from T.B., R.G., M.H., L.M.; P.V. and S.J. performed the analysis, designed the figures, and wrote the first draft of the paper; T.B., R.G., and P.V. commented on each draft; M.H. and L.M. commented on most drafts; and R.G., M.A., and D.M. helped with reserve and BBS data provision and analysis. All authors commented on later versions of the manuscript.

Supporting Information

Appendix S1 Number of sites or grids per year available in the reserve and counterfactual indices. Appendix S2 Percentage of data points (site or grid × year) observed and the total number of estimated and observed data points used to create the reserve and counterfactual trends. Appendix S3 The percentage of transects from all five target species used in the counterfactuals which contains the grassland types “Other dry grassland,” “Water meadows/ grazing marsh,” “Reed swamp,” and “Other open marsh” (Wet grassland), remaining semi-natural grassland in at least one of the two primary habitat categories but none of the wet grassland habitat types (Grassland), farmland but not grassland (Farmland), and the percentage of transects which contain neither of the above habitat types (Other). Appendix S4 Proposed

relationship between four categories of wetland management and the species that benefit from these (X = benefit). Appendix S5 Reserve trends from 1994 to 2018 for Black-tailed godwit, Garganey, and Shoveler. Appendix S6 The Snipe and Yellow Wagtail reserve trend with and without the Ouse Washes. Appendix S7 Reserve and counterfactual breeding trends from 1994 to 2018 using the liberal, benchmark and the stringent matching settings as in earlier analysis. Appendix S8 Reserve and counterfactual breeding trends from 1994 to 2018. Appendix S9 Relationship between site age (using first site count as year 0) and annual change % in breeding pairs and corresponding 95% CI (gray-shaded area). Appendix S10 Reserve and counterfactual breeding trends from 1994 to 2018 using reserve sites which were under RSPB management in 1994 and excluding Ouse Washes (Number of sites: Curlew = 10; Lapwing = 22; Redshank = 22; Snipe = 19; Yellow Wagtail = 4). Appendix S11 Location of Lowland wet grassland reserves (green bird symbol) and BBS grids (red dots) used in the analysis.

Literature Cited

- Acreman M. C., Mountford J. O., Gowing D. J., Stratford C., Durrell S., Duenas M. A., Rispen E., Youngs E., & Moy I. L. 2010. Restoration of wet grasslands through re-instatement of surface grips. Extended Final report to Department for Environment, Food and Rural Affairs. Centre for Ecology and Hydrology, Wallingford.
- Adams V. M., Barnes M., & Pressey R. L. 2019. Shortfalls in Conservation Evidence: Moving from Ecological Effects of Interventions to Policy Evaluation. *One Earth*, **1**(1), 62–75.
- Ausden M., & Hirons G. J. M. 2002. Grassland nature reserves for breeding wading birds in England and the implications for the ESA agri-environment scheme. *Biological Conservation*, [https://doi.org/10.1016/S0006-3207\(01\)00254-3](https://doi.org/10.1016/S0006-3207(01)00254-3)
- Ausden M., & Bolton M. 2012. Breeding waders on wet grassland: factors influencing habitat suitability. In (R. Fuller Ed.), *Birds and Habitat: Relationships in Changing Landscapes* (Ecological Reviews, pp. 278–306). Cambridge: Cambridge University Press.
- Ausden M., Hirons G., & White G. 2019. Wetland restoration by the RSPB – what has it achieved for birds? *British Birds*, **112**(June), 315–336.
- 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.
- Brazauskas R., & Logan B. R. 2016. Observational studies: matching or regression?, **25**(3), 289–313.
- Boatman N. D., Parry H. R., Bishop J. D., & Cuthbertson A. G. S. 2007. Impacts of agricultural change on farmland biodiversity in the UK. *Environmental Science and Technology*, **25**. <https://doi.org/10.1039/9781847557650-00001>
- Bogart P., van der Loo M., & Pannekoek J. 2020. rtrim: Trends and Indices for Monitoring Data. R package version 2.1.1. <https://CRAN.R-project.org/package=rtrim>
- 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*, **5**: 2742–2754.
- Crick H. Q. P. 1992. A bird-habitat coding system for use in Britain and Ireland incorporating aspects of land-management and human activity. *Bird Study*, **39**(1), 1–12.
- Colhoun K., Mawhinney K., McLaughlin M., Barnett C., McDevitt A. M., Bradbury R. B., & Peach W. 2017. Agri-environment scheme enhances breeding populations of some priority farmland birds in Northern Ireland. *Bird Study*, **64**(4), 545–556.
- De Palma A., Sanchez-Ortiz K., Martin P. A., Chadwick A., Gilbert G., Bates A. E., Börger L., Contu S., Hill S. L. L., & Purvis A. 2018. Challenges with inferring how land-use affects terrestrial biodiversity: study design, time, space and synthesis. In *Advances in Ecological Research*, **58**(2), <https://doi.org/10.1016/bs.aecr.2017.12.004>
- Department for Environment, Food and Rural Affairs (DEFRA). 2019. UK Biodiversity Indicators 2019. DEFRA,
- Eaton M. A., Gibbons D. W., Evans A., Gregory R. D., Brown A. F., Noble D. G., Musgrove A. J., Hearn R. D., & Aebischer N. J. 2015. Birds of conservation concern 4: the population status of birds in the united kingdom, channel islands and isle of Man. *British Birds*, **108**(6), 708–746.
- Eglinton S. M., Gill J. A., Bolton M., Smart M. A., Sutherland W. J., & Watkinson A. R. 2008. Restoration of wet features for breeding waders on lowland grassland. *Journal of Applied Ecology*, **45**: 305–314.
- Ferraro P. J. 2009. Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, **122**: 75–84.
- Ferraro P. J., McIntosh C., & Ospina M. 2007. The effectiveness of the US endangered species act: an econometric analysis using matching methods. *Journal of Environmental Economics and Management*, **54**(3), 245–261.
- Ferraro P. J., & Pattanayak S. K. 2006. Money for nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments, **4**(4), 482–488.
- Ferraro P. J., & Pressey R. L. 2015. Measuring the difference made by conservation initiatives: protected areas and their environmental and social impacts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370**(1681), 4–8.
- Field R. H., & Gregory R. D. 1999. Measuring population changes from the Breeding Bird Survey. BTO Research Report No. 217. The Nunnery, Thetford, Norfolk.
- Franks S. E., Roodbergen M., Teunissen W., Carrington Cotton A., & Pearce-Higgins J. W. 2018. Evaluating the effectiveness of conservation measures for European grassland-breeding waders. *Ecology and Evolution*, **8**(21), 10555–10568.
- Geldmann J., Manica A., Burgess N. D., Coad L., & Balmford A. 2019. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America*, **116**(46), 23209–23215. <https://doi.org/10.1073/pnas.1908221116>
- Gilbert G., Gibbons D. W., & Evans J. 1998. Bird monitoring methods. RSPB, Sandy.
- Gregory R., Baillie S., Bashford R. 2000. Monitoring breeding birds in the United Kingdom. *Bird Census News*, **13**: 101–112.
- Harris S. J., Massimino D., Eaton M. A., Gillings S., Noble D. G., Balmer D. E., Pearce-Higgins J. W., & Woodcock P. 2019. The breeding bird survey 2018. Research report 717. British Trust for Ornithology, Thetford.
- Hayhow D. B., et al. 2019. The state of nature 2019. The State of Nature partnership.
- Hernan M. A., & Robins J. M. 2020. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, Florida.
- Hoffmann M., Duckworth J. W., Holmes K., Mallon D. P., Rodrigues A. S. L., & Stuart S. N. 2015. The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology*, **29**: 1303–1313.
- Hoffmann M., et al. 2010. The impact of conservation on the status of the world's vertebrates. *Science*, **330**(6010), 1503–1509.
- Iacus S. M., King G., & Porro G. 2019. A theory of statistical inference for matching methods in causal research. *Political Analysis*, **27**: 46–68.
- IPBES 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental

- Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, et al (eds.). IPBES Secretariat, Bonn, Germany.
- Jefferson R G, & Grice P. V. 1998. The conservation of lowland wet grassland in England. In C.B. Joyce & P.M. Wade (eds) European wet grasslands: biodiversity, management and restoration: 31–48: Wiley, Chichester.
- Joppa L, & Pfaff A. 2010. Reassessing the forest impacts of protection. *Annals of the New York Academy of Sciences*, **1185**(1), 135–149.
- Malpas L. R, Kennerley R. J, Hirons G. J. M, Sheldon R. D, Ausden M, Gilbert J. C, & Smart J. 2013. The use of predator-exclusion fencing as a management tool improves the breeding success of waders on lowland wet grassland. *Journal for Nature Conservation*, **21**: 37–47.
- Margoluis R, Stem C, Salafsky N, & Brown M. 2009. Design alternatives for evaluating the impact of conservation projects. *New Directions for Evaluation*, **122**: 85–96.
- Newson S. E, Ockendon N, Joys A, Noble D. G, Baillie S. R. 2009. Comparison of habitat-specific trends in the abundance of breeding birds in the UK. *Bird Study*, **56**: 233–243.
- Newton I. 2006. Can conditions experienced during migration limit population levels of birds. *Journal of Ornithology*, **147**: 146–166.
- Pannekoek J, Bogaart P, & van der Loo M. 2018. Models and statistical methods in rtrim. Statistic Netherlands, Haga.
- Pearl J, & Mackenzie D. 2018. *The book of why*. Basic Books, New York.
- Ratcliffe N, Schmitt S, & Whiffin M. 2005. Sink or swim? Viability of a black-tailed godwit population in relation to flooding. *Journal of Applied Ecology*, **42**(5), 834–843.
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Rubin R. D. 1980. Discussion of “Randomization Analysis of Experimental Data in the Fisher Randomization Test,” by D. Basu. *Journal of the American Statistical Association*, **75**: 591–593
- Rubin D. B. 1980. Randomization analysis of experimental data: the fisher randomization test. *Journal of the American Statistical Association*, **75**:591–593.
- Schleicher J, Eklund J, Barnes M, Geldmann J, Oldekop J. A, & Jones J. P. G. 2019. Statistical matching for conservation science. *Conservation Biology*, **34**: 538–549.
- Sills E. O, de Sassi C, Jagger P, Lawlor K, Miteva D. A, Pattanayak S. K, & Sunderlin W. D. 2017. Building the evidence base for REDD+: study design and methods for evaluating the impacts of conservation interventions on local well-being. *Glob. Environ. Change*, **43**: 148–160
- Smart J, Wotton S. R, Dillon I. A, Cooke A. I, Diack I, Drewitt A. L, Grice P. V, & Gregory R. D. 2014. Synergies between site protection and agri-environment schemes for the conservation of waders on lowland wet grasslands. *Ibis*, **156**(3), 576–590.
- Smart M, & Coutts K. 2004. Footdrain management to enhance habitat for breeding waders on lowland wet grassland at Buckenham and Cantley Marshes, Mid-Yare RSPB Reserve. *Conservation Evidence*, **1**: 16–19.
- Sonter L. J, Barnes M, Matthews J. W, & Maron M. 2019. Quantifying habitat losses and gains made by U.S. Species Conservation Banks to improve compensation policies and avoid perverse outcomes. *Conservation Letters*:1–9.
- Stuart E. A. 2010. Matching methods for causal inference: a review and a look forward. *Statistical Science*, **25**(1), 1–21.
- Terraube J, Van doninck J, Helle P, & Cabeza M. 2020. Assessing the effectiveness of a national protected area network for carnivore conservation. *Nature Communications*, **11**:1–9.
- Textor J, van der Zander B, Gilthorpe M. S, Liškiewicz M, & Ellison G. T. 2016. Robust causal inference using directed acyclic graphs: The R package “dagitty”. *International Journal of Epidemiology*, **45**(6), 1887–1894.
- Verhulst J, Kleijn D, & Berendse F. 2007. Direct and indirect effects of the most widely implemented Dutch agri-environment schemes on breeding waders. *Journal of Applied Ecology*, **44**: 70–80.
- Walker L. K, Morris A. J, Cristinacce A, Dadam D, Grice P. V, & Peach W. J. 2018. Effects of higher-tier agri-environment scheme on the abundance of priority farmland birds. *Animal Conservation*, **21**:183–192.
- Wickham, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software*, **4**(43), <https://doi.org/10.21105/joss.01686>
- Wiik E, d’Annunzio R, Pynegar E, Crespo D, Asquith N, & Jones J. P. G. 2019. Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conservation Science and Practice*, **1**:1–11.
- Wilson A. M, Ausden M, & Milsom T. P. 2004. Changes in breeding wader populations on lowland wet grasslands in England and Wales: Causes and potential solutions. *Ibis*, **146**(suppl. 2), 32–40.
- Wilson A, Pendlebury C, & Vickery J. 2005. Changes in lowland wet grassland breeding wader number: the influence of site designation. Research report 365. British Trust for Ornithology, Thetford.
- Wood B. 1992. Yellow Wagtail *Motacilla flava* migration from West Africa to Europe: pointers towards a conservation strategy for migrants on passage. *Ibis*, **134**(suppl 1), 66–76.

