SM1. Technical Annex

1. Methods

1.1. Selection of the expert panel

Unlike traditional surveys, a Delphi does not depend on selecting a statistical sample that attempts to be representative of any population. Instead, it requires a sample of qualified experts that have a deep understanding of the issues. Subsequently, rigorous selection of the panel of experts is one of the most critical requirements of any Delphi study (Okoli & Pawlowski, 2004).

The experts involved in the Delphi process were primarily identified in three ways: (1) through involvement in projects run by the Cambridge Conservation Initiative (CCI), which is a unique collaboration between the University of Cambridge and many of the largest major biodiversity conservation organisations in the world; (2) through membership of the BIOECON (Biodiversity and Economics for Conservation) network, an assembly of individuals from a wide range of international research institutions and policy organisations working on the design and implementation of economic incentives for biodiversity conservation; (3) by browsing staff profiles on the websites of major international conservation NGOs and government agencies, particularly organisations that were members of the Cambridge Conservation Forum1. Experts were also identified through participation in a highly relevant workshop at the Society for Conservation Biology's (SCB) 2013 International Congress for Conservation Biology (ICCB), an international forum for addressing conservation issues. Initially the search produced a list of approximately 1600 names.

To be selected for the expert panel participants had to work for a conservation organisation as policy advisor (designing and producing conservation policy interventions) or, have experience working as a conservation practitioner (implementing and evaluating conservation policy interventions on the ground). An additional criterion was to select conservation experts with a broad range of expertise and from a large number of different organisations to

---

1 CCF is an active network of conservation-related member organisations and institutions based in and around Cambridge. Over 50 conservation organisations are currently members of the CCF (CCF, 2013).
2 SCB is a global community of conservation professionals. The Society's membership comprises a wide range of people including resource managers, educators, government and private conservation workers and students (SCB, 2013).
try and balance any possible sector bias in opinions. Subsequently, experts were evenly spread in terms of their interests and were affiliated with both conservation NGOs, NPOs and government agencies, private consultancies as well key research institutions, such as the World Wildlife Foundation (WWF), The Rainforest Alliance, The Department for Environment, Food and Rural Affairs (DEFRA), The Royal Society for the Protection of Birds (RSPB) and Natural England to name some examples. Using web searches to obtain biographical information, approximately 300 individuals, from the initial list of 1600 names, were identified as meeting the criteria and the necessary contact information could be sourced. Based on their credentials and the amount of information found, individuals were then sorted in order of preference based on their level or experience and expertise. The most preferred candidates were individuals in more senior positions, such as project managers, and individuals working in M&E divisions or those known to have specialist knowledge in conservation effectiveness or evaluation.

As mentioned, Delphi group size does not depend on statistical power but rather on group dynamics for arriving at consensus among experts, thus Delphi panel size requirements are modest and most commonly range from just 10 to 25 participants (Diamond et al., 2014). To provide representative information some studies have employed over 50 participants, e.g., Alexander & Kroposki, 1999. While experiments have found that there is reduction in group error with increasing group size, after a certain threshold the benefits of adding more and more participants quickly diminishes (Adler & Ziglio, 1996) and can lead to data handling and analysis difficulties, particularly if the data collected in the first round is mainly of a qualitative nature (Hasson, Keeney & McKenna, 2000). Based on the literature, this study aimed to recruit at least 20 experts to serve on the panel. As at least ten percent of the sample can normally be expected to respond, the top 200 experts were invited to participate in the study.

1.2. Structure of the Delphi process

The Delphi study which was entirely managed via email, consisted of two online questionnaires with both open and closed ended questions. While some Delphi studies only terminate once a consensus is reached, most Delphi studies run for a prescribed number of rounds (Diamond et al., 2014). In this case, an a priori decision was made to have two rounds of questions due to the time available and to avoid the risk of sample fatigue; maintaining participant interest in the study is known to become increasingly difficult after each
The questionnaires were created using powerful online survey software called 'Qualtrics', crucially, allows the researcher to generate personalised links (a necessary requirement for a Delphi). Once developed, the initial questionnaire was pre-tested by one of the Senior Programme Officers at the United Nations' Environment Programme-World Conservation Monitoring Centre (UNEP-WCMC). The idea was to use a member of the intended audience to test the questionnaire (and to provide feedback) in order to ensure that the questions and instructions would be understood in the manner they were intended by the members of the Delphi panel.

Each of the 200 prospective panellists were individually invited to participate in the study and were provided with their own personal link to the questionnaire that could only be used from their email domain. While it was necessary to provide each person with their own personal link in order to be able to keep track of who had responded, it was thought that the personal touch would also re-enforce the fact that the person has been specifically 'chosen' to serve as a member of the expert panel and thus hopefully increase the response rate. Importantly, invitations were also used to inform the prospective panellists of the study's aims, the structure of the Delphi process, what they would be required to do, how much time they would need to be able to commit in order to be able to properly participate in the study, what the information would be used for and, finally, assurance of their anonymity. For this study, participants were asked to commit to completing two 15-20 minute questionnaires. As panellists ultimately self-select into the study, it was thought that by providing an honest appraisal of the time commitment required that this would help sustain the response rate for the second round of questions. Participants had eight days and nine days (with a week in between) to respond to the first and second questionnaire, respectively. The online questionnaires were designed so that the participant could start the questionnaire and return to it at a later time. The second questionnaire was only sent to participants that had completed the first questionnaire. Follow-up reminders were also emailed out to people who had not responded to the questionnaires after three days and then again two days before the deadline for responses (useless they had officially opted out of the study) for both rounds to try and further enhance response rates.
The first questionnaire was accompanied by a two-page document (sent as an email attachment) which introduced respondents to the concept of the 'counterfactual' and outlined different approaches to evaluating conservation interventions (see Supplementary Material). In this case, eco-certification was used as an example. Particular attention was given to the definition of experiments and quasi-experiments in comparison to a simple 'before-after' or 'with-without' approach.

### 1.3 Delphi survey: round one

The overall aim of the Delphi was to assess the panel's perspective on the needs, opportunities and barriers to using experimental and quasi-experimental methods (hereafter referred to as rigorous evaluation methods) when evaluating the impacts and effectiveness of conservation interventions.

To better address the study's research questions, the first round (R1) of questions was structured into two parts. Questions in part one of R1 were designed to address research questions one and two, i.e., the importance of rigorous evaluation methods as well as the panel's perspective on what they considered to be 'best practice' in conservation evaluation. Questions in part two of R1 were designed to address research questions three, four and five, i.e., how far different evaluation methods are actually being used in practice, the panel's perspective on what they considered to be the most significant barriers to using rigorous evaluation methods and how far attempts were being made to improve evaluation and evidence standards in conservation policy.

Traditionally, the first round of a Delphi would begin with an open-ended set of questions designed to generate ideas and allow panellist complete freedom in their responses, a sort of 'brainstorming' exercise (Duffield, 1993; Angus et al., 2003). This round would be used to help identify issues which would then be rated (in terms of importance or acceptance) in the second round and then re-evaluated by the panel members in subsequent rounds until a consensus was reached (Jenkins & Smith, 1994). This study used a different approach, effectively combining the first two stages into one round; instead of allowing the experts complete freedom in their responses, the panel were given a pre-defined list of possible answers to choose from that had been devised from a review of papers from the published scientific literature that had either presented a definition of ‘best practice’ programme evaluation methods to define or put forward plausible examples of both incentives and
barriers for using rigorous evaluation methods in conservation such as Blackman & Rivera, 2010; Blackman, 2012; Miteva et al., 2012; Roe, Grieg-gran & Mohammed, 2013; Samii et al., 2014. Alternatively, respondents were asked to express their attitudes and opinions by selecting the most appropriate answer from a Likert scale. Likert-scales were introduced in the first round of questions (as opposed to the second round) on the basis that experts would already have a good knowledge of rigorous evaluation methods or would have gained a good enough understanding after having read the explanation and definitions provided in the briefing material. In order to ensure that all issues had been considered and the analysis was not limited by the questionnaire's design, where relevant, questions always had an 'Other' option whereby the experts could suggest a different answer to the options provided and explain their reason for doing so. Panellists were also given the opportunity to leave additional comments to further explain their answers or to indicate where the felt the questionnaire had not sufficiently addressed certain issues.

1.4. Delphi survey: round two
Preparation for the second questionnaire (R2) was devised based on the responses from R1 and was designed to provide a more detailed judgement on the issues therein. The idea is to largely re-iterate the questions asked in R1 but this time to include additional options for the experts to choose from based on the answers provided in R1 to give panellists a chance to re-evaluate their answers/opinions in light of the new options available. In accordance with the Delphi methodology, the panellists were provided with the results from R1 to aid the re-assessment procedure (Angus et al., 2003). The results were split into two documents: (1) a summary of the main results and; (2) a subsidiary document detailing all of the additional comments and suggestions left by the panel (quoted verbatim) for those who were interested. The opinions of the panel were in many cases expressed on a Likert scale which were coded (1-5 or 1-7) to produce a minimum, maximum and mean score for each. However, it was decided that the panellists would be able to interpret the information more quickly if the main results were presented as the percentage of respondents who had chosen each of the available answers (or the most suggested answers when there had been a long list of options). The percentages were summarised for the panellists in a series of tables and graphs.

While some of the questions from R1 were indeed repeated in R2, questions that had already reached clear consensus were omitted from R2. As there is no established definition of what
constitutes 'consensus' according to a review by Diamond and others (2014), it was decided that if 75% or more of the panellists agreed on one answer then a consensus had been reached, as this would represent the vast majority of the panel. In addition, where there was marked disagreement between panellists, some of the original questions were rephrased or presented differently to try and reveal a clearer, or more nuanced message. As many of the questions from R1 did not need to be asked again, a further modification in R2 was to include several new questions based on the panel's comments and feedback from R1. Two extra questions asking experts to describe their background (expertise) and their current role within their organisation were also added to the end of R2 (see Supplementary Material for both survey instruments used in R1 and R2 of the Delphi study).

2. References


