Received: 1 December 2019 | Accepted: 5 October 2020

DOI: 10.1002/pan3.10162

CONSUMING WILDLIFE – MANAGING DEMAND FOR PRODUCTS IN THE WILDLIFE TRADE

Research Article

Challenges in the impact evaluation of behaviour change interventions: The case of sea turtle meat and eggs in São Tomé

Laura Thomas-Walters[1] | Sara Vieira[2] | Victor Jiménez[3] | Domingas Monteiro[2] | Betânia Ferreira[3] | Robert J. Smith[1] | Diogo Veríssimo[4,5]

[1]Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, UK

[2]Programa Tatô, Largo do Bom Despacho, Cidade de São Tomé, São Tomé, Sao Tome and Principe

[3]Programa Tatô, Sítio da Pedragosa, Barão de São João, Portugal

[4]Department of Zoology, University of Oxford, Oxford, UK

[5]San Diego Zoo Institute for Conservation Research, Escondido, CA, USA

Correspondence
Laura Thomas-Walters
Email: lt35@stir.ac.uk

Funding information
Rufford Foundation, Grant/Award Number: 18821-1; Lisbon Oceanarium; U.S. Fish and Wildlife Service; University of Kent Research Alumni Scholarship; Oxford Martin Programme

Handling Editor: Rachel Tiller

Abstract

1. Robust impact evaluations are needed for conservation to learn and grow as a field. Currently we lack a large body of evidence on the effects of behaviour change interventions in social-ecological systems. By uncovering mechanistic relationships and establishing causality we can refine future programmes to enhance likelihood of effectiveness. Although a range of sophisticated methodological approaches to evaluation have been developed, conceptually linking project outcomes with conservation impacts remains difficult in complex systems.
2. For example, sea turtles are one taxon in which unsustainable harvesting has been a particular problem. There have been a number of campaigns to reduce demand for sea turtle products, but we still have little evidence documenting their outcomes. We conducted the first formal impact evaluation of a conservation marketing campaign aimed at reducing the consumption of sea turtle meat and eggs. The campaign took place on the island of São Tomé, Central Africa, and included traditional mass media advertisements as well as community events. This is one of few demand reduction evaluations that has assessed both human behaviour and biological conservation outcomes. It benefited from an advantageous setting for evaluation, as nesting sea turtles are relatively easy to monitor and the island's small size and geographic isolation increased the detectability of potentially confounding factors. We found a decrease in self-reported sea turtle egg consumption and a decrease in poaching of adult sea turtles.
3. However, multiple unforeseen difficulties arose which complicated attempts at causal attribution. We were hampered by spatial spillovers, design effects from the sensitive questioning technique, concurring law enforcement and changes in biological monitoring effort. These challenges highlight the difficulties faced by practitioners seeking to apply impact evaluations in the field. We reflect on what

this means for future impact evaluations of behaviour change interventions in conservation. Our recommendations include combining multiple outcome measures to triangulate hard-to-measure behaviours and theory-based evaluation methods to explore causal impacts.

KEYWORDS

aquatic, behaviour change, bushmeat, conservation social science, consumer research, demand reduction, illegal behaviour, marketing, São Tomé and Príncipe, unmatched count technique, wildlife trade, wildmeat

1 | INTRODUCTION

In a world of limited resources and pressing environmental problems, improving the efficiency and effectiveness of conservation interventions is vital (Baylis et al., 2016). We need a robust body of evidence documenting the impact of different programmes that we can draw on when designing future projects. In recognition of this, there has been a large-scale movement towards the synthesis of scientific results through tools such as systematic reviews or repositories like Conservation Evidence (www.conservationevidence.com; Dicks et al., 2014; Sutherland & Wordley, 2017). However, these are only as useful as the studies they draw on, which are often of poor quality or have a narrow focus (see, for example: Baylis et al., 2016; Josefsson et al., 2020; Thomas-Walters et al., 2019; Veríssimo & Wan, 2018). A rigourous impact evaluation should provide credible evidence by using an appropriate counterfactual to establish causal attribution (Adams et al., 2019; Ferraro, 2009). This can be accomplished through a range of methodological approaches. Randomised control trials (RCTs), for example, are considered by some to be the gold standard for determining if there is a cause–effect relationship between an intervention and any outcomes (Backmann, 2017; Kvangraven, 2020). They are not always feasible or desirable to implement however, and researchers have also developed a variety of alternative quasi-experimental designs that use techniques such as propensity score matching or regression discontinuity to control for observed and unobserved covariates (Ferraro, 2009; Margoluis et al., 2009; Schleicher et al., 2019).

Unfortunately, using these techniques to uncover mechanisms and establishing causality can be particularly difficult in socio-ecological systems (Preskill, 2009; Veríssimo et al., 2017). Many conservation issues, from the illegal wildlife trade to climate change are 'wicked' problems, characterised by their complexity, dynamism and intractability (Knight et al., 2019). This greatly complicates evaluations, which may have to deal with uncontrollable variables, multiple spatial and temporal scales and ecological thresholds leading to non-linear change (Hildén, 2009; Margoluis et al., 2009). Furthermore, while the stated goal may be to protect wild species or ecosystems, interventions frequently need to target human behaviours (Amel et al., 2017). Evaluators are then faced with the challenge of extrapolating from project outcomes (e.g. compliance with sustainable fishery management practices) to actual conservation impacts (e.g. recovery of fish populations; McDonald et al., 2020). This requires interdisciplinary expertise, with knowledge of both social and biological processes and a detailed understanding of conceptual linkages within complex systems (Ferraro et al., 2019; Margoluis et al., 2009).

While there has been considerable research into both impact evaluation best practices and barriers to conducting evaluations, there is little guidance available for practitioners in how to deal with difficulties that may arise in the field (Curzon & Kontoleon, 2016; McIntosh, 2019). Here we describe a case study of an evaluation of a behaviour change intervention project that faced multiple challenges, and reflect on what we as academics and practitioners can learn from the experience.

2 | CASE STUDY: SEA TURTLE CONSUMPTION IN SÃO TOMÉ

2.1 | Background

We conducted the first formal impact evaluation of a sea turtle demand reduction campaign, and one of the few demand reduction evaluations that has assessed both human behaviour and biological conservation outcomes (Veríssimo & Wan, 2018). Demand for wildlife products is a key driver of the exploitation of wild populations of flora and fauna (Rosen & Smith, 2010). The dominant approach to combating the illegal wildlife trade has been to restrict the supply of wildlife products, through interventions such as trade bans and anti-poaching measures (Phelps et al., 2014). However, there is now an increasing focus on demand-side interventions, with the aim of reducing the market value of illegal wildlife products by influencing consumers to voluntarily change their purchasing behaviour (Thomas-Walters, Veríssimo, et al., 2020; Veríssimo & Wan, 2018). One biological group in which unsustainable commercial trade has been a particular problem is sea turtles (Donlan et al., 2010; Frazier, 2003). Due in part to this harvest they have experienced extirpations and population declines in coastal areas globally (e.g. Mancini & Koch, 2009; Nada & Casale, 2011). The illegal trade is driven by widespread demand for sea turtle meat and eggs (Campbell, 2003), and so demand reduction interventions could potentially play an important conservation role. Yet despite multiple campaigns about the plight of sea turtles dating back at least two

decades, there have been no published evaluations of their effectiveness (Graff, 1996).

São Tomé is an island off the coast of Central Africa where sea turtles have experienced severe poaching pressure. There are four species known to nest here, each threatened with extinction according to the IUCN Red List of Threatened Species (Castroviejo et al., 1994). In 2014, the Santomean government approved national legislation (Decreto-Lei n. 8/2014, of 28 April), criminalising the possession, trade and transportation of sea turtles (Vieira et al., 2016). Consumer research shows a large-scale demand for sea turtle meat in both rural and urban communities, and a high demand for sea turtle eggs in rural communities. Sea turtle meat is seen as a delicacy by many residents, and is easily accessible (Veríssimo et al., 2020). In 2003 the non-governmental organisation Mar, Ambiente e Pesca Artesanal (MARAPA) established a conservation project called Programa Tatô (https://www.programatato.org/) to monitor sea turtle populations and aid in their conservation by assisting with law enforcement. This has been an independent organisation since 2018.

Starting in 2016, Programa Tatô began an on-going conservation marketing campaign aimed at reducing the consumption of sea turtle meat and eggs in São Tomé. The campaign was called Tataluga—Mém Di Omali, which means 'Sea Turtle—the mother of the sea' in the local Forrô dialect. Its design was informed by a study of consumer profiles, motivations and preferences conducted in 2015 (Veríssimo et al., 2020). The campaign featured community events, such as cooking contests to promote alternative food products, theatre performances and an association football championship and mass media components such as billboard advertisements and television and radio jingles (Figure 1; full list of activities in Supporting Information S1). The campaign brand aimed to associate sea turtles with motherhood, hoping to stimulate an emotional bond between the local people and sea turtles. The overall aim was to lessen poaching pressure on sea turtle populations by changing the social norms and attitudes of people living alongside the sea turtles and thus reducing consumption behaviours.

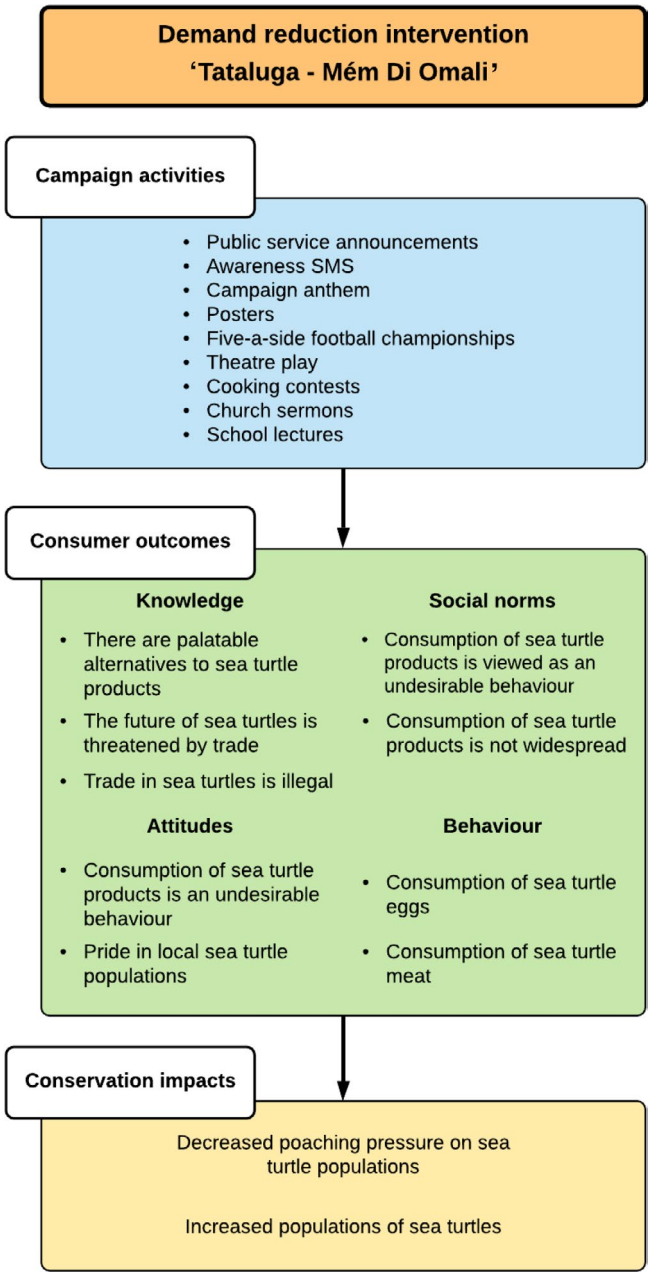

FIGURE 1 'Tataluga - Mém Di Omali' programme logic model, demonstrating the conceptualised pathway for the expected impact of the campaign activities on sea turtle populations

2.2 | Evaluation design

We initially designed the study as a BACI (before-after control-intervention) experiment, with assignment to treatment group occurring at village level. Five coastal fishing communities were assigned to receive the intervention, and a further three served as comparisons (demographic details in Supporting Information S2). All the villages were surveyed prior to the start of the intervention between May and October 2016, and then again 2 years later between May and August 2018. Full details on the development of the survey instrument can be found in Veríssimo et al. (2020). To reduce researcher bias, the questionnaires were conducted by the Santomean NGO Monte Pico, an institution with no association to Programa Tatô or sea turtle conservation. We attempted to visit each village household once and female or male respondents were identified pseudo-randomly, depending on whether the house number was even or odd. The number of households with an eligible respondent at home and willing to participate ranged from 73 to 490 per community, or ~69% of the total population (further details in Supporting Information S2). The questionnaire covered socio-demographic characteristics, consumption of sea turtle eggs and meat and attitudes and social norms around sea turtle conservation and trade (Supporting Information S3). Informed consent was given verbally, and respondents were able to withdraw at any point. The anonymity of participants was fully protected, and the questionnaire was approved by the College of Life and Environmental Sciences (Penryn) Ethics Committee at the University of Exeter (reference 2017/1755), United Kingdom.

Consumption of sea turtle eggs and meat was measured using the unmatched count technique (UCT). This has been widely used in social science to uncover the prevalence of a diverse range of sensitive and/or illegal behaviours, including logging (Wilfred et al., 2019), bushmeat poaching (Nuno et al., 2013) and dangerous driving (Sheppard & Earleywine, 2013). Survey respondents are randomly allocated into two groups: control and treatment (Hinsley et al., 2019). They are both shown the same list of non-sensitive statements, with the addition of the sensitive behaviour of interest for the treatment group list only. Each individual is asked how many, but not which, items apply to them. Prevalence of the sensitive behaviour can then be estimated by calculating the difference in means between the two groups. Treatment status for the UCT was assigned pseudo-randomly based on the time of the day. If the watch of the surveyor marked an even number of minutes, the respondent was assigned to the treatment, if it was odd the respondent was assigned to the control. UCT questions were preceded by a non-sensitive training question about household occupations.

Due to the difficulties in accurately assessing covert illegal behaviours we also tracked biological monitoring data over the same time period. Members of Programa Tatô regularly patrolled the beaches on São Tomé during the sea turtle nesting seasons. From 2012 to 2015 beaches in four of the eight study communities were monitored daily. These four communities were all areas where Programa Tatô worked, and received the full intervention. From 2015 onwards, every single beach in all of the study communities was monitored daily. Patrollers counted the number of sea turtle nests and estimated the number of harvested sea turtles based on drag marks through the sand. We conducted repeated measure ANOVAs in R v4.0 to measure any changes.

2.3 | Advantages of study context

There were several factors present in this study setting that could facilitate causal inference. São Tomé is both geographically small and isolated, allowing us to track spatial spillovers from ecological processes and behavioural responses (Baylis et al., 2016). For example, Programa Tatô is the only NGO focussed on sea turtle conservation in São Tomé, so we were confident that no external conservation campaigns were conducted concurrently. We also knew that the NGO had established strong links with key stakeholders, working since 2003 with national and local enforcement agencies as well as the Ministry of Environment. They have employed a number of residents from the local communities, whom they have been working with for many years, and have developed alternative livelihoods for former sea turtle traders.

The purpose of demand reduction campaigns is to relieve poaching pressure on threatened species by changing consumer behaviour, but few evaluations actually measure conservation impacts. In part, this is because the locations of consumer bases and harvesting sites are often geographically separated. Due again to the small size of our study location, we were able to track both consumer behaviour and conservation impacts. In addition, we benefited from the fact that during the nesting season female sea turtles are more easily accessible, whereas monitoring most animal populations is difficult and resource-intensive (Gardner et al., 2008; Sims et al., 2008). Nest counts on beaches are relatively inexpensive, do not require much expertise and have a low sampling error (Warden et al., 2017; Whiting et al., 2014). The detectability of nests is high, and it is possible to conduct a comprehensive survey of the annual nesting population. Although this does not include all demographic levels (e.g. juveniles or males), it is a useful indicator of population abundance (Whiting et al., 2013). It is also comparatively easy to detect poaching activity for adult sea turtles. Their large size means they need to be dragged away rather than carried, leaving evident marks in the sand that can be spotted during daily beach patrols, regardless of weather conditions (Lino et al., 2010; Summers et al., 2018; Troëng & González, 1998).

2.4 | Challenges in implementation and analysis

When designing project evaluations, random allocation of subjects is a simple yet effective tool to minimise bias. It theoretically establishes a robust counterfactual by controlling for both confounding factors that affect how subjects respond to the intervention, and unequal probabilities in the likelihood of being treated (Gertler et al., 2016; Pynegar et al., 2019). However, partnering with practitioners often complicates subject allocation. Prioritising the likelihood of project success over robust evaluation designs may involve building on existing relationships. This was the case with our project, where the NGO required that we assigned villages nearest to the largest nesting beaches to the intervention, as these were the ones where they had a history of working and had forged social connections. This means that the two groups of subjects, those in the target villages versus the comparison villages, were not necessarily comparable on key attributes. Initial questionnaire data showed differing levels of consumption of sea turtle products between the intervention and comparison groups (Supporting Information S2; Veríssimo et al., 2020). In particular, the intervention communities reported less sea turtle meat consumption but more sea turtle egg consumption, possibly as a function of their proximity to the largest nesting beaches and major egg sources. Additionally, those in the full intervention villages had already been exposed to environmental education messages about sea turtles in the past. This may have made the villagers potentially more resistant to change, or alternatively primed them and therefore increased their openness to change.

While the restricted size of the island meant we could track spillover effects, small geographies also make it difficult to contain interventions. The mass media messages were impossible to implement in a geographically targeted way and so went island-wide. Those in the target communities received the full intervention as planned, while contamination of comparison communities meant they essentially became a mass media treatment condition (Miteva et al., 2012). Imperfect compliance in impact evaluations, a discrepancy between

assigned and actual treatment status, is always a danger for researchers (Gertler et al., 2016). Mass media is often used in conservation campaigns as it is a cheap way of reaching a wide audience, but it is difficult to control the audience composition (Veríssimo et al., 2018). To adapt, instead of using our originally intended BACI design, we developed a before–after design with two treatment conditions; the full intervention including community events, and a mass media only treatment (formerly the comparison communities). In this situation, the unintended exposure of all study participants to the mass media portions of the campaign limits what we can say about the effectiveness of the campaign. Without an appropriate control it is much harder to determine whether changes are due solely to the full intervention and/or mass media messaging, or due to changes in other external factors during the study period.

During the analysis of the UCT responses, we used generalised linear models fitted with card type (sensitive item present or not) to estimate consumption prevalence over the previous 12 months pre- and post-intervention. The coefficient estimate for the interaction between survey date and card type was used to calculate the effect size for any changes in behaviour in the two treatment conditions (Holbrook & Krosnick, 2010). In the full intervention communities we found no significant difference between the pre- and post-survey estimated for the prevalence of sea turtle meat consumption (Figure 2). However, the consumption of sea turtle eggs significantly decreased from 40% to 11% (−0.3 ± 0.14). The longitudinal nature of the study in combination with the sensitive questioning technique limited the independent variables we could fit to the behavioural model, preventing us from exploring heterogeneous outcomes between subpopulations.

The results were not so simple in the mass media only communities. We examined whether the UCT resulted in design effects that could reduce the validity of the sensitive behaviour estimates. As described by Blair and Imai (2012), design effects occur when the response behaviour to the control items is affected by introducing the sensitive item. We used a statistical test for these effects with the LIST package of the software R (Blair et al., 2018). Unfortunately, statistical analysis of the UCT results revealed design effects in the post-intervention mass media only communities, affecting the validity of the behavioural estimates (Blair & Imai, 2012). Essentially the presence of the sensitive item (e.g. 'sea turtle meat') in the list of control items biased overall responses, with participants underreporting the numbers of species they consumed. This led to paradoxical negative percentages of behavioural prevalence, an issue that can also occur with other specialised questioning techniques (Fairbrass et al., 2016). Due to these design effects we are unable to determine whether consumption behaviour in the mass media only communities changed. One possible explanation for the over-sensitivity in responses to questions regarding sea turtle consumption could be the campaign activities altering social acceptability of consumption post-intervention. However, this also raises questions about the reliability of responses to the attitude and social norm sections where direct questioning was used (Supporting Information S4).

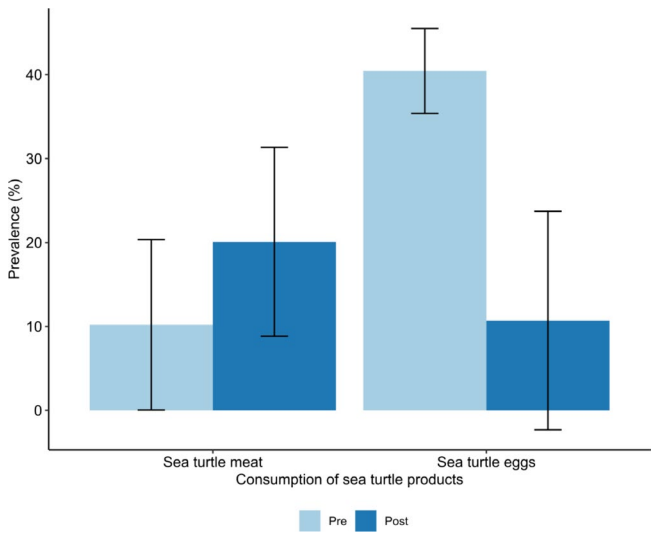

FIGURE 2 Changes in estimated prevalence (*SE*) of sea turtle meat and egg consumption amongst citizens exposed to the full social marketing intervention before and after the study

2.5 | Moving from behavioural outcomes to conservation impacts

At first glance, results from the biological monitoring data appeared positive. The nesting season on São Tomé is September to April with a peak between November and February (Vieira et al., 2016). As beaches in the mass media only communities were not patrolled regularly prior to the 2015, we focussed on just the four full intervention communities for which we had daily monitoring data. The number of sea turtles harvested decreased over the study period ($F(1) = 23.59$, $p > 0.001$), while the number of sea turtle nests recorded on nesting beaches increased ($F(1) = 13.57$, $p = 0.001$; Figure 3). Closer inspection however revealed that the decrease in poaching began before the start of the campaign. We know that there were no other concurrent demand reduction campaigns or sea turtle conservation work that could have contributed to this. There were also no macroeconomic shocks that could explain the change at the national level, as indicated by GDP varying little during the period of the intervention (The World Bank, 2018). We gathered additional data to search for further possible confounders.

Although the law criminalising the trade of sea turtles was enacted in 2014, initially it went unenforced and many citizens were not even aware of its existence (Vieira et al., 2016). To address this, Programa Tatô organised a National Workshop on Sea Turtle Law Enforcement Strategies with key stakeholders in April 2015. The first confiscation of traded sea turtles followed the next year, and systematic enforcement began with a mass seizure of sea turtle meat at the capital's municipal market in November 2017 (Supporting Information S5). The lagging enforcement by law officials and presumably growing awareness of the illegality is a likely confounder of the poaching data. By raising the salience of illegal sea turtle trade

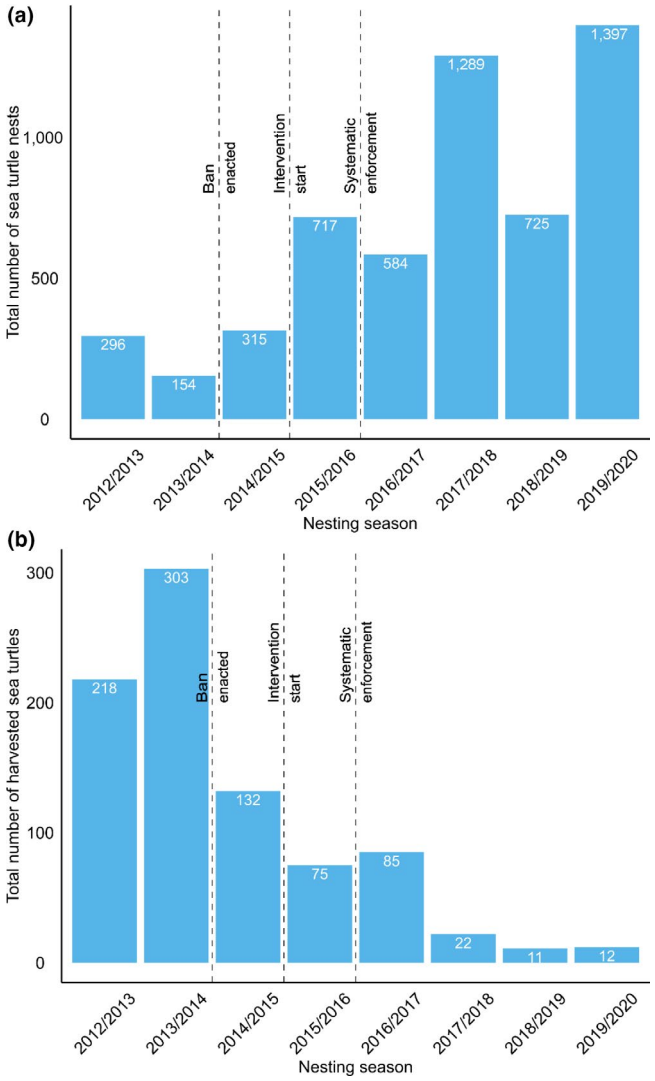

FIGURE 3 Number of sea turtle nests (a) and harvested sea turtles (b) on beaches across São Tomé, by nesting season

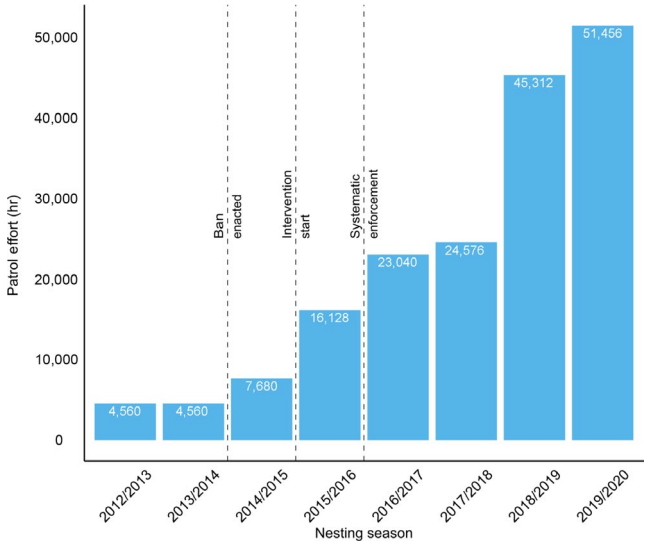

FIGURE 4 Number of hours spent patrolling beaches in the full intervention communities across São Tomé, by nesting season

amongst communities, the campaign may have prompted officials to invest more resources into preventing poaching. Indeed, after the intervention respondents in all communities were more likely to say that participants in the trade should face punitive measures (Supporting Information S4). While this would be an indirect success for the campaign, it complicates our understanding of the mechanisms of any conservation benefits. Thus, we are unable to disentangle the impacts of consumer behaviour change from changing law enforcement.

In April 2017, Programa Tatô supported the establishment of an organisation called 'Queremos ter um futuro com destino' ('We want a future with destiny'). Its aim was to help the main sea turtle traders in the capital's municipal market start an alternative, legal business, likely breaking one of the links in the trade chain. Further, the very act of Programa Tatô employees walking the beach to monitor sea turtle nests may have had a deterrent effect on any would-be poacher. If the effort put into monitoring had remained the same through the study period this could be ruled out as a potential confounder. However, records kept by monitors show a steady rise in the number of hours spent patrolling beaches (Figure 4). This increase was initiated by rangers in response to rising mortality at the nesting beaches (Vieira et al., 2016). As well as acting as a confounder, it is possible that the increase in time spent patrolling could be associated with a higher detection probability for nests. This increase, as well as potential interannual variability in nesting numbers, undermines our confidence in the meaningfulness of the nesting data (Sims et al., 2008). This is not a concern for the poaching data, which decreased in spite of the additional time spent searching.

3 | LESSONS LEARNT

Despite the advantageous context, our ability to infer causality was impacted by serious difficulties. We were hampered by spatial spillovers, design effects from the sensitive questioning technique, concurring law enforcement and changing monitoring protocols. Although we can reliably report on changes in outcomes in communities exposed to the full intervention, namely a decrease in self-reported sea turtle egg consumption and in sea turtle poaching, we were unable to confidently attribute these changes to any one factor. This is concerning, and encapsulates some of the difficulties faced by practitioners attempting to apply impact evaluation in the field.

One clear lesson that emerges is the value of triangulation between independent data sources. Impact evaluations should be approached with the mindset that you are seeking to disprove your hypothesis, rather than confirm it. At every stage of the evaluation it is important to consider rival explanations for your findings (Ferraro, 2009). This may require gathering more data than would be needed to make a convincing case for the success of an intervention. If we had compared poaching rates against only the first year preceding the start of the demand reduction campaign, then it would

have been easy to attribute the decline to our work. By including data from the widest time period possible we were able to spot a pre-existing trend, and investigate potential confounders. This led us to acquire seizure reports and beach monitoring protocols, and develop a more nuanced theory of change which included these additional inputs (Figure 5).

Theories of change can help create a broader understanding of the intervention context and causal pathways by requiring researchers to explicitly represent changes in inputs, outcomes and impacts (Adams et al., 2019; Baylis et al., 2016). They have been used successfully to understand the effects of conservation campaigns (Salazar et al., 2019). In this theory of change we hypothesise mechanistic relationships that are conceptually likely, but for which we either did not test or had unreliable data (Figure 5). This includes the potential deterrent effect of unforeseen increases in law enforcement and beach patrols on poaching rates. Many of the norms and attitudes the intervention targeted did not change, perhaps because they were already favourable to sea turtles, but respondents did develop negative attitudes towards the consumption of sea turtle products (Supporting Information S4). However, we cannot confidently attribute this to the campaign, as media coverage of sea turtle seizures could also be responsible for changing attitudes. The cyclical feedback loops within the system are notable. For instance, lowered poaching rates could cause lower availability of sea turtle products, which may lead to decreased consumption. A reduction in demand for sea turtles could again negatively impact poaching.

We used qualitative interviews to help develop the survey instrument, but they could also have provided valuable insights throughout the evaluation (Abu-Taleb & Murad, 1999; Drury et al., 2011). For example, regular semi-structured interviews or focus groups may have highlighted attitudes towards increasing law enforcement, refining our theory of change (Audrey et al., 2006;

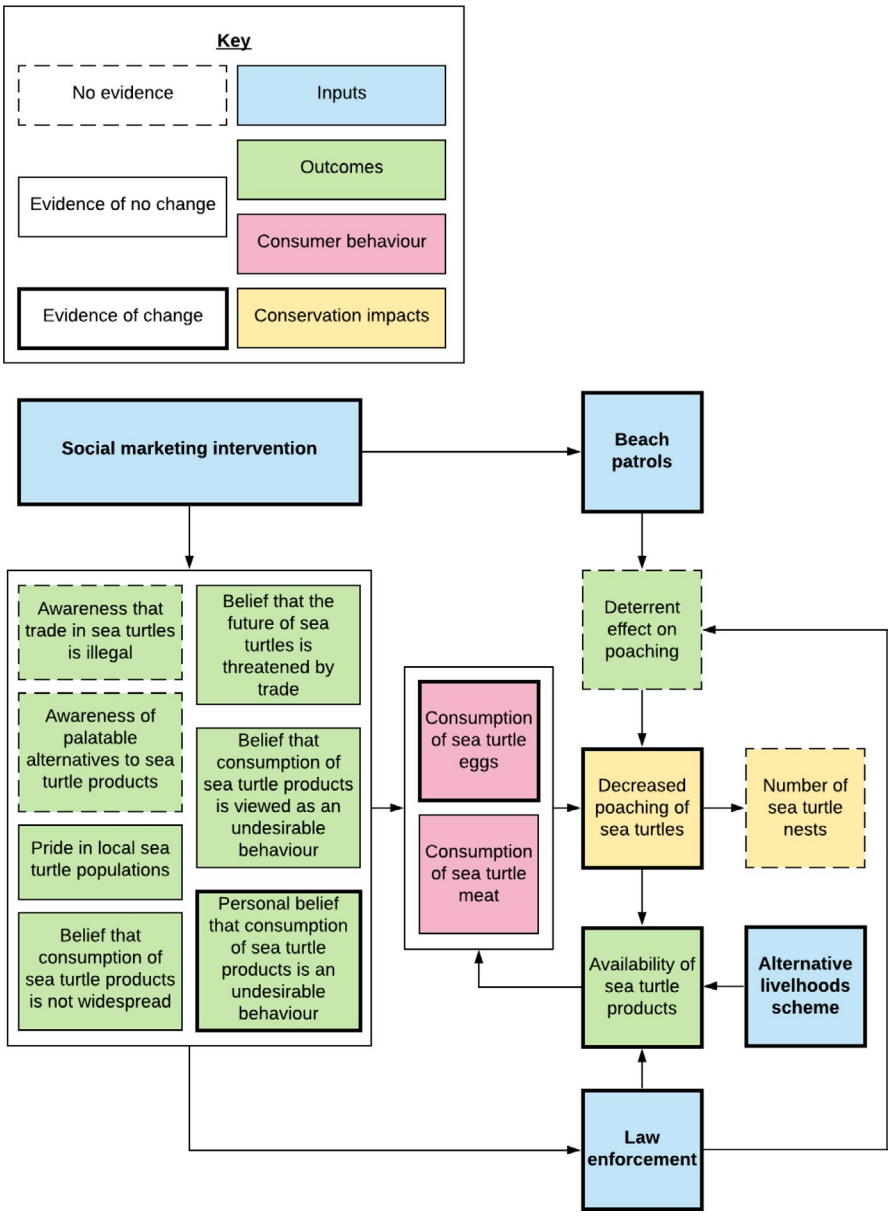

FIGURE 5 Overall theory of change for the change in poaching levels between 2012 and 2020

Reynolds et al., 2014). Continued monitoring in this form can feed into adaptive management of a campaign, enabling the refinement of messages and approaches to increase impact (Audrey et al., 2006; Murtagh et al., 2007). Theory-based evaluations which incorporate qualitative data gain greater ability to examine the context in which any changes take place (Stem et al., 2005).

The UCT has not yet been used in a longitudinal study (Hinsley et al., 2019). We had to be explicit about the theoretical assumptions it relies upon to ensure the method of statistical analysis was valid for our purposes (Blair & Imai, 2012; Nuno & St. John, 2014). This is how we identified the design effects biasing the sea turtle consumption estimates in the mass media only communities (Blair & Imai, 2012). Unfortunately, many questioning techniques can fail in differing ways (Moshagen et al., 2014). Collecting additional data does have a cost (e.g. time, money, the goodwill of participants), but combining more than one questioning technique within a survey could compensate for individual limitations. There may be large standard errors and potential design effects with the UCT or the risk of bias with direct questioning (Nuno & St. John, 2014). Researchers and practitioners working on sensitive behaviours would benefit from investing more research into alternative sensitive questioning techniques such as the ballot box method, which do not require significantly larger sample sizes in order to avoid large standard errors (Bova et al., 2018; St. John et al., 2010). Further, multiple indicators like self-reports and biological monitoring data are useful to help us triangulate changes in hard-to-measure variables like illegal behaviours (Veríssimo et al., 2017).

The difficulty in moving from behavioural outcomes to conservation impacts should not discourage future researchers. Instead it should catalyse greater effort into developing approaches to bridge that divide. For example, using structural equation models to explore pathways between different factors in a theory of change can help better understand the linkages between different indicator types (McDonald et al., 2020). We must remember that conservation is not just about documenting biodiversity declines; it ought to help safeguard it also (Williams et al., 2020). While human behaviour is at the centre of all key threats to biodiversity, addressing those threats should not stop at measuring behavioural trends but should look into understanding if and how specific interventions benefit the species we are trying to protect (Veríssimo et al., 2017).

ACKNOWLEDGEMENTS

The authors would like to thank Aidan Keane and Ana Nuno for their advice on the statistical analysis. This research was supported by a Rufford Small Grant (18821-1), Lisbon Oceanarium, and U.S. Fish and Wildlife Service. L.T.-W. acknowledges the support of the University of Kent Research Alumni Scholarship. D.V. acknowledges the support of the Oxford Martin Programme on the Illegal Wildlife Trade. We would like to thank Hipólito Lima, Gabriel Oquiongo, Aristides Monteiro, Antunes Pina, Desidério Paquete, Litoney Oliveira, Maria Branco, Gabriela Fernandes and Yedda Oliveira for the support during field work. We would also like to thank the Programa Tatô team responsible for monitoring the beaches.

CONFLICTS OF INTEREST

S.V., V.J., D.M. and B.F. are employees of Programa Tatô. D.V. is a Guest Editor for the 'Consuming wildlife – managing demand for products in the wildlife trade' Special Feature, which this article is part of, but he was not involved in the peer review and decision making process for this manuscript. There are no other conflicts of interest to declare.

AUTHORS' CONTRIBUTIONS

D.V., S.V. and V.J. conceived the intervention and designed methodology; S.V., D.M. and V.J. collected the data; L.T.-W. analysed the data; L.T.-W. conceptualised the manuscript framing; L.T.-W., R.J.S. and D.V. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

The survey data underlying this research can be found in anonymised form at https://doi.org/10.6084/m9.figshare.12559466 (Thomas-Walters, Vieira, et al., 2020).

ORCID

Laura Thomas-Walters https://orcid.org/0000-0002-3250-2799
Diogo Veríssimo https://orcid.org/0000-0002-3519-6782

REFERENCES

Abu-Taleb, M. F., & Murad, M. M. (1999). Use of focus groups and surveys to evaluate water conservation campaign. *Journal of Water Resources Planning and Management*, *125*, 94–99. https://doi.org/10.1061/(ASCE)0733-9496(1999)125:2(94)

Adams, V. M., Barnes, M., & Pressey, R. L. (2019). Shortfalls in conservation evidence: Moving from ecological effects of interventions to policy evaluation. *One Earth*, *1*(1), 62–75. https://doi.org/10.1016/j.oneear.2019.08.017

Amel, E., Manning, C., Scott, B., & Koger, S. (2017). Beyond the roots of human interaction: Fostering collective effort toward ecosystem conservation. *Science*, *356*, 275–279.

Audrey, S., Holliday, J., Parry-Langdon, N., & Campbell, R. (2006). Meeting the challenges of implementing process evaluation within randomized controlled trials: The example of ASSIST (A Stop Smoking in Schools Trial). *Health Education Research*, *21*(3), 366–377. https://doi.org/10.1093/her/cyl029

Backmann, M. (2017). What's in a gold standard? In defence of randomised controlled trials. *Medicine, Health Care and Philosophy*, *20*(4), 513–523. https://doi.org/10.1007/s11019-017-9773-2

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A., & Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, *9*(1), 58–64. https://doi.org/10.1111/conl.12180

Blair, G., Chou, W., Imai, K., Park, B., & Coppock, A. (2018). *list: Statistical methods for the item count technique and list experiment* (R package version 9.0). Retrieved from https://cran.r-project.org/web/packages/list/list.pdf

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, *20*(1), 47–77. https://doi.org/10.1093/pan/mpr048

Bova, C. S., Aswani, S., Farthing, M. W., & Potts, W. M. (2018). Limitations of the random response technique and a call to implement the ballot box method for estimating recreational angler compliance using surveys. *Fisheries Research*, *208*, 34–41. https://doi.org/10.1016/j.fishres.2018.06.017

Campbell, L. (2003). Contemporary culture, use, and conservation of sea turtles. In P. L. Lutz, J. A. Musick, & J. Wyneken (Eds.), *Biology of sea turtles, Volume II* (pp. 307–338). CRC. https://doi.org/10.1201/9781420040807.ch12

Castroviejo, J., Juste B, J., Del Val Pérez, J., Castelo, R., & Gil, R. (1994). Diversity and status of sea turtle species in the Gulf of Guinea islands. *Biodiversity and Conservation*, *3*(9), 828–836. https://doi.org/10.1007/BF00129661

Curzon, H. F., & Kontoleon, A. (2016). From ignorance to evidence? The use of programme evaluation in conservation: Evidence from a Delphi survey of conservation experts. *Journal of Environmental Management*, *180*, 466–475. https://doi.org/10.1016/j.jenvman.2016.05.062

Dicks, L. V., Walsh, J. C., & Sutherland, W. J. (2014). Organising evidence for environmental management decisions: A '4S' hierarchy. *Trends in Ecology & Evolution*, *29*(11), 607–613. https://doi.org/10.1016/j.tree.2014.09.004

Donlan, C. J., Wingfield, D. K., Crowder, L. B., & Wilcox, C. (2010). Using expert opinion surveys to rank threats to endangered species: A case study with sea turtles. *Conservation Biology*, *24*(6), 1586–1595. https://doi.org/10.1111/j.1523-1739.2010.01541.x

Drury, R., Homewood, K., & Randall, S. (2011). Less is more: The potential of qualitative approaches in conservation research. *Animal Conservation*, *14*, 18–24. https://doi.org/10.1111/j.1469-1795.2010.00375.x

Fairbrass, A., Nuno, A., Bunnefeld, N., & Milner-Gulland, E. J. (2016). Investigating determinants of compliance with wildlife protection laws: Bird persecution in Portugal. *European Journal of Wildlife Research*, *62*(1), 93–101. https://doi.org/10.1007/s10344-015-0977-6

Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, *122*, 75–84. https://doi.org/10.1002/ev.297

Ferraro, P. J., Sanchirico, J. N., & Smith, M. D. (2019). Causal inference in coupled human and natural systems. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5311–5318. https://doi.org/10.1073/pnas.1805563115

Frazier, J. (2003). Prehistoric and ancient historic interactions between humans and marine turtles. In P. L. Lutz, J. A. Musick, & J. Wyneken (Eds.), *Biology of sea turtles, Volume II* (pp. 1–38). CRC. https://doi.org/10.1201/9781420040807.ch1

Gardner, T. A., Barlow, J., Araujo, I. S., Ávila-Pires, T. C., Bonaldo, A. B., Costa, J. E., Esposito, M. C., Ferreira, L. V., Hawes, J., Hernandez, M. I. M., Hoogmoed, M. S., Leite, R. N., Lo-Man-Hung, N. F., Malcolm, J. R., Martins, M. B., Mestre, L. A. M., Miranda-Santos, R., Overal, W. L., Parry, L., … Peres, C. A. (2008). The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, *11*(2), 139–150. https://doi.org/10.1111/j.1461-0248.2007.01133.x

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact evaluation in practice* (2nd ed.). Inter-American Development Bank and World Bank. https://doi.org/10.1596/978-1-4648-0779-4

Graff, D. (1996). Sea turtle nesting and utilization survey in São Tomé. *Marine Turtle Newsletter*, *75*, 8–12.

Hildén, M. (2009). Time horizons in evaluating environmental policies. *New Directions for Evaluation*, *122*, 9–18. https://doi.org/10.1002/ev.291

Hinsley, A., Keane, A., St. John, F. A. V., Ibbett, H., & Nuno, A. (2019). Asking sensitive questions using the unmatched count technique: Applications and guidelines for conservation. *Methods in Ecology and Evolution*, *10*(3), 308–319. https://doi.org/10.1111/2041-210X.13137

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, *74*(1), 37–67. https://doi.org/10.1093/poq/nfp065

Josefsson, J., Hiron, M., Arlt, D., Auffret, A. G., Berg, Å., Chevalier, M., Glimskär, A., Hartman, G., Kačergytė, I., Klein, J., Knape, J., Laugen, A. T., Low, M., Paquet, M., Pasanen-Mortensen, M., Rosin, Z. M., Rubene, D., Żmihorski, M., & Pärt, T. (2020). Improving scientific rigour in conservation evaluations and a plea deal for transparency on potential biases. *Conservation Letters*, e12726. https://doi.org/10.1111/conl.12726

Knight, A. T., Cook, C. N., Redford, K. H., Biggs, D., Romero, C., Ortega-Argueta, A., Norman, C. D., Parsons, B., Reynolds, M., Eoyang, G., & Keene, M. (2019). Improving conservation practice with principles and tools from systems thinking and evaluation. *Sustainability Science*, *14*(6), 1531–1548. https://doi.org/10.1007/s11625-019-00676-x

Kvangraven, I. H. (2020). Impoverished economics? A critical assessment of the new gold standard. In *World Development* (Vol. 127). Elsevier Ltd. https://doi.org/10.1016/j.worlddev.2019.104813

Lino, S., Gonçalves, E., & Cozens, J. (2010). The loggerhead sea turtle (*Caretta caretta*) on Sal Island, Cape Verde: Nesting activity and beach surveillance in 2009. *Arquipelago*, *27*, 59–63.

Mancini, A., & Koch, V. (2009). Sea turtle consumption and black market trade in Baja California Sur, Mexico. *Endangered Species Research*, *7*(1), 1–10. https://doi.org/10.3354/esr00165

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. https://doi.org/10.1002/ev.298

McDonald, G., Wilson, M., Veríssimo, D., Twohey, R., Clemence, M., Apistar, D., Box, S., Butler, P., Cadiz, F. C., Campbell, S. J., Cox, C., Effron, M., Gaines, S., Jakub, R., Mancao, R. H., Rojas, P. T., Tirona, R. S., & Vianna, G. (2020). Catalyzing sustainable fisheries management through behavior change interventions. *Conservation Biology*. https://doi.org/10.1111/cobi.13475.This

McIntosh, E. J. (2019). Barriers to the evaluation of systematic conservation plans: Insights from landmark Australian plans. *Biological Conservation*, *237*, 70–80. https://doi.org/10.1016/j.biocon.2019.06.029

Miteva, D. A., Pattanayak, S. K., & Ferraro, P. J. (2012). Evaluation of biodiversity policy instruments: What works and what doesn't? *Oxford Review of Economic Policy*, *28*(1), 69–92. https://doi.org/10.1093/oxrep/grs009

Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2014). An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology*, *61*(1), 48–54. https://doi.org/10.1027/1618-3169/a000226

Murtagh, M. J., Thomson, R. G., May, C. R., Rapley, T., Heaven, B. R., Graham, R. H., Kaner, E. F., Stobbart, L., & Eccles, M. P. (2007). Qualitative methods in a randomised controlled trial: The role of an integrated qualitative process evaluation in providing evidence to discontinue the intervention in one arm of a trial of a decision support tool. *Quality and Safety in Health Care*, *16*(3), 224–229. https://doi.org/10.1136/qshc.2006.018499

Nada, M., & Casale, P. (2011). Sea turtle bycatch and consumption in Egypt threatens Mediterranean turtle populations. *Oryx*, *45*(1), 143–149. https://doi.org/10.1017/S0030605310001286

Nuno, A., Bunnefeld, N., Naiman, L. C., & Milner-Gulland, E. J. (2013). A novel approach to assessing the prevalence and drivers of illegal bushmeat hunting in the Serengeti. *Conservation Biology*, *27*(6), 1355–1365. https://doi.org/10.1111/cobi.12124

Nuno, A., & St. John, F. A. V. (2014). How to ask sensitive questions in conservation: A review of specialized questioning techniques. *Biological Conservation*, *189*, 5–15. https://doi.org/10.1016/j.biocon.2014.09.047

Phelps, J., Carrasco, L. R., & Webb, E. L. (2014). A framework for assessing supply-side wildlife conservation. *Conservation Biology*, *28*(1), 244–257. https://doi.org/10.1111/cobi.12160

Preskill, H. (2009). Reflections on the dilemmas of conducting environmental evaluations. *New Directions for Evaluation*, *122*, 97–103. https://doi.org/10.1002/ev.299

Pynegar, E. L., Gibbons, J. M., Asquith, N. M., & Jones, J. P. G. (2019). What role should randomized control trials play in providing the evidence

base for conservation? *Oryx*, 1–10. https://doi.org/10.1017/S0030605319000188

Reynolds, J., Diliberto, D., Mangham-Jefferies, L., Ansah, E. K., Lal, S., Mbakilwa, H., Bruxvoort, K., Webster, J., Vestergaard, L. S., Yeung, S., Leslie, T., Hutchinson, E., Reyburn, H., Lalloo, D. G., Schellenberg, D., Cundill, B., Staedke, S. G., Wiseman, V., Goodman, C., & Chandler, C. I. (2014). The practice of 'doing' evaluation: Lessons learned from nine complex intervention trials in action. *Implementation Science*, *9*(75). https://doi.org/10.1186/1748-5908-9-75

Rosen, G. E., & Smith, K. F. (2010). Summarizing the evidence on the international trade in illegal wildlife. *EcoHealth*, *7*(1), 24–32. https://doi.org/10.1007/s10393-010-0317-y

Salazar, G., Mills, M., & Veríssimo, D. (2019). Qualitative impact evaluation of a social marketing campaign for conservation. *Conservation Biology*, *33*(3), 634–644. https://doi.org/10.1111/cobi.13218

Schleicher, J., Eklund, J., D. Barnes, M., Geldmann, J., Oldekop, J. A., & Jones, J. P. G. (2019). Statistical matching for conservation science. *Conservation Biology*, *34*(3), 1–12. https://doi.org/10.1111/cobi.13448

Sheppard, S. C., & Earleywine, M. (2013). Using the unmatched count technique to improve base rate estimates of risky driving behaviours among veterans of the wars in Iraq and Afghanistan. *Injury Prevention*, *19*(6), 382–386. https://doi.org/10.1136/injuryprev-2012-040639

Sims, M., Bjorkland, R., Mason, P., & Crowder, L. B. (2008). Statistical power and sea turtle nesting beach surveys: How long and when? *Biological Conservation*, *141*(12), 2921–2931. https://doi.org/10.1016/j.biocon.2008.07.021

Stem, C., Margoluis, R., Salafsky, N., & Brown, M. (2005). Monitoring and evaluation in conservation: A review of trends and approaches. *Conservation Biology*, *19*(2), 295–309. https://doi.org/10.1111/j.1523-1739.2005.00594.x

St. John, F. A. V., Edwards-Jones, G., Gibbons, J. M., & Jones, J. P. G. (2010). Testing novel methods for assessing rule breaking in conservation. *Biological Conservation*, *143*(4), 1025–1030. https://doi.org/10.1016/j.biocon.2010.01.018

Summers, T. M., Martin, S. L., Hapdei, J. R., Ruak, J. K., & Jones, T. T. (2018). Endangered green turtles (*Chelonia mydas*) of the Northern Mariana Islands: Nesting ecology, poaching, and climate concerns. *Frontiers in Marine Science*, *4*(Jan). https://doi.org/10.3389/fmars.2017.00428

Sutherland, W. J., & Wordley, C. F. R. (2017). Evidence complacency hampers conservation. *Nature Ecology and Evolution*, *1*(9), 1215–1216. https://doi.org/10.1038/s41559-017-0244-1

The World Bank. (2018). *GDP per capita (current US$)—Sao Tome and Principe*. Retrieved from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=ST

Thomas-Walters, L., Vieira, S., Jiménez, V., Monteiro, D., Ferreira, B., Smith, R. J., & Veríssimo, D. (2020). Data from: Challenges in the impact evaluation of behaviour change interventions: The case of sea turtle meat and eggs in São Tomé. *Figshare*, https://doi.org/10.6084/m9.figshare.12559466

Thomas-Walters, L., McNulty, C., & Verissimo, D. (2019). A scoping review into the impact of animal imagery on pro-environmental outcomes. *Ambio*, *49*(6), 1135–1145. https://doi.org/10.1007/s13280-019-01271-1

Thomas-Walters, L., Veríssimo, D., Gadsby, E., Roberts, D., & Smith, R. J. (2020). Taking a more nuanced look at behavior change for demand reduction in the illegal wildlife trade. *Conservation Science and Practice*, *2*(9). https://doi.org/10.1111/csp2.248

Troëng, S., & González, T. A. R. (1998). Illegal harvest of nesting green turtles Chelonia Mydas in Tortuguero National Park, Costa Rica. *Proceedings of the 18th International Sea Turtle Symposium, June* (p. 312). https://conserveturtles.org/illegal-harvest-nesting-green-turtles-tortuguero-national-park-costa-rica/

Veríssimo, D., Bianchessi, A., Arrivillaga, A., Cadiz, F. C., Mancao, R., & Green, K. (2017). Does it work for biodiversity? Experiences and challenges in the evaluation of social marketing campaigns. *Social Marketing Quarterly*, *24*(1), 18–34. https://doi.org/10.1177/1524500417734806

Veríssimo, D., Schmid, C., Kimario, F. F., & Eves, H. E. (2018). Measuring the impact of an entertainment-education intervention to reduce demand for bushmeat. *Animal Conservation*, *21*(4), 324–331. https://doi.org/10.1111/acv.12396

Veríssimo, D., Vieira, S., Monteiro, D., Hancock, J., & Nuno, A. (2020). Audience research as a cornerstone of demand management interventions for illegal wildlife products: Demarketing sea turtle meat and eggs. *Conservation Science and Practice*, *2*(3), 1–14. https://doi.org/10.1111/csp2.164

Veríssimo, D., & Wan, A. K. Y. (2018). Characterizing efforts to reduce consumer demand for wildlife products. *Conservation Biology*, *33*(3), 623–633. https://doi.org/10.1111/cobi.13227.This

Vieira, S., Jiménez, V., Besugo, A., Costa, S., Miranda, F., Hancock, J., Loloum, B., & Oliveira, L. (2016). Participative approach to discuss novel law implementation strategies in São Tomé and Príncipe. *African Sea Turtle Newsletter*, *5*, 15–21.

Warden, M. L., Haas, H. L., Richards, P. M., Rose, K. A., & Hatch, J. M. (2017). Monitoring trends in sea turtle populations: Walk or fly? *Endangered Species Research*, *34*, 323–337. https://doi.org/10.3354/esr00855

Whiting, A. U., Chaloupka, M., & Limpus, C. J. (2013). Comparing sampling effort and errors in abundance estimates between short and protracted nesting seasons for sea turtles. *Journal of Experimental Marine Biology and Ecology*, *449*, 165–170. https://doi.org/10.1016/j.jembe.2013.09.016

Whiting, A. U., Chaloupka, M., Pilcher, N., Basintal, P., & Limpus, C. J. (2014). Comparison and review of models describing sea turtle nesting abundance. *Marine Ecology Progress Series*, *508*, 233–246. https://doi.org/10.3354/meps10832

Wilfred, P., Milner-Gulland, E. J., & Travers, H. (2019). Attitudes to illegal behaviour and conservation in western Tanzania. *Oryx*, *53*(3), 513–522. https://doi.org/10.1017/S0030605317000862

Williams, D. R., Balmford, A., & Wilcove, D. S. (2020). The past and future role of conservation science in saving biodiversity. *Conservation Letters*, *13*(4), 1–7. https://doi.org/10.1111/conl.12720

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Thomas-Walters L, Vieira S, Jiménez V, et al. Challenges in the impact evaluation of behaviour change interventions: The case of sea turtle meat and eggs in São Tomé. *People Nat*. 2020;2:913–922. https://doi.org/10.1002/pan3.10162