Social Marketing Quarterly
1-17

DOI: 10.1177/1524500417734806
journals.sagepub.com/home/smq
SAGE

Does It Work for Biodiversity? Experiences and Challenges in the Evaluation of Social Marketing Campaigns

Diogo Veríssimo[1,2], Annalisa Bianchessi[3], Alejandro Arrivillaga[1], Fel Ceasar Cadiz[4], Roquelito Mancao[4], and Kevin Green[1]

Abstract

There is a growing realization among conservationists that human behavior is the main driver of all key threats to biodiversity and the environment. This realization has led to an escalation of the efforts to influence human behavior toward the adoption of more sustainable alternatives, more recently through the use of social marketing theory and tools. However, these initiatives have traditionally suffered from a lack of robust impact evaluation, which limits not only accountability but also a practitioner's ability to learn and improve over time. We evaluated three social marketing campaigns conducted in the Philippines, which aimed at increasing the sustainability of local fisheries. To achieve this, we used the results not only from questionnaire surveys but also from biological and enforcement data. We found that although there is some evidence of impact around human behavior and perceptions of conservation results, those changes did not translate into biological outcomes during the 2-year time frame considered in this evaluation. We discuss many of the barriers to causal inference that still remain, particularly if causal links between outcomes and specific interventions are to be drawn, but also showcase how this current methodology can help us go further than the more basic approaches to impact evaluation commonly used. Lastly, we highlight a number of lessons learned from this experience in seeking a practical, ethical, and effective approach to impact evaluation.

Keywords

behavior change, community-based conservation, coral triangle, fisheries, impact evaluation, learning, marine conservation, marine-protected area, Philippines, social marketing

[1] Rare, Arlington, VA, USA
[2] Department of Geography and Environmental Engineering, Whitley School of Engineering, Johns Hopkins University, Baltimore, MD, USA
[3] Microinsurance Network, Luxembourg, Luxembourg
[4] Rare Philippines, Cebu City, Philippines

Corresponding Author:
Diogo Veríssimo, Department of Geography and Environmental Engineering, Whitley School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.
Email: verissimodiogo@gmail.com

Human behavior is the driver of all key threats to biodiversity (Balmford & Cowling, 2006). This realization has led those working to conserve biodiversity to increase their efforts to influence human behavior, most recently through the use of social marketing theory and tools (Veríssimo, 2013; Wright et al., 2015). However, these efforts have traditionally suffered from a lack of robust impact evaluation (Baylis et al., 2015; Ferraro & Pattanayak, 2006; Pullin & Knight, 2001). Understanding the impact of a social program or intervention serves two objectives: learning and accountability. In terms of learning, a well-designed impact evaluation can answer questions about program design, by revealing whether a program is achieving its objectives or not. This information can then be used strategically, for example, to assist in decisions about scaling up. Regarding accountability, it provides evidence to donors, members, and other constituencies that the actions of an organization are aligned with its mission.

A meaningful impact evaluation measures the causal effect of a specific intervention and requires information on the counterfactual: What the outcomes would have been in the absence of the intervention (Ferraro & Hanauer, 2014; Rosenbaum, 2010). The objective is therefore to be able to rule out other alternative explanations for the observed outcomes (Ferraro, 2009; Rosenbaum, 2010). This focus on rigorously understanding the causal relationships between intervention and outcome is particularly important in biodiversity conservation, given the increasingly stringent budgets conservationists have to operate within and the renewed emphasis on efficiency metrics such as return on investment (Baylis et al., 2015; Veríssimo, 2013).

Impact evaluation studies can be experimental or observational (Frondel & Schmidt, 2005). The main difference is that in experiments (e.g., randomized controlled trials), the researcher has control over which units (e.g., villages or participants) are assigned to the intervention and comparison groups. That is not the case in observational studies (Frondel & Schmidt, 2005; Rosenbaum, 2010). These can be divided into case studies, which are typically qualitative, and quasi-experiments, which generally have a quantitative focus (White & Phillips, 2012). Quasi-experiments are most commonly used in situations where randomization of the treatment assignment is not possible for practical or ethical reasons (Margoluis, Stem, Salafsky, & Brown, 2009).

One popular approach to quasi-experimental causal inference is the before–after–control–impact (BACI) or difference-in-differences framework. Under this experimental design, data are collected prior to and after an intervention in both treatment and comparison (i.e., control) groups (Frondel & Schmidt, 2005). The impact of the intervention is then defined as the difference between the outcome (i.e., the difference before and after the intervention) for those receiving the treatment (i.e., the conservation intervention) and those in a comparison group (Frondel & Schmidt, 2005). One important consideration is that the intervention and comparison groups must be comparable at baseline. In order to ensure this, researchers have developed a series of matching techniques which try to mitigate selection bias. They attempt this by identifying the observable biases that led to the selection of a particular treatment group or influence the outcome of the treatment and adopt those same biases as criteria for identifying a suitable comparison group (Khandker, Koolwal, & Samad, 2009). By using a combination of the BACI experimental design and matching techniques, we can mitigate the biases due to selection and simultaneous changes, two of the most common challenges to meaningful impact evaluation of conservation interventions (Khandker et al., 2009).

In sectors such as health or education, the evaluation of social marketing efforts has reached high standards with, for example, the widespread use of experimental research designs (Ferraro & Pattanayak, 2006; Gordon, McDermott, Stead, & Angus, 2006). Meanwhile, social marketing efforts aimed at conserving biodiversity have lagged behind despite some progress (Jenks, Vaughan, & Butler, 2010), with simplistic experimental designs, emphasis on indicators that are poor proxies for behavior change and lack of measurement of biological outcomes. These shortcomings are echoed across the wider conservation sciences, which have largely failed to adopt many of the best practices in impact evaluation showcased in other fields (Baylis et al., 2015).

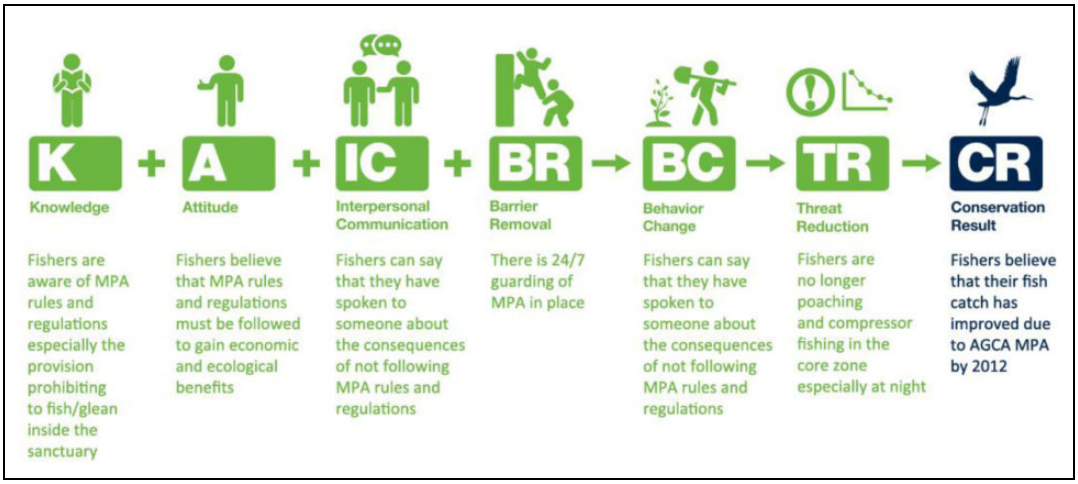

Figure 1. Locations of the Rare Pride campaign and comparison sites (*) in the Philippines. Adapted from Atrigenio et al. (2012).

In this study, we focus on the work of Rare, a U.S.-based nongovernmental organization (NGO), that has over the last 30 years conducted more than 300 social marketing interventions in over 50 countries and across topics such as fisheries, bushmeat, and watershed management (Andriamalala, Peabody, Gardner, & Westerman, 2013; Green, DeWan, Arias, & Hayden, 2013). Rare builds capacity in local partners to apply the Pride methodology, which aims to inspire people to take pride in the biodiversity around their communities while also introducing viable alternatives to environmentally destructive practices (see Butler, Green, & Galvin, 2013). The Pride method is based on social marketing principles which recognize that to change behavior, we must first understand both the motivations of the people whose behavior we want to change and the barriers that may prevent them from changing their behavior (Butler et al., 2013). The implementation of Pride campaigns is guided by a "theory of change" (TOC; Figure 1) that conceptually maps how changes in knowledge, attitudes, and interpersonal communication contribute to changes in behavior and ultimately help achieve conservation results (Butler et al., 2013).

From September 2010 to September 2012, Rare carried out its first cohort of 12 campaigns in the Philippines, which set out to transform 16 marine-protected areas (MPAs) into well-managed entities that can be a source of food security for local communities (Day et al., 2014). In all sites, the intervention consisted of both a social marketing campaign and a supporting barrier removal (BR) strategy. The campaign aimed at creating community ownership of the MPA and driving the community to manage it and voluntarily comply with existing regulations. Each campaign carried out its independent market research, consisting of focus groups and semistructured interviews, to ensure messaging and branding were tailored to the specific characteristics of the different communities. The campaign branding elements included not only a logo but also a mascot and song. The campaigns used channels such as billboards, calendars, T-shirts, stickers, school buses, murals, and radio together with activities such as school visits, religious sermons, public meetings, and sports events to spread their message to their target audiences. All campaigns had a primary target audience, focused on the fishers that used the area around the target marine protected area, and a secondary audience, focused on the remaining community members (only for the closest community).

In terms of the BR strategy, it focused on establishing and improving the required MPA governance and enforcement structures (Rare, 2012). This was achieved through participatory coastal resource assessment (PCRA) activities such as including underwater community

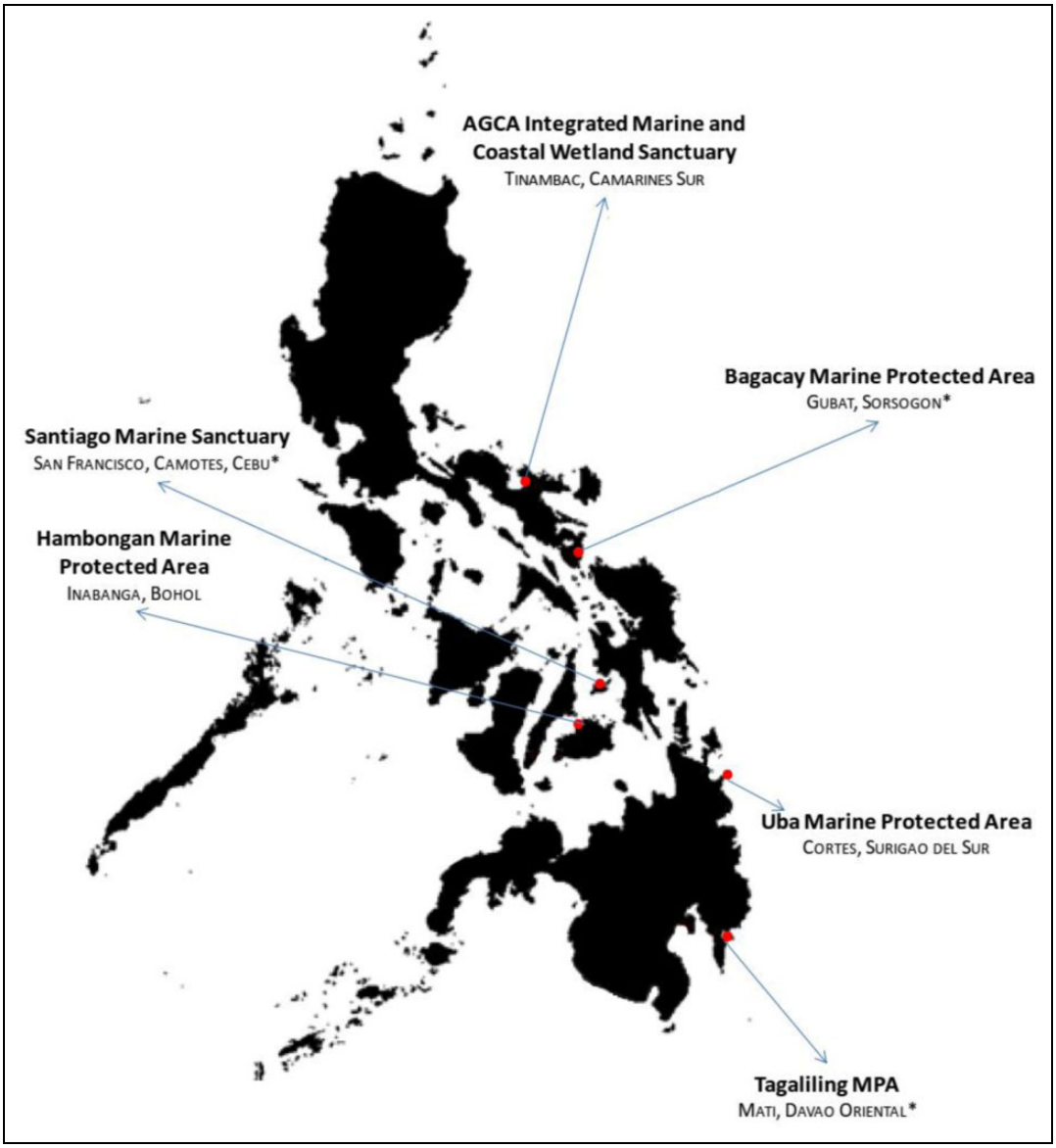

Figure 2. Example of the general conceptual model followed by Rare to structure a theory of change (TOC) for the social marketing campaigns (top) and example of a general TOC, for a fishers target audience, based on a marine social marketing campaign in the Philippines. Adapted from Demesa (2012).

monitoring and workshops to improve representation and functionality of an MPA Technical Working Group and an MPA Management Committee at each site. The latter group was instrumental to driving improvements around community enforcement. This included the installation and maintenance of MPA boundary buoys and guardhouses, the implementation of a 24/7 guarding system, the setting up of a reporting system, and training for *Bantay Dagat* (deputized fish wardens) in enforcement procedures (Rare, 2012).

Three of these 12 campaigns were selected across the three main regions of the Philippines to trial the use of a quasi-experimental BACI evaluation design (Figure 2). These were the AGCA-Integrated

Marine and Coastal Wetland Sanctuary (AGCA MPA) in Tinambac, Camarines Sur; Uba MPA in Cortes, Surigao del Sur; and Hambongan MPA in Inabanga, Bohol. The AGCA MPA was established in 2006 through a municipal ordinance. This MPA is the result of local collaborative efforts toward greater sustainability in fisheries, involving local fishers, the barangay councils of Agay-Ayan and Caloco, and the NGO Network of Sustainable Livelihoods Catalysts, Inc. (AGCA MPA MPA Technical Working Group and AGCA MPA Management Committee, 2012). The primary livelihoods in this area are fishing, farming, or seaweed farming (AGCA MPA MPA Technical Working Group and AGCA MPA Management Committee, 2012). The Uba MPA was also established in 2006 through a municipal ordinance. It was part of a regional network of MPAs implemented by the municipal government to tackle the severe depletion of the marine resources of the region, driven largely by unsustainable fishing practices such as dynamite fishing (Uba MPA Technical Working Group and Uba Management Committee, 2012). Most inhabitants of this area are full-time fishers (Uba MPA Technical Working Group and Uba Management Committee, 2012). Lastly, the Hambongan MPA, which was established in 2000 through a municipal ordinance. This was part of an internationally sponsored response by the regional government to the depletion of the marine resources that sustain much of the local population. Although it was established prior to the other two MPAs being considered in this study, it has historically suffered from poor governance and lack of support from the government (Hambongan MPA Technical Working Group and Hambongan Management Committee, 2012). The primary livelihoods in the area are fishing, seaweed farming, and the trade in ornamental fish.

In this article, we describe the impact evaluation of three social marketing campaigns implemented in the Philippines to improve marine conservation outcomes. We focus particularly on the challenges encountered in adopting a quasi-experimental approach to evaluation of behavior change interventions and discuss how we tried to overcome them. While we found some evidence that the campaigns achieved goals related to human behavior and perceptions of conservation results, we did not find any evidence that these impacts translated into the targeted biological outcomes within the 2-year time frame of this analysis. This highlights the challenges of achieving and demonstrating evidence of impact within the limited time frame frequently required by donors and other stakeholders of conservation interventions. We also discuss the implications for causal inference when thinking about behavior change in the context of biodiversity conservation, particularly as attribution becomes increasingly difficult over longer time frames of analysis. Lastly, we describe how these lessons learned can be used to improve the future impact evaluations of social marketing campaigns aimed at conserving biodiversity.

Material and Method

Experimental Design

The initial selection of 12 campaign sites in the Philippines was done according to an internal site selection process. This is a multistage assessment that takes into account the characteristics of the potential site (e.g., biological, demographic, and conservation threats), the support from local governance, and the capacity of the institutional partners and of the campaign manager. The goal is therefore to select the sites with the greatest likelihood of success. All 12 Pride campaigns had a shared TOC model (Figure 1), and implementers received similar training, mentoring, and implementation support (Rare, 2012). Three of these MPAs, Uba, Hambongan, and AGCA (Agay-ayan and Caloco), were then selected to be included in a BACI evaluation framework. These sites were selected to represent the three key regions of the Philippines.

Through an expert-led coarse matching exercise involving Rare Philippines staff and an academic from the University of the Philippines Diliman, a series of candidate matching sites were selected. This first involved selecting the variables upon which the matching would be made (Table 1). These

Table 1. Variables Used in the Matching Campaign Sites to Comparison Sites, the Rationale for Their Selection, and How They Were Measured.

Variable	Rationale
Region	Broad cultural and biological differences within the Philippines. Based on country's three main island groups: Luzon, Visayas, and Mindanao. Measured as nominal scale. Only variable for which exact matching was required.
MPA effectiveness	Differences in marine-protected area (MPA) governance can impact conservation outcomes. Based on MPA Management Effectiveness Assessment Tool (MEAT) score (Maypa et al., 2012). Measured as 1–4 ordinal scale
Climate change	Differences in potential climate change impacts (e.g., coral bleaching). Measured as nominal scale for most likely disturbances (storminess, sea surface temperature, runoff/rainfall)
MPA size	Impacts of MPA size on enforcement effort and biological recovery. Measured as area in hectares
Population	Indirect impacts of human population on nearby marine ecosystems (e.g., pollution). Measured in number of inhabitants rounded to the nearest hundred
Baseline biomass	Impacts of initial biomass on biological recovery. Measured in three ordinal levels (low, medium, and high)
Fishing intensity	Impact of fishing effort on recovery of fish populations. Measured in three ordinal levels (low, medium, and high)
Proximity to urban areas	Proxy for extent of the commercial nature of the fishery. Measured in three qualitative levels (rural, suburban, and urban)
Years established	Time for biological recovery and MPA governance stability. Measured as number of years since MPA was officially established

variables were selected to ensure that the sites were comparable at baseline and in terms of characteristics influencing social marketing and fisheries outcomes. Afterward, taking into account logistic and financial constraints, eight potential comparison sites were rated, mostly qualitatively, for each of the variables. Given the lack of published literature on many of these variables, the rating was based largely on gray literature and the professional experience of the participants at each location. From that pool of sites, three pairs were formed, involving three Pride and three comparison sites.

Data Collection

A knowledge, attitude, and practices questionnaire survey was used as the main data collection instrument. The survey contained approximately 75 questions, the majority of which measured indicators of every stage of the TOC (Figure 1). The data collection was conducted by Rare Philippines at the three Pride campaign sites, while a third-party contractor, the Asia Center for Sustainable Future, collected the data at comparison sites. Both groups used the same survey and data collection protocols. Data collection took place on February or May 2011 (presurveys) and May/June 2012 (postsurveys), with campaigns lasting between 12 and 16 months. The discrepancy in the timings of data collection for presurveys was due to logistical difficulties felt at some of the sites. Surveys were cross-sectional and used random sampling, alternating between genders in successive households. These data included data on the perceptions of fisher communities closest to the target MPAs across all stages of the TOC at Pride and comparison sites through knowledge, attitude, and practices questionnaire surveys. Although the surveys encompassed all components of the TOC (Figure 1), we focus here on the *threat reduction* and *conservation result* stages of the TOC because these stages were the only ones with standardized questions across all sites and are the closest to the conservation impact. Although data were collected from several groups within a community, we focus on the data from fishers as they would be better placed within a community to notice changes to the fishery. All surveys were anonymous and verbal consent was obtained from all respondents.

Table 2. Knowledge, Attitude, and Practices Survey Sample Sizes and Estimated Population Sizes for Campaign ([1]) and Comparison Sites ([2]).

Site	Registered Fisher Population	Pre-/Postsample
AGCA Integrated Marine and Coastal Wetland Sanctuary[1]	243	358/328
Bagacay marine-protected area[2]	138	114/120
Hambongan marine-protected area[1]	1629	350/412
Santiago Marine Sanctuary[2]	126	109/171
Uba marine-protected area[1]	86	178/140
Tagaliling marine-protected area[2]	90	78/88

Note. For Hambongan, the sample includes fishers from neighboring communities of Tungod, Cagawasan, Lawis, and Ondol. The estimated population sizes take into account the number of registered fishers, and so when the actual sample is larger than the population, this means that unregistered individuals were also sampled. Superscript indicates campaign ([1]) and control ([2]) sites, with order indicating matched pairs.

The threat reduction indicator was the number of fishers who have seen members of their village fishing in the no-take zone in last 6 months. It should be noted that due to a lapse in survey implementation, for the paired sites of Tinambac and Sorsogon, this question was asked using a 12-month time frame. The possible answers to this question were seen, not seen, and unsure/don't remember. Given the small number of respondents selecting the last category, and the difficulty in relating this category to the first two, we have not included these data in our analysis. For conservation result, the indicator used was number of fishers who say their catch has decreased, stayed the same, or increased as a result of the MPA. The survey sample sizes and estimated population sizes for Pride and comparison sites are reported in Table 2.

In an effort to triangulate these results, Rare Philippines collected data on MPA compliance through log books filled by the community members participating in MPA enforcement activities. These data included the proportion of time an MPA was guarded, the number of reports of MPA infringements, and the number of resulting arrests. Unfortunately, this was only possible at Pride sites.

Rare Philippines also collected data on fish biomass, an indicator that is able to account for variations in both size and number of fish and thus can be used to understand the ecological status of fish populations. Biomass can be estimated by recoding the number of fish of each species and their respective length and then using species-specific length–weight curves to estimate the species total biomass. Data were collected through underwater surveys both inside and outside MPAs, with 10 replicate 50 m transects per site (five inside and five outside the MPA; Atrigenio et al., 2012). A transect is line across a habitat along which data are recorded at regular intervals. Marker blocks were used as reference points for laying the four permanent transects (two inside and two outside the MPA). The remaining transects were initially randomly placed with the follow-up transects in the vicinity of the baseline location to prevent too much variation (Atrigenio et al., 2012). To make our analysis relevant to local livelihoods, which are focused on fishing, our analysis focused on fish species that are of economic importance as those would be theoretically expected to benefit the most from an improvement in MPA management. Lastly, in an effort to understand any possibility of contagion or contamination, where the comparison site is affected by other interventions with similar outcomes, data were collected ad hoc at comparison sites on any visible signs of other related programs.

Data Analysis

To understand differences at baseline between comparison and Pride sites, we calculated mean standardized differences for all matching variables. In terms of threat reduction, and given the differences in the question used at two of the six sites, each pair of treatment/control site was analyzed separately.

As one site had perfect separation between treatment and outcome, with all respondents in the posttreatment survey giving the same answer, we used Bayesian logistic regression to explore the impact of the social marketing campaign on this outcome. We did this by looking at the interaction term between the binary variable defining treatment/control status and the variable defining pre-/poststudy phase. We also include in the regression the variables *MPA size*, *MPA effectiveness,* and *years established* (Table 1), for which we found differences at baseline between Pride and comparison sites. Regarding conservation result, we used ordinal regression to examine the impact of the social marketing campaign. All sites were pooled in the same regression model, built in the same way described above.

Lastly, we investigated the impact of the social marketing campaign on fish biomass both outside and inside the MPAs. For this, we used linear mixed effects models, using the individual transect as the unit of analysis and including the same variables detailed above.

Results

In terms of differences between Pride and comparison sites, we used a general rule that considers differences above 5% to be meaningful (Glew, Mascia, & Pakiding, 2013). We found important differences in the case of the variables *MPA size, MPA effectiveness, and years established* (Table 1). MPAs at Pride sites were on average twice as large and 76% older than those in comparison sites (AGCA MPA MPA Technical Working Group and AGCA MPA Management Committee, 2012; Atrigeno & Deocadez, 2012; Hambongan MPA Technical Working Group and Hambongan Management Committee, 2012; Jesus, 2012; Martinez, 2011; Uba MPA Technical Working Group and Uba Management Committee, 2012).

In terms of threat reduction, the detection of fishermen from the target community fishing in the no-take zone, two of three Pride sites saw a decrease in the proportion of respondents who had seen members of their own community fishing in the no-take zone (Figure 3). The pattern was reversed in comparison sites with two seeing an increase and one a decrease (Figure 3). When comparing with control sites and controlling for differences in baseline and other relevant MPA traits, we only detected an impact of the Pride campaign for the Uba MPA (Table 3). Community enforcement at Pride sites improved, with the mean percentage of time an MPA was guarded increasing, while the number of reported infractions stayed constant and the number of arrests decreased (Table 4).

Regarding conservation results, the perception fishermen had of their catch increasing/staying the same/decreasing, the proportion of fishermen reporting decreased catches, was slightly reduced at Pride sites while it increased at control sites (Figure 4). When comparing with control sites and controlling for differences in baseline and other relevant MPA traits, we detected an impact of the Pride campaign (Table 3).

Concerning actual fish biomass, there was an average increase both inside and outside for Pride and comparison sites (Figure 5). However, no effect of the Pride campaign was detected, when comparing with control sites and controlling for differences in baseline and other relevant MPA traits (Table 3).

Discussion

The evaluation of social marketing campaigns targeting biodiversity is particularly challenging as the outcomes of interest extend beyond human behavior change into how these changes translate into biological results. This means that more data have to be collected and that logistics of data collection have an added level of complexity and cost. This is why it is particularly important to use the data collected to draw lessons learned and ensure that future resources are used more effectively and efficiently to achieve biodiversity outcomes. Such lessons will be valuable not only broadly in terms of social marketing implementation but also in the context of the Philippines, where MPAs have taken

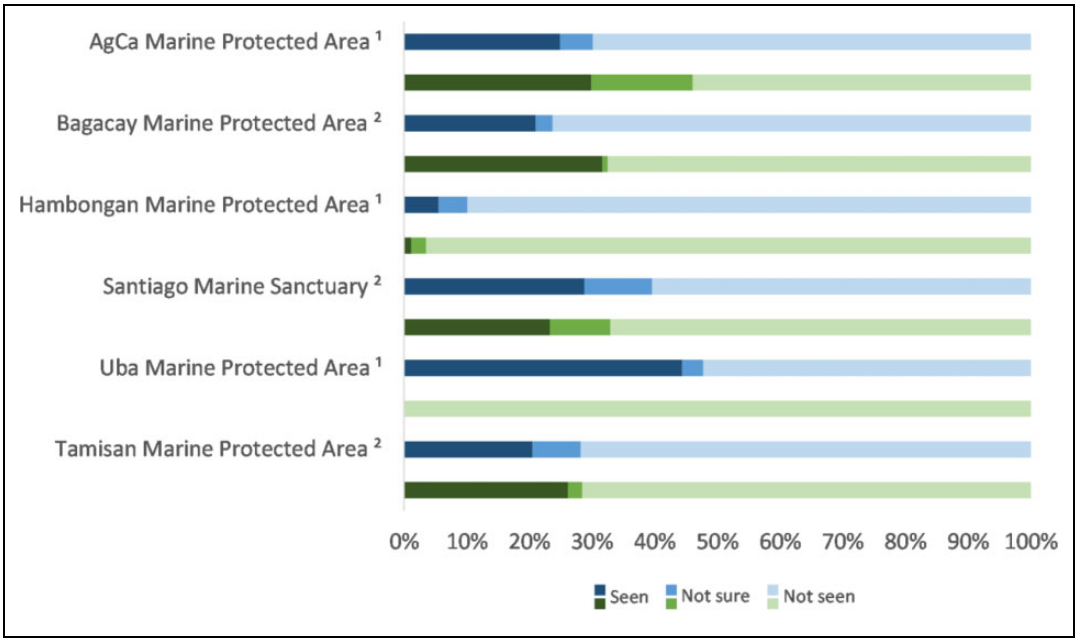

Figure 3. Changes in perceived *threat reduction* at Rare Pride campaign and comparison sites. Threat reduction indicator was number of fishers who have seen members of their community fishing in the no-take zone in last 6 months (except for AGCA-Integrated Marine and Coastal Wetland Sanctuary and Bagacay Marine–Protected Area where the time frame was 12 months). Superscript indicate campaign (^{1}) and control (^{2}) sites, with order indicating matched pairs. Blue indicates the period prior to the intervention while green indicates the period after the intervention.

on an important role in the conservation of marine resources and the maintenance of the livelihoods associated with them (Christie, White, & Deguit, 2002; Weeks, Russ, Alcala, & White, 2010).

The transparent discussion of evaluation practices and process is a key part of a dialogue around accountability between those implementing conservation interventions and those constituencies supporting them such as donors, members, and government partners. Yet, implementing organizations often does not have adequate incentive to truly evaluate the outcomes of their efforts, and even when they do, significant challenges persist, not only in ensuring that the outcomes measured reflect actual change but also in linking the changes in outcomes to the organization's own intervention or program. It is here that a more thorough implementation of impact evaluation principles can yield key improvements in the way we evaluate the impact of social marketing interventions targeted at biodiversity conservation objectives.

Experimental Design

By using a BACI design that estimates the counterfactual using matched control sites, the current research is moving beyond the before–after comparisons that have commonly been used in evaluating conservation outreach. This is because the use of control sites improves our ability to take into account, for instance, changes in the environmental and macroeconomic conditions taking place while the intervention is ongoing. Such changes can be particularly important when interventions are implemented over periods of several years, as is the case with Pride campaigns. In addition, the use of a matching process provides a transparent and data-driven process to ensure the comparison sites selected are a credible counterfactual, which is critical for a valid estimation of the impact

Table 3. Estimates of Impact for Perceived Threat Reduction (TR), Conservation Result (CR), and Changes in Fish Biomass Inside and Outside of the Marine-Protected Area (MPA).

	Treat		Stage		Treat: Stage		MPA effectiveness		MPA Size		Years Established	
Indicator	β	*p* Value	β	*p* Value	β	*p* Value	β	*p* Value	β	*p* Value	β	*p* Value
TR Hambongan	0.0004	1	−0.043	.359	−0.012	.829	−.006	.993	.003	.998	0.002	.999
TR AGCA	0.00001	1	0.1	.086	−0.009	.897	.0001	.999	.001	.999	0.00004	1
TR Uba	0.003	.999	0.044	.481	−0.52	>.001***	.016	.988	−.013	.993	−0.005	.998
CR	−0.084	.62	−0.453	.003**	0.536	.004**	.0445	>.001***	.016	>.001***	0.066	>.001***
Biomass inside MPA	−19.643	.552	7.350	.413	18.168	.122	.1	.92	.337	.542	1.086	.817
Biomass outside MPA	−8.356	.481	3.347	.554	8.814	.263	−.1179	.743	.254	.306	−0.204	.9

Note. Values below the significance level of .05 are marked with an asterisk: *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Table 4. Community Enforcement Effort, Outputs, and Outcomes for Marine-Protected Areas in Pride Campaign Sites.

Indicator	Pre-Pride	Post-Pride
Mean percentage of guarding per month	60%	81%
Mean number of reports	5	5
Mean number of arrests	4	0

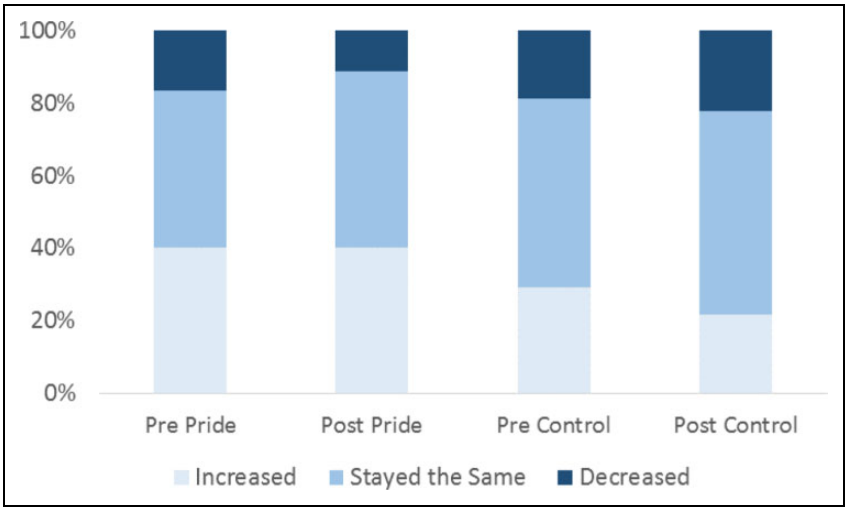

Figure 4. Changes in perceived *conservation result* at Rare Pride campaign and comparison sites. Conservation result indicator was number of fishers who say their catch has decreased/stayed the same/increased as a result of the local marine–protected area.

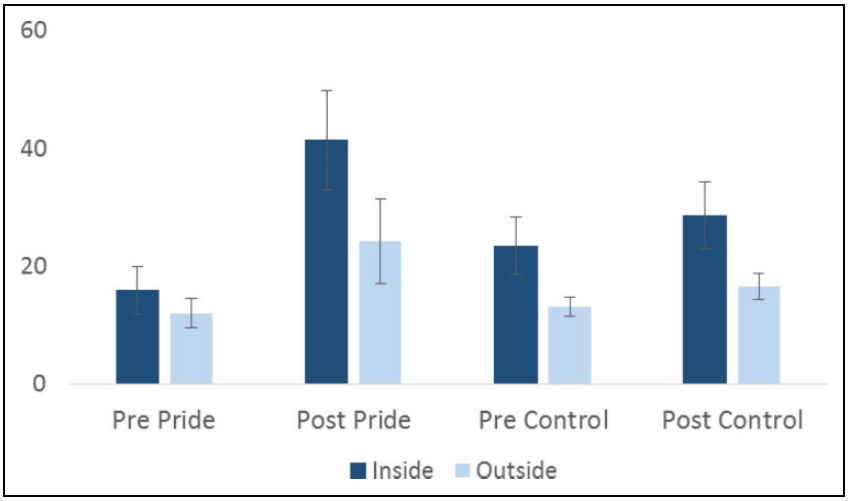

Figure 5. Changes in fish biomass (mt/km^2) inside and outside the marine-protected areas at Rare Pride campaign and control sites.

of an intervention. However, this substantial improvement still meets a number of challenges and limitations.

One relates to the process followed to match Pride and comparison sites. The expert-led process, although cost-effective and advantageous for data-poor contexts, is also prone to bias, particularly when it is not feasible to involve a wider group of experts. Further, a largely qualitative matching process offers the advantage of relative simplicity and flexibility to imperfect realities on the ground but at the same time allows greater vulnerability to variable imbalances between Pride and comparison sites. The differences uncovered between Pride and comparison sites are not negligible in the context of sustainable fisheries, as they put the Pride sites at an advantage in relation to comparison sites. For example, larger MPAs are expected to be able to recover more quickly from past impact of unsustainable fishing. This means these variables are rival explanations for a potential positive impact at Pride sites, and thus a degree of selection bias is still present. At the same time, this exercise allowed for a more precise understanding of what biases may be impacting outcomes, making it possible to investigate those aspects more closely.

One way to try to address the limitations described above would be to adopt an experimental design that assigns sites randomly to treatment (i.e., Pride campaigns) and comparison groups. This however does not necessarily mean adopting an unrestricted randomized approach, for which there may be both practical and ethical obstacles (Baylis et al., 2015). A more limited restricted randomization process could also help reduce selection bias, among others, while making use of existing process and institutional pathways. Currently, Rare has a multistep structured process that it uses to select the sites at which it will implement, in partnership, its interventions. This process starts from a large pool of initial sites and progressively reduces the pool of candidates. Introducing randomization at the last step to divide the final candidates into comparison and intervention sites would provide practitioners with more robust impact evaluation. Another, less robust, quasi-experimental alternative would be to use the same criteria used in the site selection process to select comparable comparison sites. This alternative is already being piloted in Rare's most recent project to increase the sustainability of small-scale coastal fisheries across five countries.

Another constraint was the small number of sites for which data were available. Rare invested a lot of resources in achieving a robust sampling frame within each site, but this meant that only three Pride sites where evaluated to balance the additional cost involved. Given that it is often simply impossible for organizations to accommodate the costs of more extensive data collection, one option would be to use already available data sets collected by third parties. One example of this is the many research projects that use data from the Demographic and Health Surveys Program run by U.S. Agency for International Development to evaluate the impact of health projects. Another crucial, and often overlooked, source of economy when it comes to evaluation is to ensure data sharing between organizations working in the same regions.

When thinking of the experimental design, another necessary consideration refers to the possibility of spillover effects between intervention and control sites. These effects happen when the impact of the intervention spreads to units that were not part of the intervention. This can happen, for example, through formal and informal human networks but also through the use of mass media to spread campaigns messages. The former makes social marketing interventions particularly vulnerable as these interventions often rely on mass media and other channels that reach multiple locations simultaneously. Spillover effects can undermine the impact evaluation by having control sites benefit from the intervention. Although the literature is equivocal on how buffers against spillover should be set, given the minimum straight-line distance between comparison and Pride sites was 67 km, it seems unlikely spillovers would be an issue. The considerations over physical distance are less relevant, as they would of course not apply to messaging transmitted over mass media such as radio, where some spillover could have happened.

In terms of contagion, where the comparison site is affected by other interventions with similar outcomes, all three comparison sites saw some degree of activity by other NGOs around marine

conservation, but these were largely limited to sporadic visits for data collection purposes or one-off assistance with specific issues such as MPA enforcement. Although the data collection on other interventions at comparisons sites was largely opportunistic, we feel confident that any major initiatives would have been easily detected. As such, given the small and ad hoc scale of all interventions detected, we believe the likelihood of contagion was small.

Understanding Impact

Regarding the outcomes of the current study, only in Uba did threat reduction see a statistically significant change when comparing Pride and comparison sites and taking into account baseline levels. This difference could be due to the small fisher population, which meant that the social marketing effort was higher per target audience member, therefore increasing the potential for impact. It should be noted in addition that the lack of further statistically significant results could be due to the small sample size used, which reduces the statistical power and thus the ability to detect changes, especially if they are small. To be able to foresee these issues and adequately prepare in advance, Rare could undertake, prior to data collection, a power analysis, which determines the sample size required to detect an effect of a given size with a given degree of confidence. In this way, it would be possible to better plan the sampling strategy to ensure it collects data in a way that enables the organization to detect the type of change that the project hopes to achieve.

In terms of catches, the perceptions of fishermen were found to have significantly improved at Pride sites when compared to control sites. This is an encouraging result, particularly considering how far down the TOC this indicator is and the important dimension that perceptions can adopt in sustaining local support for the management of natural resources (Bennett, 2016). Yet, an issue to bear in mind when interpreting these results is the potential bias that can impact self-reported indicators such as the one used in this study. Although a recent review comparing self-reported and actual behavior found that self-reported indicators were on average 9% higher than actual behavior, it did not find evidence of systematic biases (Kormos & Gifford, 2014). This makes it difficult to correct self-reported indicators to reflect actual change. Unsurprisingly, the same study finds that self-reported indicators are poor predictors of actual behavior and that their interpretation should be cautious (Kormos & Gifford, 2014). For this reason, there is great need for using independent data sources such as enforcement data to help triangulate the results and understand if at least the broader trends are likely to be robust. If several independently collected data sources point toward the same trend, then it is more likely that this trend is real and not the result of a measuring bias. Considering both the increase in enforcement effort and the increase in awareness generated by Pride, it would be expected that the number of reports and arrests would increase if compliance with the law remained the same. The similar number of reports and decrease in arrests found in the current study therefore supports the assertion that no take zone violations did indeed decrease in the Uba MPA. Lack of enforcement data for comparisons sites means that the usefulness of the enforcement data at Pride sites is much reduced since we cannot know, for example, if broader trends across the region or countries are driving these changes registered. Another aspect to bear in mind originates from the fact that data collection at treatment and control sites was done by two different groups. Despite the efforts to standardize questionnaires and survey protocols, there is the potential for the different rapport of different enumerators to influence the survey outcomes. At the same time, it should be acknowledged that there is a potential conflict of interest in Rare collecting the data to be used in the evaluation of its impact. The solution to this is however not as immediate as it may seem as contractors also have the incentive to try and produce the expected results under the penalty of not being selected for future contracts.

The statistically significant difference in perceived fish catch (Table 3) however was not supported by the biological monitoring of fish biomass, where no statistically significant changes were detected inside or outside of the MPAs (Table 3). This discrepancy suggests that some self-reporting bias, either

through social desirability or with the fishermen changing their perception of their catch, may have occurred. Nonetheless, the result of the biological monitoring is aligned with our current knowledge of fish population dynamics which asserts a recovery cannot be expected in such a short time frame (MacNeil et al., 2015; Russ & Alcala, 2004). Although recovery times are often context-specific, a recent study argues that species recoveries triggered by MPAs only start to be noticeable on average 5.1 ± 1.9 years after protection (Babcock et al., 2010). This illustrates another complexity of working with biological outcomes, as there is often large uncertainty around their timescale for recovery. It is therefore crucial to align the expectations of donors and other constituencies as to what constitutes a realistic time line for reporting on intervention outcomes.

Lastly, it is worth referring that there are other outcomes that future impact evaluation research on this topic could also consider. These include participation of fishers in the management of MPAs, increased civil society participation in fisheries-related public discussions, and greater gender balance in the decision-making around the management of natural resources. While these less tangible outcomes represent an important outcome of the empowering of communities to manage their own natural resources, they are often more difficult to capture with qualitative indicators. A potential way of addressing this challenge would to use qualitative data collection techniques such as ethnographies, to explore in more depth the social impact of social marketing interventions. The challenge to evaluators will continue to be, acknowledging that any evaluation is limited in the number of indicators measured, to ensure that the most relevant indicators are prioritized.

Conclusion

The implementation of meaningful impact evaluation of social marketing interventions is a complex venture. This article documents many of the barriers that still remain, particularly if causal links between outcomes and specific interventions are to be drawn, and highlights a number of lessons learned from this experience in seeking a practical, ethical, and effective approach to meaningful impact evaluation.

Rare has historically been concerned with monitoring and evaluation (Jenks et al., 2010), and this article is a result of the organization's commitment toward accountability and learning. Nonetheless, steps such as the adoption of comparison sites at a much larger number of sites, the use of a systematic method for site matching, and the move beyond direct self-reported indicators of behavior change are critical to ensure NGOs remain accountable to their constituents and learn how to adapt interventions to achieve best possible outcomes with partners. One source of experience in this could come from both the health and international development fields which not only often work with similar audiences to those targeted by conservation programs but which have used robust experimental evaluation methods for many years. It is clear that in a world where conservation is often presenting its benefits in terms of human well-being and improvement in livelihoods, and competing for funding with fields like health and development, not reaching that standard will in the long run put conservationists at a disadvantage the field may simply not be able to afford.

Acknowledgments

This research was only possible due to the efforts of the Rare Philippines staff and campaign managers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

AGCA MPA MPA Technical Working Group and AGCA MPA Management Committee. (2012). *AGCA Marine Sanctuary MPA Management Plan (2013–2015)*. Retrieved from http://www.rareplanet.org/sites/rareplanet.org/files/agca_mpa_management_plan_ver_082912.pdf

Andriamalala, G., Peabody, S., Gardner, C. J., & Westerman, K. (2013). Using social marketing to foster sustainable behaviour in traditional fishing communities of southwest Madagascar. *Conservation Evidence*, *1*, 37–41.

Atrigeno, M., & Deocadez, M. (2012). Tagaliling MPA—MPA Management Effectiveness Assessment Tool (MEAT). Quezon, Philippines: University of Philippines Marine Science Institute.

Atrigenio, M. P., Martinez, R. J. S., Panga, F. M., Deocadez, M. R., Aliño, P. M., Mamauag, S. S., & Muallil, R. (2012). *Philippines 1 Biophysical Monitoring Report*. Quezon City, Philippines: Marine Environment and Resources Foundation, Marine Science Institute, University of the Philippines Diliman.

Babcock, R., Shears, N., Alcala, A., Barrett, N., Edgar, G., Lafferty, K., . . . Russ, G. (2010). Decadal trends in marine reserves reveal differential rates of change in direct and indirect effects. *Proceedings of the National Academy of Sciences*, *107*, 18256–18261.

Balmford, A., & Cowling, R. M. (2006). Fusion or failure? The future of conservation biology. *Conservation Biology*, *20*, 692–695. doi:10.1111/j.1523-1739.2006.00434.x

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., . . . Wunder, S. (2015). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, *9*, 58–64.

Bennett, N. J. (2016). Using perceptions as evidence to improve conservation and environmental management. *Conservation Biology*, *30*, 582–592

Butler, P., Green, K., & Galvin, D. (2013). *The Principles of Pride: The science behind the mascots*. Arlington, VA: Rare. Retrieved from http://www.rare.org/publications

Christie, P., White, A., & Deguit, E. (2002). Starting point or solution? Community-based marine protected areas in the Philippines. *Journal of Environmental Management*, *66*, 441–454.

Day, B. A., DeWan, A., Cadiz, F. C., Jakosalem-Balane, J., Dueñas, V., & Trinidad, P. M., Jr. (2014). Rare social marketing for sustainable fishing in Cortes, Surigao del Sur, Philippines. *Applied Environmental Education & Communication*, *13*, 56–65.

Demesa, C. B. (2012). *AGCA Marine Sanctuary Campaign Campaign Learning Report*. Arlington, VA: Rare. Retrieved from http://s3.amazonaws.com/rarect_prod/rareplanet.org/files/agca_clr_final_draft_submitted.pdf

Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, *122*, 75–84.

Ferraro, P. J., & Hanauer, M. M. (2014). Advances in measuring the environmental and social impacts of environmental programs. *Annual Review of Environment and Resources*, *39*, 495–517.

Ferraro, P. J., & Pattanayak, S. K. (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biol*, *4*, e105. doi:10.1371/journal.pbio.0040105

Frondel, M., & Schmidt, C. M. (2005). Evaluating environmental programs: The perspective of modern evaluation research. *Ecological Economics*, *55*, 515–526.

Glew, L., Mascia, M. B., & Pakiding, F. (2013). *Solving the mystery of MPA performance: Social impacts of MPAs in the bird's head seascape*. Washington, DC: World Wildlife Fund and Universitas Negeri Papua, United States, and Manokwari, Indonesia.

Gordon, R., McDermott, L., Stead, M., & Angus, K. (2006). The effectiveness of social marketing interventions for health improvement: What's the evidence? *Public Health*, *120*, 1133–1139.

Green, K. M., DeWan, A., Arias, A. B., & Hayden, D. (2013). Driving adoption of payments for ecosystem services through social marketing, Veracruz, Mexico. *Conservation Evidence*, *10*, 48–52.

Hambongan MPA Technical Working Group and Hambongan Management Committee. (2012). *Hambongan MPA management plan (2012–2017)*. Retrieved from http://www.rareplanet.org/sites/rareplanet.org/files/mpa_governance_and_management_plan_brop_hambongan.docx

Jenks, B., Vaughan, P. W., & Butler, P. J. (2010). The evolution of Rare Pride: Using evaluation to drive adaptive management in a biodiversity conservation organization. *Evaluation and Program Planning*, *33*, 186–190.

Jesus, D. O. D. (2012). Ragnas Marine Protected Area—MPA Management Effectiveness Assessment Tool (MEAT). Quezon, Philippines: University of Philippines Marine Science Institute.

Khandker, S., Koolwal, B., & Samad, H. (2009). *Handbook on impact evaluation*. Washington, DC: World Bank.

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology*, *40*, 359–371.

MacNeil, M. A., Graham, N. A., Cinner, J. E., Wilson, S. K., Williams, I. D., Maina, J., . . . Polunin, N. V. (2015). Recovery potential of the world's coral reef fishes. *Nature*, *520*, 341–344.

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.

Martinez, R. J. (2011). Santiago Marine Sanctuary—MPA Management Effectiveness Assessment Tool (MEAT). Quezon, Philippines: University of Philippines Marine Science Institute.

Maypa, A. P., White, A. T., Caňares, E., Martinez, R., Eisma-Osorio, R. L., Aliňo, P., & Apistar, D. (2012). Marine protected area management effectiveness: Progress and lessons in the Philippines. *Coastal Management*, *40*, 510–524.

Pullin, A. S., & Knight, T. M. (2001). Effectiveness in conservation practice: pointers from medicine and public health. *Conservation Biology*, *15*, 50–54.

Rare. (2012). Cohort Learning Report: Philippines 1 (PEP2). Cebu, Philippines: Rare Philippines.

Rosenbaum, P. R. (2010). *Design of observational studies*. New York, NY: Springer.

Russ, G. R., & Alcala, A. C. (2004). Marine reserves: Long-term protection is required for full recovery of predatory fish populations. *Oecologia*, *138*, 622–627.

Uba MPA Technical Working Group and Uba Management Committee. (2012). *Uba MPA management plan (2013–2015)*. Retrieved from http://www.rareplanet.org/sites/rareplanet.org/files/agca_mpa_management_plan_ver_082912.pdf

Veríssimo, D. (2013). Influencing human behaviour: an underutilised tool for biodiversity management. *Conservation Evidence*, *10*, 29–31.

Weeks, R., Russ, G. R., Alcala, A. C., & White, A. T. (2010). Effectiveness of marine protected areas in the Philippines for biodiversity conservation. *Conservation Biology*, *24*, 531–540.

White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework*. New Delhi, India: International Initiative for Impact Evaluation.

Wright, A. J., Veríssimo, D., Pilfold, K., Parsons, E. C. M., Ventre, K., Cousins, J., . . . McKinley, E. (2015). Competitive outreach in the 21st century: Why we need conservation marketing. *Ocean & Coastal Management*, *115*, 41–48. doi:10.1016/j.ocecoaman.2015.06.029

Author Biographies

Diogo Veríssimo, PhD, specializes in the design and evaluation of interventions to change environmentally relevant human behaviors, particularly through the use of marketing theory and tools. Throughout his 15 years of experience in the fields of social marketing and community-based conservation, Diogo has authored more than 60 peer-reviewed publications and book chapters. You can learn more about Diogo's work at www.diogoverissimo.com

Annalisa Bianchessi is a senior communications manager at the Microinsurance Network in Luxembourg.

Alejandro Arrivillaga is a marine biologist and oceanographer with over twenty-five years of practical work experience in marine resource management, assessment, and monitoring, in regional and

international settings. His expertise includes the implementation of marine conservation projects, working with governments, local NGOs and stakeholders, and collaborating on alliances and initiating conservation mechanisms.

Fel Ceasar Cadiz is the director for Program Implementation at Rare Philippines where he is managing a cohort of 12 behavior change campaigns addressing various forms of overfishing in nearshore communities in the Philippines. He has presented and co-authored several international papers on the application of behavior design and social marketing in the Philippines.

Roquelito Mancao is the current lead of the Technical Team of Rare in the Philippines. He joined Rare in 2011 after serving as the associate for partnership for the recruitment of the Philippines 2 Cohort. He is a Licensed and Registered Fisheries Technologist, and has worked for more than 15 years on coastal and fisheries management in more than 100 coastal LGUs in the Philippines.

Kevin Green, Center for Behavior & the Environment, Rare, works closely with the organization's staff and partners to integrate state-of-the-art science about human motivation and decision-making into the execution of conservation programs. He is a faculty member of the Kinship Conservation Fellows program and holds an MA in Economics and International Development from Johns Hopkins University and a BA in Anthropology and Sociology from Washington and Lee University.