PHILOSOPHICAL
TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Ahmadia GN, Glew L, Provost M, Gill D, Hidayat NI, Mangubhai S, Purwanto, Fox HE. 2015 Integrating impact evaluation in the design and implementation of monitoring marine protected areas. *Phil. Trans. R. Soc. B* **370**: 20140275. http://dx.doi.org/10.1098/rstb.2014.0275

Accepted: 15 August 2015

One contribution of 16 to a theme issue 'Measuring the difference made by protected areas: methods, applications and implications for policy and practice'.

Subject Areas:
environmental science, ecology

Keywords:
conservation impact, counterfactual, monitoring, Bird's Head Seascape, Indonesia, coral reefs

Author for correspondence:
Gabby N. Ahmadia
e-mail: gabby.ahmadia@gmail.com

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2014.0275 or via http://rstb.royalsocietypublishing.org.

THE ROYAL SOCIETY
PUBLISHING

Integrating impact evaluation in the design and implementation of monitoring marine protected areas

Gabby N. Ahmadia[1], Louise Glew[2], Mikaela Provost[1], David Gill[3,4], Nur Ismu Hidayat[5], Sangeeta Mangubhai[6,7], Purwanto[6] and Helen E. Fox[8]

[1]Oceans, and [2]Science and Innovation, World Wildlife Fund, 1250 24th Street, Washington, DC 20037, USA
[3]National Socio-Environmental Synthesis Center (SESYNC), 1 Park Place, Suite 300, Annapolis, MD 21401, USA
[4]Luc Hoffmann Institute, WWF International, Avenue du Mont-Blanc, 1196 Gland, Switzerland
[5]Conservation International, Raja Ampat Marine Program, Jl. Kedondong, Puncak Vihara, Sorong, 98414 West Papua, Indonesia
[6]The Nature Conservancy, Indonesia Marine Program, Jl. Sultan Hasanudin No. 31, Sorong 98414 West Papua, Indonesia
[7]Wildlife Conservation Society, Fiji Country Program, 11 Ma'afu Street, Suva, Fiji
[8]Research and Monitoring; Rare, Inc. 1310 N. Courthouse Rd, Ste 110, Arlington, VA 22201, USA

GNA, 0000-0002-6489-7377; LG, 0000-0001-6981-1918; DG, 0000-0002-7550-1761

Quasi-experimental impact evaluation approaches, which enable scholars to disentangle effects of conservation interventions from broader changes in the environment, are gaining momentum in the conservation sector. However, rigorous impact evaluation using statistical matching techniques to estimate the counterfactual have yet to be applied to marine protected areas (MPAs). While there are numerous studies investigating 'impacts' of MPAs that have generated considerable insights, results are variable. This variation has been linked to the biophysical and social context in which they are established, as well as attributes of management and governance. To inform decisions about MPA placement, design and implementation, we need to expand our understanding of conditions under which MPAs are likely to lead to positive outcomes by embracing advances in impact evaluation methodologies. Here, we describe the integration of impact evaluation within an MPA network monitoring programme in the Bird's Head Seascape, Indonesia. Specifically we (i) highlight the challenges of implementation 'on the ground' and in marine ecosystems and (ii) describe the transformation of an existing monitoring programme into a design appropriate for impact evaluation. This study offers one potential model for mainstreaming impact evaluation in the conservation sector.

1. Introduction

Marine ecosystems are under threat, with increasing pressure from coastal development, over-exploitation of resources and increasing frequency of large-scale natural disturbances associated with climate change [1–3]. Marine protected areas (MPAs) are a widely used, spatially explicit conservation tool to mitigate these threats, enhance the resilience of marine ecosystems to disturbances, as well as protect biodiversity and enhance fisheries and the livelihoods of those who depend on marine resources [4,5]. The rapid expansion of MPAs across the globe is likely to continue, given the mismatch between current global MPA coverage (3.4%) [6] and the Aichi target 11 under the Convention on Biological Diversity, which commits countries to conserve and effectively manage at least 10% of coastal and marine areas by 2020 [6,7].

While ecological outcomes[1] of MPAs are generally positive, they vary significantly [8–12]. For example, a meta-analysis of biological outcomes of MPAs found positive trends attributed to MPAs with variation among sites and across indicators that in some cases exceeded an order of magnitude [10]. Variation in MPA outcomes has been linked to the biological, social and physical contexts in which they are established [13–15], as well as an array of attributes linked to management effectiveness [16] and marine resource governance [17]. To inform decisions about MPA placement, design and implementation, we need to expand our understanding of the conditions under which MPAs are likely to lead to positive ecological outcomes. This requires well-designed studies using best-practice evaluation methods designed to explicitly measure the impact (see endnote 1) of MPAs on target outcomes [18,19]. However, the ability of scholars to draw robust inferences about MPA impacts is constrained by the current limitations of the MPA evidence base, with many MPAs lacking an appropriate research design for monitoring, as well as a substantial disconnect between MPA objectives and the outcomes monitored [19,20]. In particular, relatively few studies have appropriate spatial and temporal replication (including baseline monitoring), monitor appropriate indicators or control for confounding factors, risking inaccurate or misleading results [20,21]. To understand the ecological impacts of MPA establishment, scholars and practitioners are increasingly adopting more robust monitoring approaches, by embracing advances in impact evaluation methodologies [22].

Impact evaluation is designed to measure the intended and unintended consequences of an intervention [23]. When applied in the conservation sector, impact evaluation focuses on disentangling the effects attributable to a particular policy intervention (e.g. protected areas) on a variable of interest (e.g. deforestation) from broader changes in a region (e.g. widespread development or government policies) [19,24]. Causal inference in impact evaluation rests on the Neyman–Rubin model [23,25], which compares the outcomes observed in an intervention with an explicit estimate of the outcomes in the absence of that intervention (i.e. the counterfactual [18,23]). In randomized experiments, random assignment to treatment and control groups allows for the isolation and identification of the treatment effect, which represents the differences in observations between the two groups. In the conservation sector, where management or policy interventions (e.g. protected areas, payments for ecosystem services schemes) target either regions with high biodiversity value [26] or those with lower economic opportunity costs [27], random treatment assignment is seldom feasible. Consequently, the majority of impact evaluations for conservation have adopted quasi-experimental designs to construct a counterfactual (e.g. [27,28]). Quasi-experiments apply one of a suite of statistical techniques (e.g. instrumental variables [29], matching [30]) that account for the biases in the placement of interventions (or 'observable bias'), to construct a counterfactual (see endnote 1) [23]. For the remainder of this paper, the term 'impact evaluation' refers specifically to studies that explicitly estimate the counterfactual using quantitative methods.

The application of quasi-experimental impact evaluation techniques to understand conservation impacts has generated a nascent evidence base that documents the ecological and social impacts of protected areas at global [27,28,31] or national scales [32–34], as well as the impacts of payments for ecosystems services [35] and certification approaches [36]. However, current impact evaluation has been constrained to only a small subset of outcomes (e.g. habitat cover [27]): aggregate poverty metrics (e.g. [32,37]) and policy interventions (e.g. terrestrial protected areas [28]; terrestrial ecosystem system services schemes [38]) arising from management decisions. The first generation of impact evaluations, for example, typically generated broad policy insights at national or global scales [27], based on the retrospective analysis of secondary datasets. The reliance on secondary data that can be remotely sensed (e.g. forest cover [27,28,32,33]) or understood from national-scale data collection methods has skewed the existing evidence base towards a subset of interventions, outcomes and policy decisions. Impact evaluation is in its infancy in the conservation sector and there remains ample opportunity for the 'second-generation' impact evaluations to build upon existing studies and focus on new interventions (e.g. community-based natural resource management, fisheries certification schemes), providing increased insights for policy and practice.

Currently, a considerable disconnect exists between the implementation of conservation interventions and the scientific literature on impact evaluation, stemming from mismatches between the spatial, temporal and conceptual resolutions of impact evaluation evidence, and the information required for adaptive management at the site or regional level [39]. For MPA managers and others tasked with adaptive management, who require data at fine spatial resolution and in near real-time to inform management actions [40], the existing quasi-experimental evidence base probably represents a small fraction of the information needed to guide adaptive management decisions [41]. For example, an MPA manager may need to understand the distribution of specific social impacts (e.g. food security or income) across social groups (e.g. fishers versus non-fishers) over relatively short time-frames, information that is seldom available from secondary sources [42]. Similarly, management decisions about harvest rules may require specific, site-level information on the status of key fisheries species or functional groups, calling for an ongoing monitoring effort that allows for causal inference with an appropriate level of confidence [41]. While scale mismatches between conservation science and practice are not unique to impact evaluation [43], they pose a considerable barrier to widespread adoption of evidence-based conservation practices at the local scale, and represent a missed opportunity to use robust evidence to inform adaptive management.

For impact evaluation to transform the conservation sector into an evidence-based discipline, there is an urgent need to reconcile the scale mismatches that limit the adoption of evidence on impact and expand the application of impact evaluation methodologies across interventions, outcome types and geographies [19,39]. Integration of impact evaluation into the day-to-day implementation of conservation actions is one of a suite of approaches that could help to bridge this gap. However, this integration does pose challenges that require careful consideration. In the following sections, we describe the theoretical and practical challenges of transforming conventional 'on the ground' MPA monitoring programmes to robust impact evaluation frameworks that explicitly control for observable bias in the placement or outcomes of MPAs, with streamlined data collection and control sites. We then provide an illustrative example to demonstrate one potential solution that in the long term will enable causal inferences to be made between the establishment of the MPAs and an array of ecological outcomes. Specifically, we describe the quasi-experimental impact evaluation approach developed to monitor the ecological impacts of MPAs in Birds' Head Seascape (BHS) in West Papua, Indonesia. The BHS presents an ideal case study, as there was an existing, ongoing large-scale ecological monitoring programme in place, implemented while the MPA network was in the process of being established. The BHS case study provides one example of an approach to evaluating the impact of a conservation intervention that allows us to bridge the scale mismatches between evidence and decisions, embeds the potential for quasi-experimental impact evaluation into the day-to-day implementation of conservation interventions, and offers a model for mainstreaming impact evaluation across the conservation sector.

2. Challenges

While many challenges to impact evaluation are not unique to the conservation sector or marine systems, special considerations include: (i) monitoring ecological outcomes in marine ecosystems; (ii) selection of appropriate and meaningful indicators; (iii) selection of appropriate research designs; and (iv) controlling for confounding factors.

(a) Monitoring ecological outcomes in marine ecosystems

Documenting the status and trends of many ecological attributes is challenging owing to the stochasticity and heterogeneity of both terrestrial and marine ecosystems. High replication and statistical power are often required to accurately capture the spatial heterogeneity of ecosystems [44–46], with long time-series required to disentangle cyclical or directional changes [47,48] from those changes attributed to conservation interventions. While a subset of ecological outcomes can be remotely sensed (e.g. deforestation rates), which enables scholars to examine lengthy time-series across many replicates, many management-relevant ecological outcomes require on-the-ground measurement to understand ecosystem function and populations. Monitoring MPA ecological impacts, for example, typically requires *in situ* data collection underwater (e.g. fish biomass) or observed fishing at sea or at landing sites (e.g. fish catch), with implementation spanning the period before and after MPA establishment and including sites outside of MPAs. The substantial and sustained funding required to maintain monitoring at the appropriate spatial and temporal scales means that rigorous impact evaluation remains rare in marine ecosystems [20].

(b) Selection of appropriate and meaningful indicators

Impact evaluation lies at the intersection between science and policy [18,39], with the goal of determining whether a specific intervention is achieving its desired outcome. Consequently, ecological outcomes that are measured need to be salient for adaptive management at the local scale, broader conservation policy or both (i.e. if an MPA's objective is to sustain biodiversity, then the outcomes should reflect biodiversity metrics). Where possible, impact evaluations monitor outcomes that align with the full range of intended MPA management objectives, as well as include metrics capable of detecting unintended changes in an ecosystem. However, conservation goals for marine reserves are often poorly defined [49], leading to different interpretations of appropriate indicators and 'success'.

Monitoring marine ecological systems may entail data collection on a range of indicators ranging from physical (e.g. sedimentation, habitat complexity) to biological (e.g. density and biomass of fish populations). Indicator selection, scale, data collection methods and analytical approaches can substantially influence trends detected, leading to very different interpretations of MPA impact. Consequently, several criteria need to be carefully considered when selecting appropriate indicators of MPA ecological impact, including sensitivity to change and management relevance [9,50].

(c) Selection of appropriate research designs

Substantial literature exists on the design of monitoring efforts to document the impacts of MPAs and similar policy interventions, drawing on both econometric and ecological theory [51–53]. Scholars in both disciplines emphasize the need for appropriate controls to support causal inference [19,54]. Advances in the identification of controls (e.g. matching to control for biases in MPA placement) developed by econometricians [51] can be adopted to complement research designs developed to account for temporal stochasticity in ecological systems (e.g. Before-After-Control-Impact Paired Series (BACIPS) [54]). The identification of appropriate research designs to enable causal inference varies, depending on local context (e.g. secondary data available), the characteristics of the intervention (e.g. level of replication, establishment date [24,54]), and the level of rigour required to support decision-making [41]. At a minimum, causal inference with a high degree of confidence requires the ability to disentangle the effects of MPA establishment from other policy interventions [19] or natural perturbations through time [53], in a manner that avoids confounding factors owing systematic biases in the placement of MPAs [51].

(d) Controlling for confounding factors

The establishment of an MPA is non-random, generating systematic biases between the characteristics of protected and non-protected areas. The specific attributes of these biases are contingent on the process of MPA establishment itself, the criteria used to identify MPA location and boundaries, decision-making involved, governance and goals [50,55]. MPAs are often designed with the intention of protecting areas of high biodiversity, critical habitats and/or processes that maintain populations and ecosystem stability (i.e. larval sources, spawning aggregations [50,55]). For political, societal and economic reasons, MPAs are often located where the marginal costs of protection are low, owing to characteristics inherent to the area (e.g. far from markets, low population density, difficult to access [56]). This trend can also be observed in terrestrial protected area establishment, as protected areas are more likely to be established where existing uses and potential for extractive purposes are low, establishment is politically feasible and management costs are low [31,56,57]. If MPA placement is biased towards less threatened areas (e.g. little historical exploitation), then most conventional methods (i.e. those that do not explicitly control for observable bias in MPA placement or outcomes) may overestimate the impact of protection [28,31].

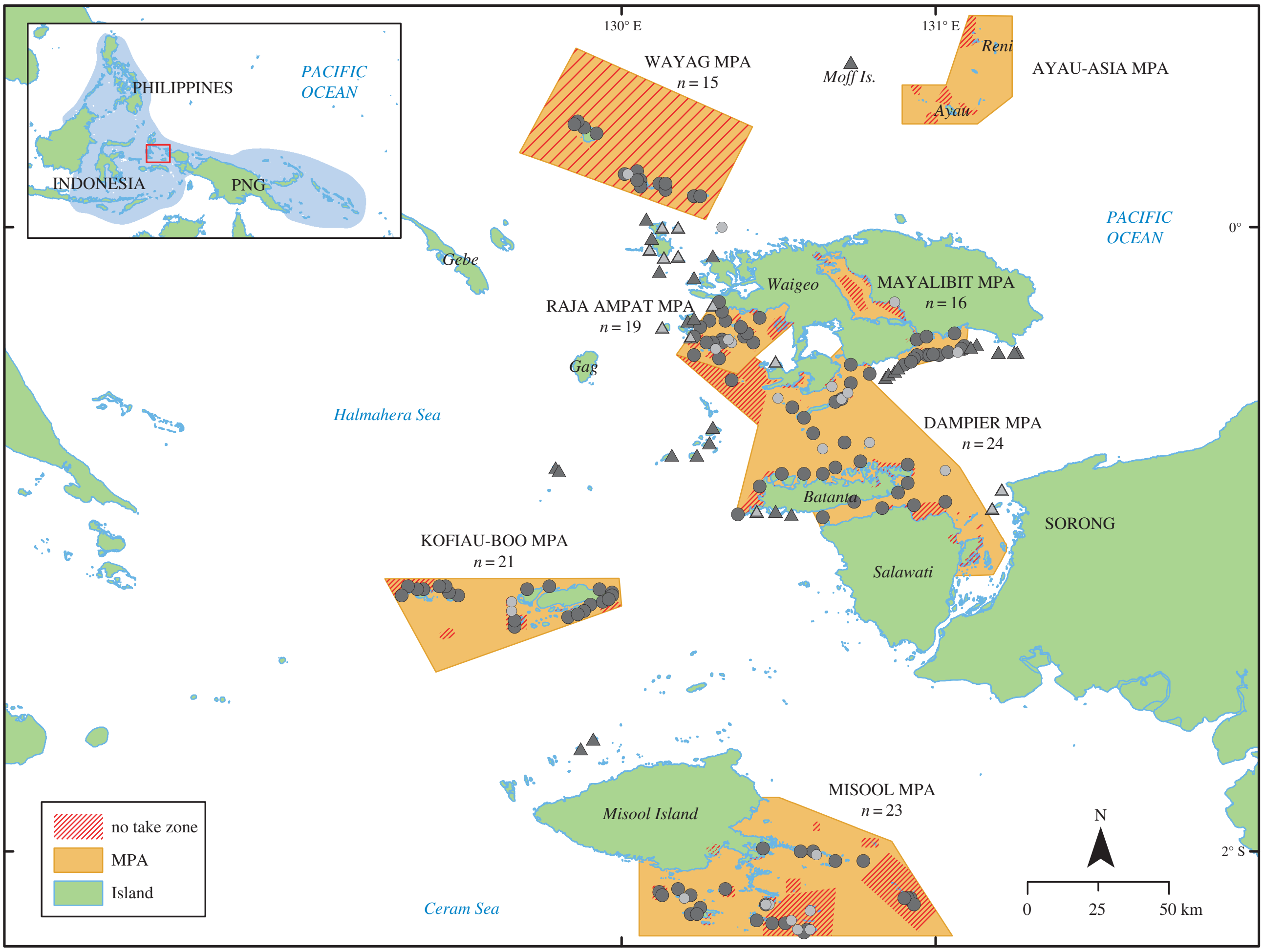

Figure 1. Map of monitoring sites inside and outside of MPAs in the BHS. Triangles indicate sites sampled outside of MPAs ('control') and circles indicate sites sampled inside MPAs ('treatment'). Black shapes were matched and will be included in subsequent impact analyses, while grey shapes are sites that were dropped post-matching and will therefore not be included future impact analyses. Map does not include all MPAs in the region and is limited to MPAs included in the analyses. PNG, Papua New Guinea.

Ecological and social processes link protected areas with their unprotected surroundings. The establishment of conservation interventions can modify the magnitude, direction or variation in these processes, generating spillover effects [58]. In marine ecosystems, ontogenic and adult migratory behaviours may span MPA boundaries, allowing individuals from within an MPA to be observed and/or harvested outside the MPA boundaries [59]. Consequently, dispersal and migratory behaviours can result in spillover effects from MPA to control sites that vary on a species-specific basis [60]. This poses significant challenges for impact evaluation, because causal inference requires independence between treatment (MPA) and control samples. Evaluators can attempt to account for spillover effects by creating a buffer within which sites are not considered to be valid matches to those inside in the immediate vicinity of protected areas, to reduce the likelihood of non-independent MPA and control samples (see [28] for terrestrial example of this approach). In the majority of cases, however, it is not possible to create a buffer that extends beyond the migratory or dispersal range that encompasses all species within an MPA. The potential for spillover effects to bias the magnitude or direction of MPA impacts requires that scholars interpret their findings with caution when examining outcomes for highly mobile species, or ecological processes that could be affected by interactions between MPA and control sites.

3. Bird's Head Seascape case study

(a) Bird's Head Seascape

The Bird's Head Seascape (BHS), located in West Papua, Indonesia, in the heart of the Coral Triangle, has the richest diversity of corals and reef fish species in the world [61–63]. The Seascape also provides critical habitat for migratory species such as turtles, cetaceans and whale sharks [64,65]. While no coastal marine ecosystems there remain pristine, the Seascape's low human population density and relative remoteness have ensured that its coastal marine ecosystems are relatively healthy compared with other parts of Southeast Asia [66]. Over the past decade, considerable investments by government and non-government organizations (NGOs) have sought to protect the globally significant biodiversity in the region, primarily by establishing a network of 12 multi-use MPAs across the Seascape (figure 1). The MPAs range in size from 5000 to 1 453 500 ha and cover a total area of 3 594 702 ha, representing approximately 29% of Indonesia's total MPA estate. The majority of the MPAs were established by local communities, with the support of NGOs through traditional ('*adat*') and Regency legal frameworks and reinforced by national legislation [64]. The BHS MPAs are also recognized under a broader provincial spatial plan for the West Papua ('*Papua Barat*') Province. This study focuses on six of the 12 MPAs in the BHS as these are the MPAs in which the initial

Table 1. BHS MPAs ecological monitoring site summary. The table includes the total number of baseline reef sites monitored in BHS MPAs, the number of sites remaining within each MPA following statistical matching, and the timelines of establishment, effective enforcement and baseline monitoring of each MPA. The second column lists the number of sites in the MPA prior to statistical matching, while the third column lists the number of 'matched' sites from each MPA remaining following statistical matching (some sites within the MPA were dropped if there were no suitable control sites identified); numbers in parentheses indicate how many sites were dropped. The fourth column is the year of effective enforcement in the MPA, which meets the following parameters: (i) patrol is established (both community-based patrol or joint patrol), (ii) prosecution processes and sanctions are happening in the local community and (iii) patrol log/ records are present and used for future enforcement strategy. The fifth column indicates the year the MPA zoning plan was formalized and recognized.

MPA	total no. sites	no. 'matched' sites	year of baseline monitoring	year of effective enforcement	MPA zoning plan formalized
Dampier	28	23 (5)	2012	2012	2013
Kofiau-Boo	19	18 (1)	2010	2011	2012
Mayalibit	12	11 (1)	2012	2011	2012
Misool	24	18 (6)	2011	2012	2012
Raja Ampat	16	16 (0)	2011–2012	n.a.	2009
Wayag	9	9 (0)	2012	2008	2009
Controls	53	31 (22)	2012–2014	n.a.	n.a.

monitoring approach and design were sufficient to adapt into an impact evaluation framework.

(b) Implementation in the Bird's Head Seascape

(i) Monitoring methods and design

The initial ecological monitoring programme developed for the Seascape in 2009 by international NGOs focused on providing insights on the status and trend of coral-reef habitats and fish populations within MPAs. The programme and subsequent protocols were designed to provide salient information on status and trends to MPA managers and others, at relatively low cost, at relevant spatial and temporal scales, and at appropriate levels of statistical power, over the long-term. Well-designed impact evaluation shares many of these attributes (e.g. appropriate statistical power, long time-series [51]), but imposes additional data collection and management requirements to enable quasi-experimental causal inference (e.g. data on observable biases influencing participation in, or outcomes of MPA establishment; outcomes for untreated units [51]).

In 2011, scholars and practitioners recognized the potential to modify the ecological monitoring programme implemented in the BHS to enable quasi-experimental causal inference at the Seascape scale. In effect, the existing performance measurement systems were nested within a broader impact evaluation framework that transformed the BHS MPAs into a replicated set of policy experiments, mirroring an ongoing effort to document the MPAs' social impacts (described in [67]). This nested approach enabled the revised BHS monitoring system [68] to retain its ability to inform adaptive management, while simultaneously creating the opportunity for causal inference.

Considerable monitoring efforts took place from 2009 to 2014 inside both no-take and use zones of the six MPAs, and in 2012, in areas outside of MPAs, to document baseline ecological conditions (fish and benthic attributes) of coral reefs (figure 1 and table 1). Ideally, all MPA sites would have been monitored prior to the intervention. However, baseline data were not always available, because data collection was adopted from an existing monitoring programme without an impact evaluation lens. Initial baseline conditions of most MPAs were monitored either prior to or within one year of the intervention (year in which MPA zoning plans were formalized by the government; table 1). The exceptions are Wayag and Raja Ampat MPAs. In these MPAs, MPA zoning plans were finalized in 2009 but baseline conditions were not monitored until 2012. To address this inconsistency, time since effective enforcement and/or finalization of MPA zoning plans will be taken into consideration in post-hoc analyses. In addition, even without achieving the ideal scenario, we anticipate that inclusion of control sites will be highly informative as a functional baseline for a long-term monitoring programme. Monitoring was done using SCUBA at 10–12 m depth following standard protocols [68]. At all sites, data were collected on environmental conditions (e.g. wave exposure, currents) and general reef characteristics (reef slope, reef type).

Ecological indicators were selected to reflect management goals, inform policy-makers, and be useful as indicators of ecosystem health and fish populations. With the exception of Raja Ampat, goals and objectives were developed for the five other MPAs, which were gazetted as a network under fisheries legislation. Indicators were aligned with the Indonesian MPA Management Assessments [69], including condition of the coral reef and populations of key fisheries species and non-target fish species. Other criteria included characteristics of the ecological indicators (i.e. different trophic and functional groups, life-histories and home-ranges). Taking all of this information into consideration, the following indicators were selected for inclusion: (i) overall biomass of key fisheries species (Lutjanidae, Haemulidae, Serranidae), (ii) biomass of herbivorous fish species (Acanthuridae, Scaridae, Siganidae) and (iii) habitat quality (ratio of hard coral cover to rubble and algae cover) (NB: we do not present hard coral data in this manuscript).

(ii) Matching methods

MPAs in the BHS have been strategically designed and (non-randomly) placed [70]. To avoid observable selection bias, we adopted a quasi-experimental design, and applied a tiered matching approach (coarse matching followed by statistical matching) to identify comparable control sites, which allows us to estimate the counterfactual (i.e. changes in fish biomass that would have occurred if no MPA were established). Coarse matching selected reef areas to survey that were thought to be the most similar to reefs within MPAs; this was based on available reports from rapid expeditions [71] and expert opinion (M. V. Erdmann 2012, personal communication).

Statistical matching was then used in effect to 'reverse engineer' a randomized controlled trial by reducing observable biases (i.e. differences between treatment and control groups arising from non-random assignment), generating a matched set of MPA and non-MPA sites with similar contextual conditions.

The contextual factors or covariates used in the statistical matching model were selected based on published literature in coral-reef ecology, best available data and recommendations from experts who ranked variables most pertinent to BHS reef ecosystems. We chose 10 contextual variables that encompassed structural, biophysical and social features of coral-reef sites that influence ecosystem structure (table 2, and also see the electronic supplementary material). Structural variables included: reef exposure, slope and type; and distance to deep water (50 m isobaths). Biophysical variables included: frequency of sea surface temperature anomalies (SSTAs) in degree-heating-weeks; exposure to either northwest or southeast monsoon winds; and distance to nearest mangrove habitat. Social variables, associated with fishing and resource use, included: distance to nearest fishing settlement; distance to primary market; and pollution risk.

We performed statistical matching procedures in R [72] using the Matching package [73]. We assessed the covariate balance (i.e. the differences between distributions of covariates across treated and control sites) achieved by multiple matching algorithms, including propensity scores and covariate matching. After reviewing covariate balance produced by the different matching algorithms, we selected nearest-neighbour covariate matching using the Mahalanobis distance metric, which produced the smallest mean differences between MPA and control sites.

All covariates were weighted equally. As there were fewer control sites than MPA sites, we matched with replacement (control sites being returned to the pool of potential matches) so that a control site could be matched with multiple MPA sites. We required that each MPA site be matched with exactly two control sites. If an MPA site matched with more than two equally good control sites, two control sites were selected randomly. Approximately 40% of control sites were used 1–2 times, another 40% were used 3–9 times, and 20% of control sites were matched with MPA sites 10–18 times. We investigated increasing the number of control sites required to match with each MPA site from 3 to 10, but this results in reduced covariate balance, meaning the average quality of matched pairs decreases if MPA sites are forced to match with more than two control sites.

To ensure high-quality matches, we imposed restrictions (i.e. calipers) on the maximum difference between matched MPA-control site covariate values [74]. Calipers pose a certain trade-off: overall match quality (i.e. covariate balance) increases when calipers are tightened (i.e. lower maximum differences), but the number of possible matches is reduced. We imposed calipers on: reef slope, reef type and reef exposure because fish and coral-reef communities are often structured along these habitat gradients [75]. For reef slope, 'flat' sites match with other flat sites, but can also match with 'slope' sites. Flat reef sites, however, are not allowed to match with 'wall' sites. Similarly, for reef exposure, 'exposed' sites match with 'exposed' or 'semi-exposed' sites, and 'sheltered' sites match with 'sheltered' or 'semi-exposed' sites. Reef types may match with identical reef types, or atolls may match with barrier reefs, barrier with fringing, and fringing with patch. Additionally, sites with a pollution risk of 1 can match with sites of a pollution risk of 1 or 2, sites with pollution risk of 3 may match with either 3 or 2, but never 1. After multiple iterations considering the trade-offs between dropped MPA sites and achieving covariate balance, the optimal matching model dropped 13 of the 108 MPA sites and 22 of the 53 control sites (table 1). Trade-offs were resolved by selecting the optimal matching model that would ensure the highest quality of matched pairs, without substantially reducing sample size so much as to compromise the ability of post-hoc analyses to detect changes in ecological outcomes.

Quasi-experimental causal inference rests on the assumption that, after statistical matching procedures, systematic differences between treated and untreated units (i.e. observable biases) are negligible. Statistical matching seldom *eliminates* observable bias, but rather reduces it to within acceptable bounds. Typically, scholars employ 'rules of thumb' to assess how well treatment-control pairs match, and specifically if covariate balance was achieved. For example, standardized mean differences of less than 5% are typically considered acceptable for robust causal inference [76]. This threshold, however, may vary in the marine environment for certain covariates. For example, at sites relatively close to fish markets, fish biomass and distance to market are tightly correlated whereas, at relatively far distances, the relationship between biomass and distance to market breaks down [13]. The nonlinearity in the relationship between observed covariates and outcomes, and the prevalence of incomplete or biased datasets, suggests that a threshold of standardized mean differences of less than 5% may not always be meaningful or achievable when applied to marine ecosystems. This would result in MPA practitioners having to accept higher levels of difference between treatment and control groups, and these differences would need to be accounted for in future analytical models that measure MPA impact.

Post-matching covariate balance between BHS MPA and control sites is variable (table 2). Standardized mean differences of less than 5% are achieved for six covariates (reef exposure, type, slope, pollution risk, monsoon direction and SSTA; see standardized mean difference column and the values in 'matched' rows, table 2). There were larger differences between MPA and control sites post-matching for the remaining four covariates (distance to market, fishing settlement, deep water and mangrove habitat), suggesting there still are substantial systematic biases between protected and unprotected sites for these covariates [23]. Matched MPA sites are, on average, further from fish markets, fishing settlements, mangroves and deep water compared to control sites (table 2), reflecting the well-documented biases in the placement of MPAs [56]. In the future, we will specifically account for these biases in the analytical model that measures the effect of MPAs on ecological outcomes and use caution when interpreting the treatment effect.

Preliminary analysis of two ecological outcomes of interest in the BHS—biomass of key fisheries families (Serranidae, Lutjanidae and Haemulidae) and ecologically important

Table 2. Covariate balance pre- and post-matching. Differences were assessed between MPA and control (non-MPA) sites, before and after matching. The first column lists covariates used to match MPA and control sites, and for each covariate match statistics are provided before and after matching, indicated in the 'unmatched' and 'matched' rows, to show how well the matching model performed. The third column presents mean covariate values for MPA sites. The fourth column presents mean covariate values for control sites. The fifth column shows the standardized mean difference between treatment and control means (mean difference between treatment and control sites divided by the standard deviation of the treatment observations multiplied by 100); values near or at zero indicate that sites inside and outside protected areas are very similar to each other. The sixth and seventh columns respectively show mean and maximum differences in each covariate Quantile–Quantile (plot QQ plot), with lower values indicating a better match. The covariates reef exposure, slope and type were split out to ensure that sites match exactly within the calipers we specified. For example, for the covariate 'exposed: semi-exposed', sites were assigned 1 if they were exposed or semi-exposed. We then required standardized mean difference to equal zero post-matching, which guarantees that exposed sites matched with other exposed or semi-exposed sites. DHW, degree heating weeks.

covariate		treatment mean	control mean	std. mean diff	mean eQQ diff.	max eQQ diff.
distance to deep water (m)	*unmatched*	627	656	-3.14	131	2164
(50m depth contour)	*matched*	633	384	26.2	251	4341
SSTA	*unmatched*	25.4	24.5	18.3	2.40	10
	matched	25.3	26	-12.6	0.881	9
exposure	*unmatched*	1	1	0	0	0
(exposed : semi-exposed)	*matched*	1	1	0	0	0
exposure	*unmatched*	0.148	0.334	-53.6	0.188	1
(semi-exposed : sheltered)	*matched*	0.126	0.126	0	0	0
reef slope	*unmatched*	0.925	0.981	-21.0	0.056	1
(flat : slope)	*matched*	1	1	0	0	0
reef slope	*unmatched*	0.910	0.870	13.6	0.038	1
(slope : wall)	*matched*	0.935	0.937	0	0	0
reef type	*unmatched*	0.907	0.85	20.0	0.056	1
(patch : fringing)	*matched*	0.905	0.905	0	0	0
reef type	*unmatched*	0.953	1	-21.9	0.056	1
(fringing : barrier)	*matched*	1	1	0	0	0
reef type	*unmatched*	0.093	0.151	-20.0	0.056	1
(barrier : atoll)	*matched*	0.095	0.095	0	0	0
distance to mangroves (m)	*unmatched*	4618	18439	-267	13846	89428
	matched	4409	3500	17.1	1544	11933
distance to fishing	*unmatched*	35460	27741	21.3	8151	62438
settlement (m)	*matched*	36223	26036	27.2	12139	77385
distance to market (m)	*unmatched*	146809	124152	21.5	36220	809911
	matched	147333	142528	4.41	35144	812645
pollution risk	*unmatched*	1.23	1.37	-32.7	0.170	1
	matched	1.25	1.21	9.14	0.0417	1
monsoon direction (NW, SE)	*unmatched*	0.57	0.36	43.4	0.208	1
	matched	0.568	0.489	16.0	0.0773	1
year	*unmatched*	2012	2012	-17.2	0.792	2
	matched	2012	2012	-28.0	0.794	2

[a]SSTA-Freq is the frequency of sea-surface temperature anomaly (SSTA), which is the number of times over the previous 52 weeks that SST was greater than or equal to 1°C above that week's long-term average value (electronic supplementary material).
[b]Sites were classified as exposed to either the northwest (NW) or southeast (SE) monsoon winds, depending on their placement in relation to nearby islands (electronic supplementary material).

herbivorous fish (Acanthuridae, Scaridae and Siganidae)—shows there is variation among both MPAs and outcomes (figure 2). Since we do not have time-series data collected at control sites yet, our analyses are strictly limited to summarizing baseline differences in biomass between matched MPA and control sites, and not carrying out a complete impact evaluation of MPAs. Ecological data presented here simply provide the baseline conditions that will be used once monitoring is complete to measure the difference-in-difference (i.e. the difference between biomass rate of

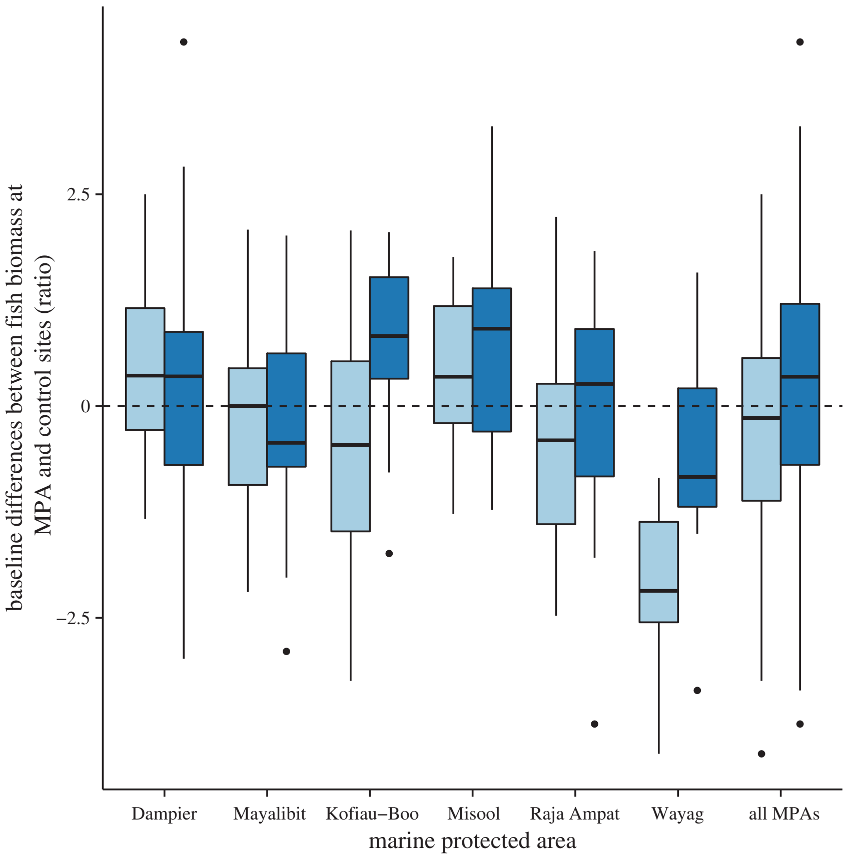

Figure 2. Baseline differences in the fish biomass of key fisheries (dark blue) and functional fish groups (light blue) within MPAs versus outside MPAs in the BHS, shown as ratios of biomass from matched MPA (inside) and control (outside) site pairs. Ratio more than 0 indicate biomass at MPA sites more than control sites. Each boxplot shows the distribution of ratios for individual MPAs; ratios were also pooled to show the overall distribution across the seascape (All MPAs). The shaded box represents the interquartile range; the black line within the shaded box is the median value; whiskers indicate maximum and minimum values excluding outliers; dots represent outliers (more than 1.5 times upper quartile).

change at MPA sites and biomass rate of change at control sites) once repeat monitoring is complete at all sites (a cycle of every 2–3 years). In post-hoc analyses, we will consider baseline fish biomass as a potential factor that could influence recovery within MPAs.

To isolate treatment effects, the potential influence of spillover was considered. This was accounted for in selections of control sites and will also be incorporated in post-hoc analyses. We identified an appropriate buffer of 5 km to account for the home range of most of the species of interest [47]. While most of the control sites are located at distances more than 10 km from MPA boundaries, four non-MPA sites are located less than 5 km from the boundaries of MPAs. However, they are all located more than 10 km from the MPA ‘no-take’ zones, which are expected to have the greatest positive outcomes via potential for spillover effects [5,59,60]. When short-term impacts of MPAs are analysed in future work, distance from MPA boundary will be considered in the model and weighted appropriately for each ecological outcome.

(iii) Repeat monitoring

Efforts to document and explain the ecological impacts of the BHS MPAs are in their early stages. Repeat ecological monitoring will generate longitudinal data in six MPAs by the end of 2015, enabling us to explore the short-term impacts of MPA establishment on habitat condition and the status of key fisheries species. We will document MPA impacts, and examine the ecological and social factors that explain variation in impacts in a series of hierarchical mixed effects models, using the *relative* difference between changes in MPA and control sites over time (described as the treatment effect in econometric literature, or the response ratio in the ecological literature) as our dependent variable. For those observable biases that cannot be adequately controlled by our statistical matching model (e.g. distance to fishing settlement, mangrove habitat and deep water, and year), we account for the magnitude and direction of this bias in our mixed model, in a procedure known as post-hoc regression adjustment [23]. Unlike BACIPs models, where the magnitude and direction of systematic differences between MPA and control groups may not be explicitly incorporated into the analysis, statistical matching procedures enable us to document and correct for observed biases in our estimates of MPA impacts. Statistical matching does not, however, eliminate the potential for unobserved bias (i.e. the presence of a variable that influences MPA placement or outcomes, but that was not included in the matching model). Unobserved bias, by its nature, cannot be detected directly, but sensitivity analysis techniques are available [77] that will allow us to understand the vulnerability of our treatment effects to such bias, should it exist.

As ecological monitoring in the BHS is repeated at relatively high frequencies (i.e. every three years), we will test, and if necessary control for, serial correlation effects (where error terms for a given variable over various time intervals are correlated) [53]. Initial treatment effects, computed 3–4 years post-baseline, are likely to capture a subset of the full suite of MPA ecological impacts, namely those with relatively rapid response times. Existing literature [9] suggests that these short-term treatment effects may detect impacts on aggregate measures, although they may be insufficient to capture impacts that are slower to emerge. For example, peak recovery of fish groups may to vary from 7 years to 37 years [78]. We anticipate that ecological monitoring in the BHS will continue beyond these initial short-term analyses, allowing scholars to document MPA impacts and understand the synergies and trade-offs across domains (e.g. benthic habitats versus fisheries impacts; fisheries versus non-fisheries species; piscivores versus herbivores), and spatial (e.g. between MPAs, between MPA management zones) or temporal scales (e.g. short versus long terms).

4. Discussion

Over the past decade, after rallying cries on the need for rigorous impact evaluation of conservation interventions (e.g. [19]), novel work has generated an emerging evidence base for their application (e.g. [19,79,80]). The application of impact evaluation is, however, constrained by the availability of long-term datasets, in both treatment and control sites, collected under protocols designed to enable causal inference. In some cases, remote-sensing data enable scholars to circumvent the paucity of on-the-ground monitoring (e.g. [81]) to understand the conservation impacts related to specific subsets of outcomes (e.g. forest cover) at broad scales. For many policy-relevant outcomes and interventions, however, remotely sensed indicators are uninformative [75]. Consequently, evidence-based conservation faces an urgent need to expand the scope of impact evaluations to include outcomes of interventions that are not easily evaluable using secondary or remotely sensed data [39]. This paper presents one possible pragmatic solution, namely the integration of impact evaluation into the ongoing monitoring efforts of a large-scale conservation intervention.

In the BHS, existing monitoring systems have been modified to nest the conventional ‘performance measurement’ MPA monitoring needed for adaptive management [22] within a quasi-experimental framework designed to provide robust evidence for long-term MPA impacts (or lack thereof). Ecological outcomes indicators shared across performance measurement and impact evaluation systems ensure that insights are salient for potential adaptive management and can inform local policy. At the same time, integrating the information needed to support causal inference into routine monitoring ensures that data-collection efforts are ongoing and sustained. We anticipate that both in the near-term (our initial impact data of +3 years) and in the longer-term (over the next decade or more), these fine-resolution time-series datasets will allow scholars to explore the temporal and spatial patterns of MPA impacts, as well as document variation across key fisheries species and functional fish groups.

With productive marine habitats and populations declining from a number of causes [2,48,66], understanding the impacts of interventions aimed at preserving the ecosystem services that flow from those habitats and populations is key for both science and management. Without the counterfactual provided by impact evaluation, seeing no change or a decline in outcomes before and after monitoring could lead to an erroneous conclusion about the effectiveness of management, if (as is often the case) the areas outside MPAs are in steeper decline [48]. While it can be difficult to argue for spending money on monitoring sites outside the areas where the interventions take place [39], in fact doing so can inform the efficacy of the (vastly larger) funding for those interventions. In the BHS, streamlining monitoring focused around impact evaluation gave a much greater power to detect change, for the same total monitoring budget.

Integrating impact evaluation techniques into the monitoring of the BHS MPA network posed both technical and logistical challenges. While the BHS experience highlights appropriate methodologies for handling some of these challenges, it also provides insights into improving the design of future MPA impact evaluation studies. In the BHS, an existing and ongoing large-scale ecological monitoring programme already included appropriate and meaningful outcome indicators. Thus, the primary challenge was to control for confounding factors in order to make causal inference between the establishment of the MPAs and an array of ecological outcomes. Because this challenge entailed adapting an existing monitoring programme into an impact evaluation framework, many processes were carried out post-hoc of monitoring design, resulting in dropping many of the monitoring sites from the analyses. Future considerations to maximize use of monitoring data would include more robust coarse matching procedures to select inside/outside sites (i.e. collecting biophysical and social data prior to selecting areas for potential monitoring). Another consideration when designing an impact evaluation study is to ensure that the scale and rigour of the monitoring programme are sufficient for the sample size to have the power to detect changes in a dynamic ecosystem within the time frame that matches the scope/goals of the monitoring programme, while also being able to adapt to other unforeseen changes (i.e. such as expansion of MPA boundaries or changes in zonation). Not without its limitations, the BHS process did demonstrate that ‘on-the-ground’ impact evaluation can successfully be implemented in an MPA monitoring programme.

Beyond advancing theory and answering scientific questions around MPA impacts, data collected for impact evaluation can also directly inform management on the ground. Much of the initial ecological baseline monitoring informed design of the BHS MPA network and the no-take zones within MPAs (J. Wilson 2012, personal communication). We anticipate that the value of this baseline data, now combined with control site data, will only increase with time and as more geographies use comparable monitoring methods and well-designed sampling. The approaches in the BHS have also been replicated in the Sunda Banda Seascape of Indonesia; working across these seascapes has provided a network to facilitate learning and shared insights. While impact evaluation is not necessary or practical in many cases, it is in general underused [19]. To support policy-makers, researchers and practitioners who might draw insights from, or adopt the methods developed in these seascapes, the data products and methods are available to others (www.mpamystery.org). Our quasi-experimental approach will enable evaluating the impact of the conservation interventions in the BHS, bridging the scale mismatches between evidence and decisions by providing information at a scale relevant to management in the region. As the BHS impact evaluation study progresses, we anticipate these monitoring efforts will generate novel insights to better understand when, where and why MPAs lead to positive or negative impacts.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material. R Code available on request from L.G.

Authors' contributions. G.N.A., L.G. and H.E.F. conceived and designed the project; G.N.A., L.G., H.E.F., N.I.H., M.P., P. and S.M. collected the data; G.N.A., L.G. and M.P. analysed the data; L.G. and M.P. developed the R code; all authors wrote the paper.

Competing interests. We have no competing interests.

Funding. G.N.A., L.G., H.E.F., N.I.H., P. and S.M. are all supported by the Walton Family Foundation. Additional support for G.N.A., L.G., H.E.F. and M.P. is from Margaret A. Cargill Foundation. D.G. is supported by Luc Hoffmann Institute and the National Socio-Environmental Synthesis Center (SESYNC).

Acknowledgements. We would like to thank C. N. Ashcroft for her input on early drafts of this manuscript, J. R. Wilson and M. V. Erdmann for contributing to discussions and sharing their knowledge of the BHS, F. Pakiding, Universitas Papua, H. Dennis and M. Wang for helping to gather covariate data and E. Selig for insightful input on covariate selection. We would also like to acknowledge the contribution of M. B. Mascia to efforts to evaluate the impacts of the BHS MPAs. This is contribution #7 of the research initiative *Solving the Mystery of MPA Performance*.

Endnote

[1]Using standard impact evaluation terminology, the *impact* of the intervention/MPA can be defined as the difference between the ecological *outcomes* for those receiving the 'treatment' (or sites within an MPA) and those in the control group (outside the MPA). An *outcome* is the change in the variable of interest over the period of the MPA. The *counterfactual* is what would have happened in the absence of an intervention.

References

1. Bellwood DR, Hughes TP, Folke C, Nystrom M. 2004 Confronting the coral reef crisis. *Nature* **429**, 827–833. (doi:10.1038/nature02691)
2. Halpern BS *et al.* 2008 A global map of human impact on marine ecosystems. *Science* **319**, 948–952. (doi:10.1126/science.1149345)
3. Hughes TP *et al.* 2003 Climate change, human impacts, and the resilience of coral reefs. *Science* **301**, 929–933. (doi:10.1126/science.1085046)
4. Tundi Agardy M. 1994 Advances in marine conservation: the role of marine protected areas. *Trends Ecol. Evol*. **9**, 267–270. (doi:10.1016/0169-5347(94)90297-6)
5. Gaines SD, White C, Carr MH, Palumbi SR. 2010 Designing marine reserve networks for both conservation and fisheries management. *Proc. Natl Acad. Sci. USA* **107**, 18 286–18 293. (doi:10.1073/pnas.0906473107)
6. Thomas HL, Macsharry B, Morgan L, Kingston N, Moffitt R, Stanwell-Smith D, Wood L. 2014 Evaluating official marine protected area coverage for Aichi Target 11: appraising the data and methods that define our progress. *Aquat. Conserv. Mar. Freshwater Ecosyst*. **24**, 8–23. (doi:10.1002/aqc.2511)
7. Toonen RJ *et al.* 2013 One size does not fit all: the emerging frontier in large-scale marine conservation. *Mar. Pollut. Bull*. **77**, 7–10. (doi:10.1016/j.marpolbul.2013.10.039)
8. Claudet J, Pelletier D, Jouvenel J-Y, Bachet F, Galzin R. 2006 Assessing the effects of marine protected area (MPA) on a reef fish assemblage in a northwestern Mediterranean marine reserve: Identifying community-based indicators. *Biol. Conserv*. **130**, 349–369. (doi:10.1016/j.biocon.2005.12.030)
9. Halpern BS. 2003 The impact of marine reserves: do reserves work and does size matter? *Ecol. Appl*. **13**, 117–137. (doi:10.1890/1051-0761(2003)013[0117:TIOMRD]2.0.CO;2)
10. Lester SE, Halpern BS, Grorud-Colvert K, Lubchenco J, Ruttenberg BI, Gaines SD, Airamé S, Warner RR. 2009 Biological effects within no-take marine reserves: a global synthesis. *Mar. Ecol. Progr. Ser*. **384**, 33–46. (doi:10.3354/meps08029)
11. Gaston KJ, Jackson SF, Cantú-Salazar L, Cruz-Piñón G. 2008 The ecological performance of protected areas. *Annu. Rev. Ecol. Evol. Syst*. **39**, 93–113. (doi:10.1146/annurev.ecolsys.39.110707.173529)
12. Halpern BS, Warner RR. 2002 Marine reserves have rapid and lasting effects. *Ecol. Lett*. **5**, 361–366. (doi:10.1046/j.1461-0248.2002.00326.x)
13. Cinner JE, Graham NA, Huchery C, Macneil MA. 2013 Global effects of local human population density and distance to markets on the condition of coral reef fisheries. *Conserv. Biol*. **27**, 453–458. (doi:10.1111/j.1523-1739.2012.01933.x)
14. Edgar GJ *et al.* 2014 Global conservation outcomes depend on marine protected areas with five key features. *Nature* **506**, 216–220. (doi:10.1038/nature13022)
15. Grober-Dunsmore R, Frazer T, Lindberg W, Beets J. 2007 Reef fish and habitat relationships in a Caribbean seascape: the importance of reef context. *Coral Reefs* **26**, 201–216. (doi:10.1007/s00338-006-0180-z)
16. Hargreaves-Allen V, Mourato S, Milner-Gulland E. 2011 A global evaluation of coral reef management performance: are MPAs producing conservation and socio-economic improvements? *Environ. Manag*. **47**, 684–700. (doi:10.1007/s00267-011-9616-5)
17. Horigue V, Aliño PM, Pressey RL. 2014 Evaluating management performance of marine protected area networks in the Philippines. *Ocean Coast. Manag*. **95**, 11–25. (doi:10.1016/j.ocecoaman.2014.03.023)
18. Ferraro PJ. 2009 Counterfactual thinking and impact evaluation in environmental policy. *New Dir. Eval*. **2009**, 75–84. (doi:10.1002/ev.297)
19. Ferraro PJ, Pattanayak SK. 2006 Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biol*. **4**, e105. (doi:10.1371/journal.pbio.0040105)
20. Claudet J, Guidetti P. 2010 Improving assessments of marine protected areas. *Aquat. Conserv. Mar. Freshwater Ecosyst*. **20**, 239–242. (doi:10.1002/aqc.1087)
21. Claudet J *et al.* 2010 Marine reserves: fish life history and ecological traits matter. *Ecol. Appl*. **20**, 830–839. (doi:10.1890/08-2131.1)
22. Mascia MB *et al.* 2014 Commonalities and complementarities among approaches to conservation monitoring and evaluation. *Biol. Conserv*. **169**, 258–267. (doi:10.1016/j.biocon.2013.11.017)
23. Rosenbaum PR. 2010 Design sensitivity and efficiency in observational studies. *J. Amer. Stat. Assoc*. **105**, 692–702. (doi:10.1198/jasa.2010.tm09570)
24. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CM. 2011 *Impact evaluation in practice*. Washington, DC: World Bank Publications.
25. Sekhon JS. 2009 Opiates for the matches: matching methods for causal inference. *Annu. Rev. Political Sci*. **12**, 487–508. (doi:10.1146/annurev.polisci.11.060606.135444)
26. Prendergast J, Quinn R, Lawton J, Eversham B, Gibbons D. 1993 Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337. (doi:10.1038/365335a0)
27. Joppa LN, Pfaff A. 2011 Global protected area impacts. *Proc. R. Soc. B* **278**, 1633–1638. (doi:10.1098/rspb.2010.1713)
28. Andam KS, Ferraro PJ, Pfaff A, Sanchez-Azofeifa GA, Robalino JA. 2008 Measuring the effectiveness of protected area networks in reducing deforestation. *Proc. Natl Acad. Sci. USA* **105**, 16 089–16 094. (doi:10.1073/pnas.0800437105)
29. Heckman JJ, Robb RJr. 1985 Alternative methods for evaluating the impact of interventions: an overview. *J. Econ*. **30**, 239–267. (doi:10.1016/0304-4076(85)90139-3)
30. Heckman JJ, Ichimura H, Todd P. 1998 Matching as an econometric evaluation estimator. *Rev. Econ. Stud*. **65**, 261–294. (doi:10.1111/1467-937X.00044)
31. Joppa LN, Pfaff A. 2009 High and far: biases in the location of protected areas. *PLoS ONE* **4**, e8273. (doi:10.1371/journal.pone.0008273)
32. Andam KS, Ferraro PJ, Sims KR, Healy A, Holland MB. 2010 Protected areas reduced poverty in Costa

Rica and Thailand. *Proc. Natl Acad. Sci. USA* **107**, 9996 – 10 001. (doi:10.1073/pnas.0914177107)

33. Gaveau DL, Epting J, Lyne O, Linkie M, Kumara I, Kanninen M, Leader-Williams N. 2009 Evaluating whether protected areas reduce tropical deforestation in Sumatra. *J. Biogeogr*. **36**, 2165 – 2175. (doi:10.1111/j.1365-2699.2009.02147.x)
34. Somanathan E, Prabhakar R, Mehta BS. 2009 Decentralization for cost-effective conservation. *Proc. Natl Acad. Sci. USA* **106**, 4143 – 4147. (doi:10.1073/pnas.0810049106)
35. Arriagada RA, Ferraro PJ, Sills EO, Pattanayak SK, Cordero-Sancho S. 2012 Do payments for environmental services affect forest cover? A farm-level evaluation from Costa Rica. *Land Econ*. **88**, 382 – 399. (doi:10.3368/le.88.2.382)
36. Weber JG, Sills EO, Bauch S, Pattanayak SK. 2011 Do ICDPs work? An empirical evaluation of forest-based microenterprises in the Brazilian Amazon. *Land Econ*. **87**, 661 – 681. (doi:10.3368/le.87.4.661)
37. Gurney GG, Cinner J, Ban NC, Pressey RL, Pollnac R, Campbell SJ, Tasidjawa S, Setiawan F. 2014 Poverty and protected areas: an evaluation of a marine integrated conservation and development project in Indonesia. *Glob. Environ. Change* **26**, 98 – 107. (doi:10.1016/j.gloenvcha.2014.04.003)
38. Ferraro PJ, Hanauer MM, Miteva DA, Nelson JL, Pattanayak SK, Nolte C, Sims KRE. 2015 Estimating the impacts of conservation on ecosystem services and poverty by integrating modeling and evaluation. *Proc. Natl Acad. Sci. USA* **112**, 7420 – 7425. (doi:10.1073/pnas.1406487112)
39. Miteva DA, Pattanayak SK, Ferraro PJ. 2012 Evaluation of biodiversity policy instruments: what works and what doesn't? *Oxf. Rev. Econ. Policy* **28**, 69 – 92. (doi:10.1093/oxrep/grs009)
40. Fox HE *et al*. 2012 Reexamining the science of marine protected areas: linking knowledge to action. *Conserv. Lett*. **5**, 1 – 10. (doi:10.1111/j.1755-263X.2011.00207.x)
41. Cook CN, Carter RB, Fuller RA, Hockings M. 2012 Managers consider multiple lines of evidence important for biodiversity management decisions. *J. Environ. Manag*. **113**, 341 – 346. (doi:10.1016/j.jenvman.2012.09.002)
42. Mascia MB, CLAUS C, Naidoo R. 2010 Impacts of marine protected areas on fishing communities. *Conserv. Biol*. **24**, 1424 – 1429. (doi:10.1111/j.1523-1739.2010.01523.x)
43. Sanchirico JN. 2000 *Marine protected areas as fishery policy: a discussion of potential costs and benefits*. Discussion Paper, 00 – 23. Washington, DC: Resources for the Future.
44. García-Charton J, Pérez-Ruzafa A, Sánchez-Jerez P, Bayle-Sempere J, Reñones O, Moreno D. 2004 Multi-scale spatial heterogeneity, habitat structure, and the effect of marine reserves on Western Mediterranean rocky reef fish assemblages. *Mar. Biol*. **144**, 161 – 182. (doi:10.1007/s00227-003-1170-0)
45. Edmunds PJ, Bruno JF. 1996 The importance of sampling scale in ecology: kilometer-wide variation in coral reef communities. *Mar. Ecol. Progr. Ser. Oldendorf* **143**, 165 – 171. (doi:10.3354/meps143165)
46. Schutte VG, Selig ER, Bruno JF. 2010 Regional spatio-temporal trends in Caribbean coral reef benthic communities. *Mar. Ecol. Progr. Ser*. **402**, 115 – 122. (doi:10.3354/meps08438)
47. Green AL, Maypa AP, Almany GR, Rhodes KL, Weeks R, Abesamis RA, Gleason MG, Mumby PJ, White AT. 2014 Larval dispersal and movement patterns of coral reef fishes, and implications for marine reserve network design. *Biol. Rev*. (doi:10.1111/brv.12155)
48. Selig ER, Bruno JF. 2010 A global analysis of the effectiveness of marine protected areas in preventing coral loss. *PLoS ONE* **5**, e9278. (doi:10.1371/journal.pone.0009278)
49. Willis T, Millar R, Babcock R, Tolimieri N. 2003 Burdens of evidence and the benefits of marine reserves: putting Descartes before des horse? *Environ. Conserv*. **30**, 97 – 103. (doi:10.1017/S0376892903000092)
50. Roberts CM *et al*. 2003 Ecological criteria for evaluating candidate sites for marine reserves. *Ecol. Appl*. **13**, 199 – 214. (doi:10.1890/1051-0761(2003)013[0199:ECFECS]2.0.CO;2)
51. Rosenbaum PR. 2010 *Design of observational studies*. New York, NY: Statistics SSi.
52. Underwood A. 1992 Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *J. Exp. Mar. Biol. Ecol*. **161**, 145 – 178. (doi:10.1016/0022-0981(92)90094-Q)
53. Stewart-Oaten A, Murdoch WW, Parker KR. 1986 Environmental impact assessment: 'Pseudoreplication' in time? *Ecology* **67**, 929 – 940. (doi:10.2307/1939815)
54. Osenberg CW, Shima JS, Miller SL, Stier AC. 2011 *Assessing effects of marine protected areas: confounding in space and possible solutions. Marine protected areas: a multidisciplinary approach*, p. 143. Cambridge, UK: Cambridge University Press.
55. Stewart RR, Possingham HP. 2005 Efficiency, costs and trade-offs in marine reserve system design. *Environ. Model. Assess*. **10**, 203 – 213. (doi:10.1007/s10666-005-9001-y)
56. Devillers R, Pressey RL, Grech A, Kittinger JN, Edgar GJ, Ward T, Watson R. 2014 Reinventing residual reserves in the sea: are we favouring ease of establishment over need for protection? *Aquat. Conserv. Mar. Freshwater Ecosyst*. **25**, 480 – 504. (doi:10.1002/aqc.2445)
57. Loucks C, Ricketts TH, Naidoo R, Lamoreux J, Hoekstra J. 2008 Explaining the global pattern of protected area coverage: relative importance of vertebrate biodiversity, human activities and agricultural suitability. *J. Biogeogr*. **35**, 1337 – 1348. (doi:10.1111/j.1365-2699.2008.01899.x)
58. Ewers RM, Rodrigues AS. 2008 Estimates of reserve effectiveness are confounded by leakage. *Trends Ecol. Evol*. **23**, 113 – 116. (doi:10.1016/j.tree.2007.11.008)
59. Roberts CM, Sargant H. 2002 Fishery benefits of fully protected marine reserves: why habitat and behavior are important. *Nat. Resour. Model*. **15**, 487 – 507. (doi:10.1111/j.1939-7445.2002.tb00099.x)
60. Abesamis RA, Russ GR. 2005 Density-dependent spillover from a marine reserve: long-term evidence. *Ecol. Appl*. **15**, 1798 – 1812. (doi:10.1890/05-0174)
61. Veron J, Devantier LM, Turak E, Green AL, Kininmonth S, Stafford-Smith M, Peterson N. 2009 Delineating the coral triangle. *Galaxea J. Coral Reef Stud*. **11**, 91 – 100. (doi:10.3755/galaxea.11.91)
62. Allen GR, Erdmann MV. 2009 Reef fishes of the bird's head peninsula, West Papua, Indonesia. *Check List* **5**, 587 – 628.
63. Allen G, Erdmann M. 2012 *Reef fishes of the East Indies,* vol. I – III, pp. 1 – 1260. Perth, Australia: Tropical Reef Research.
64. Mangubhai S *et al.* 2012 Papuan Bird's Head Seascape: emerging threats and challenges in the global center of marine biodiversity. *Mar. Pollut. Bull*. **64**, 2279 – 2295. (doi:10.1016/j.marpolbul.2012.07.024)
65. Ender AI, Mangubhai S, Wilson JR, Muljadi A. 2014 Cetaceans in the global centre of marine biodiversity. *Mar. Biodivers. Rec*. **7**, e18. (doi:10.1017/S1755267214000207)
66. Burke LM, Reytar K, Spalding M, Perry A. 2011 *Reefs at risk revisited*. Washington, DC: World Resources Institute.
67. Glew L, Mascia MB, Pakiding F. 2012 *Solving the mystery of MPA performance*. Washington, DC: Field manual.
68. Ahmadia GN, Wilson JR, Green AL. 2013 Coral reef monitoring protocol for assessing marine protected areas in the coral triangle. Coral Triangle Support Partnership.
69. Richards BL, Williams ID, Vetter OJ, Williams GJ. 2012 Environmental factors affecting large-bodied coral reef fish assemblages in the Mariana Archipelago. *PLoS ONE* **7**, e31374. (doi:10.1371/journal.pone.0031374)
70. Grantham HS *et al.* 2013 A comparison of zoning analyses to inform the planning of a marine protected area network in Raja Ampat, Indonesia. *Mar. Policy* **38**, 184 – 194. (doi:10.1016/j.marpol.2012.05.035)
71. DeVantier LM, Turak E, Allen GR. 2009 *Reef-scapes, reef habitats and coral communities of Raja Ampat, Bird's Head Seascape, Papua, Indonesia*. Bali, Indonesia: The Nature Conservancy.
72. Team RC. 2012 *R: a language and environment for statistical computing*. Vienna, Austria: the R Foundation for Statistical Computing.
73. Sekhon JS. 2011 Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J. Stat. Software* **42**, 1 – 52.
74. Rosenbaum PR, Rubin DB. 1985 Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Stat*. **39**, 33 – 38.
75. Gust N, Choat JH, McCormick MI. 2001 Spatial variability in reef fish distribution, abundance, size and biomass: a multi-scale analysis. *Mar. Ecol. Progr. Ser*. **214**, 237 – 251. (doi:10.3354/meps214237)
76. Caliendo M, Kopeinig S. 2008 Some practical guidance for the implementation of propensity

score matching. *J. Econ. Surv*. **22**, 31–72. (doi:10.1111/j.1467-6419.2007.00527.x)

77. Rosenbaum PR. 2007 Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. *Biometrics* **63**, 456–464. (doi:10.1111/j.1541-0420.2006.00717.x)
78. McClanahan T, Graham N. 2005 Recovery trajectories of coral reef fish assemblages within Kenyan marine protected areas. *Mar. Ecol. Progr. Ser*. **294**, 241–248. (doi:10.3354/meps294241)
79. Sutherland WJ, Pullin AS, Dolman PM, Knight TM. 2004 The need for evidence-based conservation. *Trends Ecol. Evol*. **19**, 305–308. (doi:10.1016/j.tree.2004.03.018)
80. Fisher B, Balmford A, Ferraro PJ, Glew L, Mascia M, Naidoo R, Ricketts TH. 2014 Moving Rio forward and avoiding 10 more years with little evidence for effective conservation policy. *Conserv. Biol*. **28**, 880–882. (doi:10.1111/cobi.12221)
81. Ferraro PJ, Hanauer MM, Miteva DA, Canavire-Bacarreza GJ, Pattanayak SK, Sims KR. 2013 More strictly protected areas are not necessarily more protective: evidence from Bolivia, Costa Rica, Indonesia, and Thailand. *Environ. Res. Lett*. **8**, 025011. (doi:10.1088/1748-9326/8/2/025011)