Research article

From ignorance to evidence? The use of programme evaluation in conservation: Evidence from a Delphi survey of conservation experts

Hannah Fay Curzon [a], Andreas Kontoleon [b], *

[a] Environment Department, University of York, UK
[b] Department of Land Economy, University of Cambridge, UK

ABSTRACT

Persistent gaps in the evidence base regarding the performance of conservation policies has put pressure on the conservation policy field to adopt 'best practice' programme evaluation methods. These are methods that account for the counterfactual and are able to attribute causality between a conservation policy and specific observable environmental and social impacts. Despite this pressure, use of such methods continues to be rare. This paper uses the Delphi technique to provide the first systematic assessment of the reasons behind the apparent hesitation of conservation practitioners to adopt rigorous policy impact evaluation methods. The Delphi study consisted of two online questionnaires conducted on conservation policy experts. The results presented confirm that the use of rigorous impact evaluation methods in conservation is still very limited but this, crucially, is not because conservationists are ignorant of these methods or their advantages. In fact, considerable effort is being made to develop and improve evidence standards but these efforts have largely been thwarted by large financial and time related constraints that mean even elementary evaluations are hard to achieve. The results from this Delphi study allow us to provide more realistic recommendations on how impact evaluation studies can be more widely embraced and implemented in conservation practice.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Conservation practitioners and policy-makers need credible information regarding the performance of conservation interventions in order to ensure that scarce funds are not wasted on ineffective policies (Sutherland et al., 2004; Stem et al., 2005; Bottrill et al., 2011). There have been numerous calls for the conservation policy field to adopt 'best practice' or 'rigorous' programme evaluation methods (e.g. Ferraro and Pattanayak, 2006; Ferraro, 2009). These methods focus on the use of experimental and quasi-experimental evaluation designs that can be used to credibly measure 'counterfactual' outcomes. It is argued that establishing this counterfactual is critical to being able to unambiguously isolate the impacts of policy interventions so as to get an unbiased estimate of a programme's performance (Berry et al., 2012).

Despite these calls, there are still large gaps in the hard fact evidence base regarding the performance of conservation policies.

Several reviews have documented the paucity of formal evaluations studies on conservation policies using experimental and quasi-experimental methods (e.g., Pattanayak et al., 2010; Blackman, 2012; Miteva et al., 2012; Adhikari and Agrawal, 2013; Roe, Greig-gran & Mohammed, 2013; Zheng et al., 2013; Alcorn, 2014; Cowling, 2014; Samii et al., 2014). This body of work has found that though monitoring and evaluation data (which only documents trends and changes in variables) is abundant and routinely collected, formal evaluation studies (which identify the causal links between a policy and specific conservation outcomes) are highly scarce.

Although the inherent financial, temporal, logistical, and sometimes ethical, challenges of conducting rigorous evaluations have been discussed in the literature, it is still conjectured that one of the main reasons for the limited use of policy evaluation methods is not through a lack of opportunity and resources but, instead, due to a lack of awareness, understanding and appreciation of the need for counterfactual thinking within the conservation policy field (e.g., Ferraro, 2006; 2009). Such assertions are, however, largely unsupported by any kind of formative assessment of the rationale behind conservation evaluation decisions in practice

and thus risk being inaccurate and out-of-date. Arguably, in order to obtain a more comprehensive understanding of the underlying reasons for the documented gaps in the evidence base, it is necessary to draw on the knowledge and experience of the actual decision-makers and practitioners working in the conservation policy field. The present study aims to fill this research gap by being the first to systematically ascertain information from experts working in conservation as to their stance with respect to the usefulness, practicality, desirability and prospects of using formal policy evaluation methods. For this purpose, our study uses the Delphi technique, an iterative survey-based research method, which allows for a systematic assessment of the conservation sector's actual knowledge, appreciation, and experience with such methods. As a result, our study will be able to more critically evaluate the commonly made assertions found in several past reviews that the conservation sector is averse to impact evaluation. Lastly, the study will provide policy relevant information on how to more rigorously determine the needs, opportunities and barriers to using 'best practice' methods to evaluate the impact of conservation interventions. These findings could significantly contribute to improving our understanding of the conservation sector's approach to evaluation and how far conservation organisations represented in this study are thinking counterfactually, thus providing a more accurate and informed assessment of the real reasons for the gaps in the evidence base.

This paper proceeds as follows: Section 2 provides some of the common critical assertions found in review literature on the paucity of impact evaluation work in the conservation field. This is followed by the rationale for this study and the specific research questions we address. Section 3 outlines the research methodology as applied in the Delphi study. The results of the study are then presented in Section 4 and are discussed and summarised in Section 5. The survey instruments that were used appear as Supplementary Materials (Appendices SM1-SM4). More details specifically on the methods used can also be found in a SM1 (Technical Annex).

2. Impact evaluation in conservation policy

2.1. The impact evaluation revolution in science

Programme evaluation is fundamentally a process of making inferences about an unobserved counterfactual outcome, i.e., what would have happened in the absence of the intervention, programme or policy. (Ferraro and Pattanayak, 2006). Without this 'counterfactual analysis' it is impossible to know how far impacts are the result of the intervention and not due to other confounding factors or biases (White, 2006; Khandker et al., 2010). However, as the counterfactual cannot be observed, the main challenge of impact evaluation is to find or construct an appropriate counterfactual in the light of the missing data.

Two common approaches to evaluation that have been used in the conservation policy field are before-after and with-without comparisons, i.e., comparisons of outcomes before and after an intervention and comparisons of outcomes in areas with and without exposure to the intervention. As before-after comparisons do not control for other time varying factors, and with-without comparisons do not control for selection bias, both methods lead to biased estimates of impacts (Khandker et al., 2010). More rigorous approaches that can be used to solve the problem of selection bias and establish a credible counterfactual broadly fall into two categories (Khandker et al., 2010). The first relies on data obtained from randomised controlled evaluations or trials (i.e. RCTs) which randomly assign study subjects into treatment and control groups. The data is collected before and after the policy leading to

the so-called Before-After-Control-Impact (or BACI) design which is widely regarded as the 'gold-standard' in programme evaluation (Frondel and Schmidt, 2005; Duflo et al., 2008; Greenstone and Gayer, 2009). By randomly allocating treatment and control groups across eligible sample units, units that do not receive the treatment will be a valid comparison group for those that did since there should be no systematic differences between their characteristics (Rossi et al., 2003).

When randomisation of the treatment is not possible, the second-best option is to rely on observational data of two samples of subjects, one that has been exposed to a policy (or treatment) and others that have not. Then practitioners use quasi-experimental statistical methods (such as propensity score matching and difference-in difference estimation) to create comparison groups that are valid under a set of underlying assumptions about the nature of potential selection bias in programme targeting and participation (Khandker et al., 2010). While these econometric methods are well-developed and firmly grounded in theory and statistics, the identifying assumptions are not always directly testable, and the validity of any particular study depends instead on how convincing the assumptions appear (Duflo et al., 2008).

The call for the use of formal impact evaluation methods that address the issue of the counterfactual is part of a broader movement towards evidence-based policy making (Gertler et al., 2011) that was first experienced in medicine in the second half of the twentieth century (Pullin et al., 2004). The resulting paradigm shift from 'experience-based' to 'evidence-based' practice that emphasized the use of clinical experiments and systematic reviews (Pullin and Knight, 2001; Stevens et al., 2001) completely revolutionised medical practice. This 'effectiveness revolution' became the archetypal method for evaluation and primary research and spread to other social policy fields such as public health, education and international development who started to build randomised evaluations into their programmes recognising the need for convincing and comprehensive evidence that could be used to inform policy making and improve the allocation on government resources (Pullin and Knight, 2004; Pullin et al., 2004; Gertler et al., 2011).

2.2. Impact evaluation in conservation policy

In contrast, the field of conservation policy did not experience the same 'effectiveness revolution' and even by the beginning of the twenty-first century the evaluation of conservation programmes continued to be rare (Kleiman et al., 2000). One of the main conclusions stemming from a global review of the evidence base known as the 'Millennium Ecosystem Assessment,' was that '[f]ew well-designed empirical analyses assess even the most common biodiversity conservation measures' (MEA, 2005, p.122). Indeed, it was widely acknowledged at the time that conservation was still largely an experience-based practice that depended on intuition and anecdote to guide the design of conservation investments as opposed to empirical evaluations (Kleiman et al., 2000; Pullin and Knight, 2001; Salafsky et al., 2002; Salafsky and Margoluis, 2003; Pullin et al., 2004; Sutherland et al., 2004). While these studies advocated the need for evidence-based conservation, interest in impact evaluation *per se* did not emerge in the conservation policy field until the mid to late 2000s (Frondel and Schmidt, 2005; Ferraro and Pattanak, 2006; Ferraro, 2009; Greenstone and Gayer, 2009; Pattanayak et al., 2010). As a result, the amount of literature on environmental impact evaluation is still limited.

Ferraro and Pattanayak (2006) was one of the first papers to call for rigorous empirical evaluation of conservation polices. The authors argued that while conservation projects had increasingly focused on 'monitoring and evaluation' since the 1990s, 'rigorous