Comparative effectiveness research: the missing link in conservation

Rebecca K. Smith[1]*, Lynn V. Dicks[1], Roger Mitchell[2] and William J. Sutherland[1]
[1]*Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK*
[2]*Amphibian and Reptile Conservation, 655a Christchurch Road, Boscombe, BH1 4AP, UK*

SUMMARY

This editorial highlights the deficit of studies that directly compare different conservation interventions for the same threat. Most studies test a single intervention (86% in Conservation Evidence), comparing it against a control that lacks the intervention. Such studies can provide evidence that a particular intervention is effective, but do not inform a practitioner whether that intervention is the best option relative to others. Comparing results from different studies is difficult, as outcomes depend on factors such as the site, species and method of measurement. We suggest that a key step to understanding the effectiveness of conservation interventions is to compare different interventions in the same context within studies. If widely adopted this could transform global conservation practice. We provide some guidance on how to design and conduct comparative studies.

DEFICIENCY OF COMPARATIVE STUDIES

After ten years of the online journal Conservation Evidence, we focus this editorial on what we consider a major gap in applied conservation research: the paucity of studies that directly compare single specific interventions against one another. Here we explain why more comparative studies are required and provide some guidance on how to design and conduct them.

Most studies testing interventions examine a single intervention. We found that of the 95 papers published in Conservation Evidence during 2009-2013, a large proportion (86%) studied a single intervention. The intervention is usually compared against a control that lacks the conservation management treatment. We suggest that although such studies are useful, this widespread experimental approach has a number of shortcomings. Consider, for example, a study of the effect of treating an invasive plant with a herbicide. The herbicide is very likely to have more impact on the plant than the control treatment, which involves not applying a herbicide. However, this result does not indicate whether the selected herbicide is the best option compared to a number of other possible intervention treatments, such as alternative herbicides, hand weeding or even applying the same herbicide at different times, in different formulations or by various methods. For this, one has to make comparisons with other studies which are very likely to have been carried out in different circumstances and therefore unsatisfactory.

As another example, many studies test the effects of providing nest boxes on bird reproductive success or population status (see, for example, several of the papers in our virtual online Bird Management issue http://www.conservationevidence.com/collection/19#spec). These studies are a very useful contribution to the literature on bird conservation. There are cases where the use of nest boxes led to measurable increases in reproductive success. Those studies provide evidence for practitioners that nest boxes should be one of the options they consider in a package of conservation measures for a particular bird species or community. However, such studies do not inform the practitioner whether nest boxes are the best option relative to other options with different costs. For example, how do the effects of nest boxes compare with the effects of providing supplementary food? Again, comparison across studies would be needed but would not necessarily be comparing like with like.

Comparing results across different conservation intervention experiments is difficult, although often necessary for pragmatic decision-making. The magnitude of the change resulting from implementing an intervention is likely to depend on the site, species, intervention and method of measurement. Consequently, environmental evidence is likely to be more variable than evidence used in medicine. There is greater heterogeneity between ecosystems and the communities of species within them than between humans in their different communities. To illustrate this, consider a pair of similar medical trials. One showed that 5% of patients recovered from a disease with treatment X carried out in the UK. The other showed that 55% of patients in the same demographic group recovered from the same disease with treatment Y in Australia. It is quite reasonable to infer from these two studies that treatment Y is likely to be more effective. Now consider a similar case of two studies in conservation. In one study, nest box design A had a 5% occupancy rate in the UK. In the other, nest box design B had a 55% occupancy rate in Australia. Many conservationists would consider that results are as likely to reflect differences in the species involved or their interaction with the surrounding environments, as they are to reflect a genuine difference in effectiveness of the two nest box designs. From the two studies, it is not reasonable to assume nest box B is a better design. As a real example, a collation of evidence by Dicks *et al*. (2010) on the effectiveness of bumblebee nest boxes found that seven trials between 1912 and 1978 in the USA, Canada and UK and one more recent trial in the USA, showed overall uptake rates of between 10% and 48%. However, three trials between 1989 and 2009 in the UK showed very low uptake rates (0-2.5%) of various nest box designs. The more recent UK studies used a narrower range of nest box designs, with a majority of nest boxes placed at the surface and no underground nest boxes. Does this mean the bumblebee nest box designs used in the American experiments are better? It could also indicate that bumblebees are now at lower densities in the UK and so less likely to find artificial nest boxes, or that natural nest sites are more widely available

* To whom correspondence should be addressed: r.k.smith@zoo.cam.ac.uk

ISSN 1758-2067

for British bumblebees, reducing their need for artificial sites. Similar problems with comparing different treatments in different locations would occur for a wide range of types of intervention, including the effectiveness of habitat manipulations, and social or educational interventions.

A possible solution is to compare different treatments for the same threat within a single study – an approach that has been called comparative effectiveness research. In the USA, there is strong policy support for a move towards comparative effectiveness research in medicine (Olfson & Marcus 2013). Traditionally the most frequent type of research into the effectiveness of medical treatments was placebo-controlled trials, in which a specific treatment was compared to a placebo, i.e. a control with no treatment. Olfson and Marcus (2013) document a long-term decline in effect sizes in placebo-controlled trials of medical treatments. They argue this makes it increasingly difficult to compare different treatments across studies. When effect sizes are small, other features of treatments, such as cost, social acceptability and side effects, become important to decision-making. These are best tested in comparative research because they are more likely to depend on context. The questions usually asked in "head-to-head" comparison trials are: which intervention works better and at what relative cost? They are designed to improve the link between research and everyday practice, as well as enabling the delivery of more cost effective treatments. For example, a double-blind randomized crossover trial in 156 adult patients suffering from migraine found that compared with a placebo, a single oral dose of either of two drugs (50 or 100 mg diclofenac-potassium or 100 mg sumatriptan) was an effective treatment for migraine, but that one of the drugs was faster-acting and had reduced side effects compared to the other (The Diclofenac-K/Sumatriptan Migraine Study Group 1999). As another example, studies had found that cognitive behaviour therapy (CBT) and graded exercise therapy (GET) can be effective treatments for chronic fatigue syndrome, but patients' organisations had reported that these treatments could be harmful and so they prefer adaptive pacing therapy (APT) and specialist medical care (SMC). A comparative randomised trial assessed the effectiveness and safety of all four treatments. The trial in 641 patients found that CBT and GET in addition to SMC moderately improved outcomes for chronic fatigue syndrome (compared to SMC alone) with no detrimental effects, but that APT was not effective (White *et al*. 2011).

We therefore suggest that, just like in medicine, a key step to understanding the effectiveness of conservation interventions is to compare different treatments in the same context within studies. For example, of the 11 bumblebee nest box studies cited previously, four directly compared nest box designs and found that wooden surface nest boxes were occupied less frequently than underground, false underground or aerial boxes (Dicks *et al*. 2010). A collation of evidence for amphibian conservation found that the majority of studies testing the effect of prescribed forest fires did not find any overall effect on amphibians, suggesting that prescribed burns are not detrimental, but not usually beneficial to amphibians (Smith & Sutherland 2014). However, only 29% of the 14 studies compared different fire frequencies (2-7 year cycles) or season of burn. Both of these variables changed the effect of prescribed fire on amphibian abundance. One study, for example, found that salamander numbers declined following spring, but not autumn or winter burns (Broadman 2010). If such management details can change the overall outcome, then it is not correct to compare across studies that vary in these details of the intervention. Table 1 provides some more examples of comparative studies of conservation interventions in the literature.

Table 1. Examples of studies that have compared interventions.

Target	**Intervention**	**Result**	**Reference**
Bats	Five different bat box designs for natterer's bat *Myotis nattereri* and brown long-eared bat *Plecotus auritus*.	Occupancy bias towards two box types, which was likely to have been influenced by seasonal bird competition.	Dodds & Bilston (2013)
Limestone grassland	Delay cattle grazing of limestone grassland from May to July.	A delay in grazing start date resulted in few pronounced changes to the vegetation.	Costley (2013)
Plants	Reintroduction of seeds or seedlings of star cactus *Astrophytum asterias* in the spring or autumn.	Less than 4% of seeds produced seedlings, but 55% of spring planted and 73% of autumn planted seedlings survived for at least one year.	Birnbaum *et al.* (2011)
Plants	*Premna serratifolia* plantlets were regenerated through shoot tip explants using four different media (MS, SH, Y3 and B5) and three different anti-oxidants.	Shoot multiplication was greatest using one of the four mediums (MS) supplemented with two specific hormones (a cytokinin and auxin) and activated charcoal between November and March. Best rooting was achieved from the medium supplemented with a different auxin.	Chinnappan *et al.* (2011)
Invasive/problem plants	Three control methods for giant reed *Arundo donax*.	Most effective control method was cutting and removal of reed stems followed by two glyphosate-based foliar herbicide applications (one in May and October).	Silva *et al.* (2011)
Heathland vegetation	Eight combinations of treatments to reinstate heathland vegetation (additions of peat dust, heath mulch, geojute, fertiliser and grass seed).	A combination of peat dust, heath mulch and geojute gave the best results with 80% cover of vascular plants (70% heather *Calluna vulgaris*).	Robertson (2010)

ISSN 1758-2067

Birds	Four establishment methods for skylark plots: undrilled or sprayed with herbicide in December, January or February.	Undrilled plots had greater vegetation cover and so were more suitable for skylarks than sprayed plots. Spraying in December resulted in greater vegetation cover and so was better than spraying in January or February.	Dillon *et al*. (2009)
Birds	Field margin plots established using one of three seed mixes and managed using one of three treatments.	Margin management affected use more than seed mix planted. In summer, bird densities were higher on disturbed plots and plots treated with herbicides than on cut plots. In winter, there were twice as many birds on cut compared to uncut margins.	Henderson *et al.* (2007)
Amphibians	Four different culvert designs for under roads.	Green frogs *Lithobates clamitans* used tunnels with soil or gravel more than concrete or PVC substrates, but showed no preference for specific width or length tunnel. Leopard frogs *Rana pipiens* showed no substrate preference, but used wider, shorter tunnels more often. Both species preferred tunnels with greater light permeability.	Woltz *et al.* (2008)
Bees	Prescribed burning and/or mechanical shrub control in forest.	More bees and bee species in plots following both prescribed burning and shrub control, compared to those with only shrub control or fires, or no fire control.	Campbell *et al.* (2007)
Mammals	Riparian field margins were mown or unmown and then food and/or cover was or was not added.	Field voles *Microtus agrestis* were most frequently found in control plots where vegetation was not mown and no supplementary food or cover was added, in both narrow and wide field margins.	Yletyinen & Norrdahl (2008)

DESIGNING STUDIES THAT COMPARE INTERVENTIONS

In consideration of the above, we offer the following guidance when designing studies that compare different interventions.

Firstly, all treatments should be compared directly with a control. A control is necessary to determine whether any measured changes, such as increases in population size, are a result of the treatments rather than other factors such as changes in weather or resource availability.

Secondly, the range of effect of each treatment (including the control) and the amount of variation in their outcomes, should be clearly reported. Often, in ecological studies, data analysis uses analysis of variance or more complex statistical modelling and results are reported as the relative significance of treatments in explaining variability in the measured effect. However, the results of pairwise comparisons between each individual treatment and the control, or each treatment and each other treatment are often not reported. This can make it difficult to assess whether one treatment is significantly better than another. For example, in a site comparison study on UK farmland, Pywell *et al.* (2005) measured numbers of bumblebees on naturally regenerated field margins, control margins sown with crops, and margins sown with a wildflower mix. The naturally regenerated field margins supported fewer bumblebees (18 individuals and 2.7 species/100 m on average) than margins sown with a wild flower seed mixture (29 bumblebees, 3.0 species/100 m; Dicks *et al.* 2010). The two treatments were not directly compared in the analysis, so the reported average numbers for each treatment are really important. Decision makers would have to run their own statistical analysis to decide with confidence whether one option was better.

Thirdly, for studies in which species are given a choice, such as between different nest box designs or habitat management options, it is important to be aware that the proximity of treatments or designs could influence the conclusions. For example, if one nest box design was selected by the majority of birds when compared to three other designs in neighbouring trees, it suggests that it is the preferred design in that situation for those species. It does not necessarily mean that the other designs would not be used if they were the only box available. Replicated controlled experiments of field margin management options for pollinators provide another example. Carvell *et al.* (2007) found that naturally regenerated field margins supported a greater number and diversity of foraging bumblebees than cropped margins (including conservation headlands), but only in the first year of the study. In the second and third years, bumblebee numbers on naturally regenerated margins were not significantly different from cropped treatments, but this may have been because the bees were more attracted to areas on the same field margin densely sown with wild flowers, as a comparative treatment in the experiment. Therefore, as with nest boxes, the proximity of a more attractive option can skew the results. This becomes important if the less attractive option is much cheaper, or easier to implement, as it is in the field margin management example.

Finally, sample sizes are particularly important to the experimental design of comparative studies. If all the interventions being compared are known to work and the expected difference between them is relatively small, the sample size in each treatment group will need to be larger to

ISSN 1758-2067

detect a difference between treatments. Statisticians sometimes say that 30 is a good sample size to aim for, so if you are testing nest boxes you would need 30 of each design. Achieving such large sample sizes is often impossible in conservation research, especially if the interventions need to be carried out at separate sites, on threatened individuals or in protected areas. The best strategy is to replicate the experiment as many times as is feasible, given time and cost constraints.

Conservation Evidence continues to welcome papers that compare individual interventions against controls, but this year we are particularly seeking submissions that carry out comparisons. In Table 2 we identify some areas in which further comparison studies for interventions would be especially helpful. We look forward to your submissions.

ACKNOWLEDGEMENTS

RKS and WJS are funded by Arcadia. LVD is funded by the Natural Environment Research Council (Grant codes NE/K015419/1, NE/J500665/1).

Table 2. Some possible options for comparisons.

Target	Example of possible comparisons
Invasive plants	Cutting, pulling, mulching, flame-gun or grazing versus different herbicides; treatment methods; times of year.
Invasive animals	Trapping versus poisoning or shooting, at different times of year.
Reintroduction or assisted colonisation of animals or plants	Translocation of wild individuals versus release of head-started, captive-bred, propagated or *in vitro* individuals; release of different life stages; season of release; single, multiple or sequential releases of individuals per unit area; with or without post-translocation management.
Habitat management	Cutting meadows on different dates; different width buffer strips/field margins; different timing or frequency for ditch clearance, pond dredging or hedge management; planting versus natural re-colonization of created habitats; different varieties of wild flower mix; season or frequency of prescribed fires; different grazing animals or grazing regimes; soil treatments; water level control.
Species management	Supplementary feeding; methods of predator control; methods of hand pollination, artificial incubation or *in vitro* fertilisation.
Shelter	Different designs or locations of nest boxes, refugia or hibernacula.
Infrastructure	Different designs of under or overpasses for roads, types of nesting platforms, exclosures or enclosures.
Disturbance	Different designs of signs aimed at adjusting routes taken by humans or modifying other potentially damaging behaviour.
Social	Establishing different land ownership rights.
Education	Classroom or on-line versus field training.

REFERENCES

Birnbaum, S.J., Poole, J.M. & Williamson, P.S. (2011) Reintroduction of star cactus *Astrophytum asterias* by seed sowing and seedling transplanting, Las Estrellas Preserve, Texas, USA. *Conservation Evidence*, 8, 43-52.

Brodman, R. (2010) The importance of natural history, landscape factors, and management practices in conserving pond-breeding salamander diversity. *Herpetological Conservation and Biology*, 5, 501-514.

Campbell, J.W., Hanula, J.L. & Waldrop, T.A. (2007) Effects of prescribed fire and fire surrogates on floral visiting insects of the blue ridge province in North Carolina. *Biological Conservation*, 134, 393-404.

Carvell, C., Meek, W.R., Pywell, R.F., Goulson, D. & Nowakowski, M. (2007) Comparing the efficacy of agri-environment schemes to enhance bumble bee abundance and diversity on arable field margins. *Journal of Applied Ecology*, 44, 29-40.

Chinnappan, R.S., Ruthar, N. & Sethu, S.S. (2011) Rapid in vitro propagation of *Premna serratifolia*, a medicinally important declining shrub, India. *Conservation Evidence*, 8, 66-73.

Costley, J. (2013) Impact of change in start date of grazing on limestone grassland communities, Deep Dale, Derbyshire, UK. *Conservation Evidence*, 10, 77-79.

Dicks, L.V. Showler, D.A. & Sutherland, W.J. (2010) *Bee Conservation: Evidence for the Effectiveness of Interventions.* Pelagic Publishing.

Dillon, I.A., Morris, A.J., Bailey, C.M. & Uney, G. (2009) Assessing the vegetation response to differing establishment methods of Skylark Plots in winter wheat at Grange Farm, Cambridgeshire, England. *Conservation Evidence*, 6, 89-97.

Dodds, M. & Bilston, H. (2013) A comparison of different bat box types by bat occupancy in deciduous woodland, Buckinghamshire, UK. *Conservation Evidence*, 10, 24-28.

Henderson, I.G., Morris, A.J., Westbury, D.B., Woodcock, B.A., Potts, S.G., Ramsay, A. & Coombes, R. (2007) Effects of field margin management on bird distributions around cereal fields. *Aspects of Applied Biology*, 81, 53-60.

Olfson, M. & Marcus, S.C. (2013) Decline in placebo-controlled trial results suggests new directions for comparative effectiveness research. *Health Affairs*, 32, 1116-1125.

Pywell, R.F., Warman, E.A., Carvell, C., Sparks, T.H., Dicks, L.V., Bennett, D., Wright, A., Critchley, C.N.R. & Sherwood, A. (2005) Providing foraging resources for bumblebees in intensively farmed landscapes. *Biological Conservation*, 121, 479-494.

ISSN 1758-2067

Robertson, J. (2010) Identifying effective treatments to reinstate heath vegetation on commercially extracted peatlands at Hobbister RSPB Reserve, Orkney, Scotland. *Conservation Evidence*, 7, 123-129.

Silva C.M.N., Silva L., Oliveira N., Geraldes P. & Hervías S. (2011) Control of giant reed Arundo donax on Vila Franca do Campo Islet, Azores, Portugal. *Conservation Evidence*, 8, 93-99.

Smith, R.K. & Sutherland, W.J. (2014) *Amphibian conservation: evidence for the effectiveness of interventions*. Pelagic Publishing.

The Diclofenac-K/Sumatriptan Migraine Study Group (1999) Acute treatment of migraine attacks: efficacy and safety of a nonsteroidal anti-inflammatory drug, diclofenac-potassium, in comparison to oral sumatriptan and placebo. *Cephalalgia*, 19, 232-40.

White, P.D., Goldsmith, K.A., Johnson, A.L., Potts, L., Walwyn, R., DeCesare, J.C., Baber, H.L., Burgess, M., Clark, L.V., Cox, D.L., Bavinton, J., Angus, B.J., Murphy, G., Murphy, M. , O'Dowd, H., Wilks, D., McCrone, P., Chalder, T. & Sharpe, M. (2011) Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *The Lancet*, 377, 823 – 836.

Woltz, H.W., Gibbs, J.P. & Ducey, P.K. (2008) Road crossing structures for amphibians and reptiles: informing design through behavioral analysis. *Biological Conservation*, 141, 2745-2750.

Yletyinen, S. & Norrdahl, K. (2008) Habitat use of field voles (*Microtus agrestis)* in wide and narrow buffer zones. *Agriculture, Ecosystems & Environment, 123, 194-200*.

ISSN 1758-2067