On the evaluation, monitoring and management of law enforcement patrols in protected areas

Anthony Dancer

Centre for Biodiversity & Environment Research, UCL & Institute of Zoology, ZSL

A dissertation submitted for the degree of Doctor of Philosophy

March 2019

I, Anthony Dancer, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

For Ben

Once there were brook trout in the streams in the mountains. You could see them standing in the amber current where the white edges of their fins wimpled softly in the flow. They smelled of moss in your hand. Polished and muscular and torsional. On their backs were vermiculate patterns that were maps of the world in its becoming. Maps and mazes. Of a thing which could not be put back. Not be made right again. In the deep glens where they lived all things were older than man and they hummed of mystery.

Cormac McCarthy, "The Road"

His talent was as natural as the pattern that was made by the dust of a butterfly's wings.

Ernest Hemingway, "A Moveable Feast"

Abstract

Ranger-led law enforcement patrols are the primary response to illegal use of natural resources in protected areas globally. To date, however, the effectiveness of patrolling as a means to reduce illegal activity has been neglected as a subject of study. Relatedly, there has been no rigorous evaluation of tools which aim to increase patrol effectiveness through patrol monitoring and management. In this thesis, I explore the use of patrols for reducing illegal activity, and evaluate a popular tool for increasing patrol effectiveness: SMART.

SMART involves ranger-based monitoring – collection of data by rangers on patrol – of both natural resource use and patrol activity. I exploit data collected via SMART to investigate the extent of patrolling conducted in terrestrial protected areas globally. I show that patrol presence within and across sites is typically very low, is constrained by limited budgets, and frequently falls short of industry targets. I also use SMART data to explore whether and in what contexts deterrence – the primary mechanism through which patrols are assumed to reduce illegal activity – operates in practice. I focus on four protected areas with relatively high patrol presence and find that patrols may have deterred illegal activity in three sites, but the effect was weak and inconsistent.

I draw on these results and guidance from other policy arenas to evaluate SMART. I illustrate the causal pathways through which SMART aims to reduce illegal activity, using a theory of change approach. I develop evidence to verify SMART's theory of change, including whether the intervention was implemented as intended, and whether the chain of expected results occurred. I also develop a novel framework for describing heterogeneity among implemented interventions. I find that patrol presence is improving in SMART sites. Yet inconsistent implementation of management activities, and mixed evidence for deterrence, precluded a causal claim for SMART at this time.

My findings suggest that patrol activity globally is insufficient to either reduce or monitor illegal activity in protected areas. SMART may improve patrol presence, and might improve it further through more faithful implementation of management. However, inconsistent evidence of deterrence, even in sites with high patrol presence, highlights the need for fundamental research into whether and how well-managed and socially just patrolling can be effective. My findings also demonstrate that robust monitoring of threats in protected areas, independent of patrolling, is essential.

Impact statement

The natural world is increasingly threatened by human activities. Overhunting and habitat destruction are driving declines in populations of endangered species. One of society's primary responses to these threats is to designate areas where such activities are prohibited by law ('protected areas'). Laws, however, do not immediately confer protection. In many protected areas proscribed activities (e.g., poaching) continue to occur. Consequently, biodiversity continues to decline.

To reduce illegal activity protected area managers implement law enforcement – strategies for punishing people who break laws. Typically, teams of rangers, with powers to enforce laws, are employed to patrol protected sites. However, patrolling has been neglected as a subject of study. Consequently, it is unclear whether patrols can reduce illegal activity in practice.

The paucity of research into patrolling stems from a lack of information required to conduct analyses. Historically, there has been minimal monitoring in protected areas, either of illegal activity or of patrols. This lack of data has also constrained management. For example, protected area managers need information on where illegal activity is occurring to efficiently target patrols in response.

Recently, conservation organisations have developed tools for improving protected area monitoring and management, which exploit collection of data by rangers. One tool, called SMART, which aims to reduce illegal activity via patrolling, is increasingly popular, with over 600 sites globally investing in its implementation as of 2018. However, whether these investments represent a wise use of resources is unclear, because there has been no rigorous evaluation of SMART.

This thesis tackles these issues, and develops insights which can be used to generate immediate and lasting improvements to nature conservation policy and practice in protected areas globally. This thesis also advances the science underlying patrolling, to enable better research in this essential yet understudied area.

Firstly, my findings show that one of the primary mechanisms by which patrols are assumed to reduce illegal activity – deterrence – does operate in practice, suggesting patrols are an effective means of achieving conservation goals, but the effect was weak and inconsistent, and varied with context. This knowledge can be used to deploy patrols in contexts in which they are more likely to work (e.g., savannah habitats or marine environments). However, my results also demonstrate that levels of patrolling in protected areas globally may be far too low for an effect to be generated, because of major shortfalls in funding.

My findings point to an urgent need to increase financial support to protected areas, but also highlight a pressing requirement to research how patrolling can be improved and combined with alternative, non-enforcement approaches to achieve consistent and cost-effective reductions in illegal activity.

Secondly, whilst my evaluation of SMART was inconclusive, my findings show SMART was not always implemented as intended, which may have constrained its ability to improve patrolling, and provide insights for how this can be rectified. My research also suggest the tool might be further improved by integrating monitoring of illegal activity, independent of collection of data by patrols. This knowledge can be used to guide efforts to conserve biodiversity via patrolling in protected areas globally.

Acknowledgements

This thesis is dedicated to my supervisor, Ben Collen, who passed away on Saturday 19th May 2018. Ben, this thesis is very much for and because of you. I will be forever grateful for your guidance, support and friendship. I wish you could be here to see it.

Thank you also to my other supervisors, both those who were there from the start and those that joined me later: Robin Freeman, Raj Amin, Kate Jones, and Marcus Rowcliffe. You have all been incredibly generous with your time, enthusiasm and expertise, for which I am extremely grateful.

The SMART Partnership were instrumental in facilitating this research, including individuals representing the Frankfurt Zoological Society, the North Carolina Zoo, Panthera, the Wildlife Conservation Society, the World Wide Fund for Nature, and the Zoological Society of London. I am particularly indebted to Emma Stokes, who has been unfalteringly supportive of this work over the last 3.5 years. I would also like to thank the following for their assistance, time and patience throughout the process: Richard Bergl, Paul Glover-Kapfer, Abishek Harihar, Barney Long, Rohit Singh and Elsabe van der Westhuizen.

Numerous additional individuals working with or for SMART Partnership organisations helped make this research possible by providing their invaluable insights and data. I am incredibly grateful to all the technical partners and patrol managers who generously gave up their time to be interviewed or helped source monitoring data. I would also like to thank all the protected area managers, staff and rangers involved in this work, whose efforts go largely unappreciated but who often bear the greatest risks.

I have also been supported by or had the opportunity to work with several fantastic researchers throughout my PhD. My thanks, in particular, to EJ Milner-Gulland, without whose kindness over the last 9 years I would not be where I am today, and Aidan Keane, whose intellectual rigour has been inspiring. Thanks also to Andrew Dobson, Colin Beale, Harriet Ibbett and Emily Woodhouse.

Not all of the research paths taken during the course of my PhD made it into this thesis. Nevertheless, I would like to thank Piero Visconti and Neil Burgess for their ideas and assistance in the early stages. I am also enormously grateful to Alasdair Davies and all at the Zoological Society of London, both in Cameroon and the UK, who supported my acoustic research. Thanks, also, to mon fils (Tom Bruce).

My special thanks and love to all of my wonderful family and friends, who have borne with me throughout these years. I'm looking forward to seeing far more of you. To my mother and father, everything I have had the opportunity to be is entirely because of you.

Finally, to my own small family: Primrose and Gertrude, we will miss you always. Cowslip, please stop yowling. Nettle, please stop flopping on my keyboard. Alice, my wife, you have been – and always shall be – my friend.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Illegal use of natural resources is a major threat to biodiversity in protected areas globally, which is countered, primarily, by ranger-led law enforcement patrols (Harrison, 2011; Henson *et al.*, 2016). Patrolling is assumed to reduce illegal activity through detection and deterrence of crime (Nagin, 2013b): rangers travel around parks, aiming to arrest offenders, remove passive hunting devices, and discourage crime from taking place. Yet patrolling has been neglected as a subject of study. For example, the extent of patrolling in protected areas is largely unknown, despite fears that enforcement may often be inadequate (Leverington *et al.*, 2010; Plumptre *et al.*, 2014). Similarly, there is limited appreciation of whether and how mechanisms such as deterrence operate (Dobson *et al.*, 2018). The paucity of patrol research stems, in part, from a lack of monitoring data from protected areas (Bertzky and Stoll-Kleemann, 2009). Illegal activity, particularly wildlife crime, is inherently difficult to study (Gavin *et al.*, 2010), but monitoring of less elusive features, such as the activity of rangers whilst on patrol, has also been largely inadequate. Consequently, it is unclear whether patrolling, as practiced, represents an effective means to reduce illegal natural resource use.

Ranger-based monitoring – the collection of data by rangers on patrol, including observations of natural resource use, wildlife, and recording of rangers' own activities – has been developed to fill this information gap (Gray and Kalpers, 2005). Concurrently, conservation organizations have invested in technology-enabled systems which facilitate ranger-based monitoring, and which also provide platforms for using data to inform management, commonly called Law Enforcement Monitoring (LEM) tools (Stokes, 2010). One such tool, SMART (Spatial Monitoring and Reporting Tool), which aims to enable adaptive management of patrols to improve effectiveness, is increasingly popular, with deployments in >600 sites globally as of 2018 (SMART Partnership, 2018). Widespread and standardised monitoring in protected areas, via SMART, provides an unprecedented opportunity to investigate patrolling and its effects on illegal activity, but this opportunity has yet to be fully exploited.

Relatedly, there has also been an absence of rigorous evaluation of interventions aiming to improve patrol effectiveness, such as SMART. Positive outcomes have been reported in a few sites in which SMART has been implemented (Hötte *et al.*, 2016), but the extent to which the intervention caused the observed changes is unexamined. Empirical evaluation of aspects such as whether SMART contributed to reduced illegal activity, or even if it was implemented as intended, is essential but lacking. Causal attribution of the effects of conservation interventions, in general, is rare, because the complex and resource-limited contexts in which conservation operates seldom lend themselves to rigid counterfactual designs (Fisher *et al.*, 2014). For example, interventions such as SMART rarely represent homogenous treatments, but vary between participants in multiple, continuous dimensions, often precluding experimental or quasi-experimental designs (Ferraro and Pressey, 2015). However, flexible, mixed-methods approaches to causal attribution are commonly used in similarly complex policy arenas, such as public health and education (Stern *et al.*, 2012), providing an opportunity to advance evaluation in conservation.

In this thesis, I fill these gaps. I exploit recent, widespread monitoring in protected areas via SMART to improve the evidence base underpinning the use of ranger-led law enforcement patrols to reduce illegal activity. I investigate how patrolling is practiced in protected areas around the world and explore whether and in what contexts patrols deter crime. I draw on the results of these analyses and guidance and methods from other policy arenas to evaluate SMART. I illustrate, interrogate, and verify the causal pathways through which SMART aims to reduce illegal activity, using a theory-based approach for causal attribution. In the process, I develop and apply a novel framework for describing heterogeneity of conservation interventions.

1.2 Background

1.2.1 Enforcement responses to illegal activity threatening biodiversity in protected areas

Protected Areas (PAs), demarcated geographical spaces within which human activities are controlled or proscribed, are the cornerstone of modern nature conservation (Dudley, 2008). Protected areas cover 14.9% of terrestrial and inland waters outside Antarctica (UNEP-WCMC *et al.*, 2018) and attract a large share of global investment in conservation (Balmford *et al.*, 2003). These spaces continue to grow in number and extent, approaching the Convention on Biological Diversity target of 17% of land under protection by 2020 (CBD, 2010) but still short of the 50% advocated by prominent conservationists (Wilson, 2016). Despite their area and expense, protected areas are failing to stem declines in endangered species (Tittensor *et al.*, 2014). For example, African parks have failed to protect large mammals from human perturbation, resulting in an average decline of 59% in population abundance from 1970 to 2005 (Craigie *et al.*, 2010). Globally, evidence for protected areas representing an effective means for maintaining species populations is equivocal (Geldmann *et al.*, 2013).

Uncertainty over the ability of protected areas to achieve conservation outcomes stems, in part, from concerns over whether they provide sufficient biodiversity coverage (Joppa and Pfaff, 2009). However, there is also doubt over the ability of protected areas to effectively conserve biodiversity within their borders (Pringle, 2017). Protected area effectiveness is constrained by multiple factors, but an overriding concern is weak management. A global meta-analysis of 4,000 management effectiveness assessments found that over half of sites showed significant or major deficiencies in management (Leverington *et al.*, 2010). In part, these deficiencies can be explained by chronic shortfalls in political and financial support (Watson *et al.*, 2014). In short, sites often lack the resources or capacity to achieve conservation objectives. Perversely, focusing international targets on area under protection, rather than increasing support to existing sites to increase their effectiveness, has encouraged the formation of 'paper parks', which create an illusion of progress whilst biodiversity continues to decline (Barnes *et al.*, 2018). Increasing support to protected areas is essential if they are to represent an effective conservation mechanism. In tandem, the conservation community must ensure that existing protected area resources are allocated and applied efficiently and effectively to address threats to wildlife.

Multiple human activities threaten biodiversity in protected areas, ranging from recreation to livestock ranching to logging, but the most frequently reported threat is unsustainable hunting and collection of terrestrial animals (Schulze *et al.*, 2018). Illegal hunting, particularly with snares, is common throughout sites in Southeast Asia and is believed to be a major driver of wide scale defaunation in the region (Gray *et al.*, 2017). In Central Africa, illegal subsistence and commercial hunting within and without protected areas has led to drastic and rapid declines in large mammals, disrupting food webs (Abernethy *et al.*, 2013; Tranquilli *et al.*, 2014). Hunting pressure across the tropics has reduced bird and mammal abundance inside and outside parks (Benítez-López *et al.*, 2017). Consequently, "Empty forest syndrome" – the extirpation of bird and mammal species >2 kg – may be common in tropical forest reserves (Harrison, 2011). Concurrently, illegal logging, burning and encroachment are also common in protected areas throughout the global south, degrading and shrinking habitats (Laurance *et al.*, 2012).

Addressing illegal natural resource use in protected areas requires an understanding of the factors which lead individuals to break or comply with rules, and implementation of effective strategies to encourage compliance and discourage infractions (Keane *et al.*, 2008). For example, normative procedural justice theories suggest that voluntary compliance with laws is encouraged by legitimate governments and legal authorities (Moreto and Gau, 2017). Interventions which aim to incentivise compliance by increasing the benefits and decreasing the costs to local communities from supporting conservation in protected areas may also hold the potential to reduce wildlife crime (Cooney *et al.*, 2017; Wilfred *et al.*, 2017). For example, interventions which aim to strengthen communities' ownership rights and capacity to use and manage wildlife can increase the benefits that communities derive from protected areas (Blackburn *et al.*, 2016). Conversely, insurance schemes can provide compensation for crop damage caused by wildlife, reducing the costs associated with area-based protection (Hoare, 2015). Concurrently, there is increasing interest in interventions which aim to reduce demand for illegal wildlife products (Veríssimo and Wan, 2018). Such strategies are rare, however, as is evidence of whether or how they work (Cooney *et al.*, 2017).

The dominant mode through which decision-makers seek to discourage illegal activity in protected areas is through law enforcement - policing, prosecution, and punishment of wildlife crime (Moreto and Gau, 2017). According to criminological theory, enforcement works to reduce crime through two mechanisms: incapacitation and deterrence. Firstly, punishment (e.g., imprisonment or fines) removes or reduces offenders' capacity to commit crime (Pratt and Cullen, 2005). Secondly, enforcement deters potential or past rule-breakers from undertaking future illegal activity by increasing the risks of punishment, assuming those punishments are sufficiently severe, certain, and swift (Nagin, 2013b). An entire literature has been devoted to understanding whether these mechanisms operate in practice and has found mixed results. For example, the relationship between enforcement and successful deterrence of crime is nuanced and varies with context (Nagin, 2013b). In protected area contexts, enforcement has hitherto received relatively little attention, but as the primary strategy for addressing crime it is essential to understand if the approach is effective, and the mechanisms responsible for effectiveness.

The principal means by which protected area managers operationalise law enforcement at site-level is through ranger-led law enforcement patrols (Henson *et al.*, 2016). Teams of park rangers police protected sites (i.e., rangers travel around parks on a regular basis), aiming to find and arrest offenders, seek out and remove passive hunting devices (e.g., snares), and/or discourage crime from taking place through their presence. Whilst the objectives of ranger patrols are broadly similar between sites, other aspects may vary, including, for example, size of patrols (number of rangers), professionalization (e.g., employed vs. community rangers), powers of arrest, whether rangers are equipped with firearms, transport (e.g., foot patrols vs. vehicle patrols), and provision of training and equipment (Henson *et al.*, 2016). Rangers may undertake daily patrols from a park headquarters or patrol post, or aim to increase their presence by carrying out multi-night patrols or by being transported to remote locations before commencing patrols (Plumptre, 2019). Specialised 'intelligence-led policing' – targeting criminal networks and prolific offenders through analysis of intelligence – is becoming popular in some contexts (Moreto, 2015), but traditional ranger patrolling is still the dominant paradigm in the majority of protected areas. Accordingly, patrol activities often constitute the single largest expenditure for protected area management authorities (e.g., >66% of park management budgets in Uganda (Plumptre, 2019)). However, the use of ranger patrolling to reduce wildlife crime has been neglected as a subject of study.

Most studies that have addressed enforcement in protected areas have examined correlations with conservation outcomes across sites. In general, the presence and a greater degree of enforcement appears to be an important factor in determining protected area effectiveness. For example, the presence of rangers is the best predictor of great ape survival in tropical African parks (Tranquilli *et al.*, 2012), and ranger density correlates with the effectiveness of tropical parks, in general (Bruner *et al.*, 2001). Comparable relationships have been documented in marine ecosystems (e.g., marine sanctuaries in the Philippines (Walmsley and White, 2003)). Comparisons within the same sites over time also suggest increased enforcement is important for effective protection. For example, increased allocation of resources to enforcement showed a strong negative relationship with illegal killing of elephants in Zambia (Jachmann and Billiouw, 1997), and reduced anti-poaching effort correlated with a reduction in the abundance of African buffalo, elephant, and black rhino in Serengeti National Park, Tanzania (Hilborn *et al.*, 2006). While such temporal relationships may be spurious autocorrelations (Dobson *et al.*, 2018), model-based approaches also suggest that funds should be disproportionately invested in enforcement rather than expansion of protected areas to achieve conservation outcomes (Kuempel *et al.*, 2018).

Whilst evidence suggests that law enforcement is important for effective protected areas, the mechanisms involved are poorly understood. For example, there is limited appreciation of whether and how patrols deter illegal activity. Rigorous empirical analyses are rare, but recent studies attempting to demonstrate a deterrence effect of patrols, conducted in individual sites, have found contrasting results (Beale *et al.*, 2018; Moore *et al.*, 2018). In part, demonstrating deterrence is difficult because little is known about the temporal and spatial scales over which the mechanism operates in protected areas, or how it might vary with context (Dobson *et al.*, 2018). The effects of fines or imprisonment arising from detection by patrols are even less well studied (Leader-Williams and Milner-Gulland, 1993). In contrast, the effects of snare removal by patrols has been assessed. For example, areas in a Sumatran park with a greater frequency of patrols showed a lower occurrence of snare traps in succeeding years (Linkie *et al.*, 2015). However, in general, rates of detection of snares by patrols are low (O'Kelly *et al.*, 2018), and patrols aiming to remove the threat of snaring have been largely ineffective (Gray *et al.*, 2017). The preponderance of sustained illegal activity in protected areas in which enforcement is present, and the attendant continued declines in biodiversity, suggests a clear and urgent need to develop a better understanding of how enforcement operates to understand why it is failing in certain contexts.

A major hindrance to research into the effects of patrolling has been limited understanding of the amount and extent of patrolling currently practiced within sites. Yet the few studies that have explicitly addressed patrol activity have reported inadequate patrolling, which may explain why patrols are often ineffective. For example, only small proportions of protected areas throughout the Greater Virunga Landscape in central Africa were regularly patrolled, and significant areas were rarely or never patrolled (Plumptre *et al.*, 2014). In these contexts, levels of arrests or snare removal are unlikely to be sufficient to reduce the threat of illegal activity to biodiversity, and deterrence may be weak. However, outside of a few studies conducted at relatively small scales, understanding is limited. Indirect evidence suggests that inadequate patrolling may be common. Protected areas frequently receive limited financial support (Watson *et al.*, 2014). Exacerbating the issue, less than 12% of global spend on PA management is in developing countries (Balmford *et al.*, 2003), where the need is arguably greatest. Consequently, sites often have insufficient funding for patrol activities (e.g., for ranger salaries, equipment or training) and managers without the capacity or skills to deploy resources effectively (Di Minin and Toivonen, 2015). Moreover, and because of lack of funding, rangers may lack motivation necessary to achieve patrol targets (Ogunjinmi *et al.*, 2009). It is imperative to develop a broader understanding of how patrolling is practiced in protected areas to enable research, assess whether activity is sufficient to reduce illegal activity and ensure limited resources are deployed efficiently.

1.2.2 Monitoring and management of law enforcement in protected areas

Understanding of the extent and effectiveness of patrolling has been constrained by historically poor monitoring of features of interest within protected areas. Adequate monitoring is crucial to inform decision-making, especially if the aim is to evaluate and improve management activities (Lyons *et al.*, 2008). In protected areas, managers and researchers need information on the state of natural resources, use of those resources (both legal and illegal), and enforcement responses to illegal use. Yet protected sites are generally large, remote and inaccessible (Joppa and Pfaff, 2009). Consequently, collecting data within sites can be challenging, and available data are scarce (Bertzky and Stoll-Kleemann, 2009). Most protected area monitoring takes place remotely using satellites and studies of effectiveness often address features that can be monitored accordingly (e.g., land-use and land-cover change (Tesfaw *et al.*, 2018)). However, the commonest and most serious threats to biodiversity in these contexts are difficult to monitor remotely and generally require data collected in situ (e.g., over-hunting of species (Schulze *et* *al.*, 2018)). Illegal activity, which is illicit and therefore conducted covertly, is inherently and particularly hard to monitor, because offenders have strong incentives to withhold information, and in protected area contexts wildlife crimes are rarely observed or reported (Gavin *et al.*, 2010). Even monitoring of seemingly innocuous activities, including that of ranger patrols, has been crude and inadequate, because managers lacked the capacity to implement more sophisticated systems. For example, until relatively recently, patrol activity was commonly recorded as number of days for which rangers were on patrol, ignoring temporal and spatial variation in activity within parks. (Leader-Williams *et al.*, 1990)

Ranger-based monitoring, defined as the collection of data – on illegal activities, focal species and law enforcement responses – by ranger teams whilst on patrol, has been developed to fill this monitoring gap (Gray and Kalpers, 2005). As rangers travel around protected areas as part of their day-to-day work of enforcement, ranger-based monitoring can, in theory, provide a regular source of data on features throughout sites, without the requirement for additional survey effort (Gavin *et al.*, 2010). In tandem, Law Enforcement Monitoring (LEM) tools, such as MIST (Management Information SysTem), have been developed to facilitate standardised data collection, storage, management, analysis and mapping of ranger-based monitoring data (Stokes, 2010). LEM tools draw on technological advances, such as the advent of cheap, accessible radio navigation-satellite services (e.g., GPS), which permit geotagging of ranger observations and fine-scale monitoring of patrol activity in space and time, and geographic information systems software for managing spatial data. LEM tools may also exploit GPS-enabled digital devices for recording observations by rangers (e.g., smartphones running specialist software, such as CyberTracker (CyberTracker Conservation, 2013)). In general, most LEM tools also explicitly aim to generate information which can be used to inform patrol management (e.g., MIST (Stokes, 2010)).

In recent years, the number and variety of LEM tools has proliferated. Several countries have adopted and adapted LEM tools for proprietary nation-wide use. For example, the LAWIN and MSTrIPES systems facilitate ranger-based monitoring in >150 forests sites threatened by deforestation throughout the Philippines (DENR, 2016; USAID, 2018), and tiger protection sites throughout India (MoEF, no date), respectively. LEM tools have also emerged with a greater focus on rapid-response patrolling, which integrate data from a variety of data sources, including ranger-based monitoring, and which provide platforms for near real-time data visualization. For example, the EarthRanger system has been implemented in 11 sites throughout sub-Saharan Africa (Vulcan Inc., 2018), and the C-more platform provides a similar service for monitoring of natural resources and threats in 21 parks across South Africa (SANParks, 2018; CSIR, 2019).

SMART (Spatial Monitoring And Reporting Tool), which exploits ranger-based monitoring in protected areas via an LEM tool, has emerged as the market leader. SMART differs from its predecessor, MIST, in that it is explicitly marketed as an approach for monitoring *and* managing law enforcement in protected areas, which aims to improve patrol effectiveness (i.e., increase the degree to which patrols are successful in reducing illegal activity) (SMART Partnership, 2018). SMART purports to be able to achieve this aim through implementation of an adaptive management cycle for continuous learning from and evaluation of ranger-based monitoring data, supported by capacity building activities and enforcement management standards (SMART Partnership, 2017). Implemented management activities may include, for example, mechanisms for motivating rangers to patrol farther or for longer durations by linking patrol effort results from the LEM tool to performance evaluations and associated incentives, or for increasing detection and deterrence of crime by targeting patrols towards areas which, according to SMART-derived estimates, experience high levels of illegal activity. Consequently, SMART is increasingly popular as an intervention for improving patrol effectiveness in protected areas globally. Since launching in 2011 SMART has grown rapidly, with >600 formally implemented terrestrial and marine sites in >55 countries as of 2018, and 11 countries adopting SMART nation-wide (i.e., across all state-managed parks) (SMART Partnership, 2018). Broadscale, standardised monitoring of ranger patrol activity and wildlife crime in protected sites around the world, via SMART, provides a unique opportunity to evaluate the extent and effectiveness of patrolling in a variety of contexts, which has yet to be realised.

1.2.3 Broadening the range of designs and methods for conservation evaluation

It is unclear whether interventions which aim to improve patrol effectiveness through patrol monitoring and management, such as SMART, are achieving their goals. Despite substantial resources expended in development and implementation of SMART in protected areas globally, there has been no rigorous evaluation. The only direct study, an analysis of outcomes in four Russian parks in which SMART had been implemented, reported that patrol effort increased across all sites over the four years following implementation (Hötte *et al*. 2016). However, the study did not examine the extent to which SMART contributed to the observed changes, nor did it include rigorous analysis of effects on levels of illegal activity. Understanding of whether and how SMART contributes to patrol effectiveness is essential to determine whether allocation of scarce resources to SMART represents a wise investment, and to understand how the intervention can be improved if it is failing (Baylis *et al.*, 2016).

Causal attribution of the role of interventions in producing change is essential but rare in conservation, because the context poses unique challenges, which render common impact evaluation methods unsuitable (Ferraro and Pattanayak, 2006). The dominant approach to causal attribution advocated in the literature, counterfactual inference, involves estimation of what would have happened in the absence of the intervention, either through experimental methods or comparison with quasi-identical participants (Ferraro, 2009; Bottrill and Pressey, 2012; Jones and Lewis, 2015). However, most conservation evaluations are conducted ex post on interventions which were implemented non-experimentally across large, complex settings, without evaluation design in mind (Margoluis *et al.*, 2009a). Rarely do such interventions lend themselves to the rigid requirements of quasi-experimental counterfactual designs. For example, interventions seldom represent homogenous treatments, but vary between participants and through time in multiple dimensions, either requiring impractically large sample sizes or confounding assumptions (e.g., finding truly quasi-identical non-treated participants may be difficult or impossible) (Ferraro and Pressey, 2015). Conservation also often acts on units several steps removed from biodiversity targets in complex environments with multiple, confounding factors, generating time lags and subtle effects which can be difficult to detect (Howe and Milner-Gulland, 2012). SMART, for example, ultimately aims to reduce criminal activity by offenders but is designed to more immediately influence the behaviour of rangers and managers, and how the intervention is implemented will vary according to the needs and capacity of participant sites (SMART Partnership, 2017). Moreover, few evaluations have the time or resources to gather primary outcomes data or have access to secondary data gathered in a standardised fashion at an appropriate scale (Margoluis *et al.*, 2009a). SMART is a case in point: protected areas without ranger-based monitoring will generally lack appropriate secondary monitoring data for comparison with SMART sites, rendering estimation of what would have happened in the absence of SMART challenging.

Whilst counterfactual approaches should be the gold standard for impact evaluation, interventions will continue to be implemented in contexts for which such approaches are unsuitable (Bonell *et al.*, 2011). Consequently, it is essential that the conservation community develops and employs rigorous yet flexible methods for causal attribution to ensure interventions are appropriately evaluated. Moreover, the assumptions underlying counterfactual approaches are easily confounded and, even when assumptions are met, such methods provide little information on how effects are produced (Stern *et al.*, 2012). Examining the mechanisms responsible for an effect, and how mechanisms interact with context, is essential to understand how an intervention might be replicated, if effects will be reproduced, or why an intervention failed to achieve an effect (Moore *et al.*, 2015)

Policy arenas operating under similar constraints, such as public health, education and development, have long employed flexible, mixed-methods designs for causal attribution, which have less stringent requirements than counterfactual approaches, and which provide insights into how effects are produced that can be used to improve interventions (Craig *et al.*, 2008; Stern *et al.*, 2012). Such designs could usefully be applied to advance impact evaluation in conservation in general and to inform rigorous evaluation of SMART specifically. For example, in development, despite a recent increase in randomised control trials (Donovan, 2018), the minority of interventions lend themselves to experimental or quasi-experimental designs (Stern *et al.*, 2012). This has led to increasing use of evaluation designs for linking causes and effects derived from the social sciences, such as theory-based, case-based and participatory approaches (Stern *et al.*, 2012).

Theory-based approaches to impact evaluation, such as contribution analysis and process tracing, which draw on generative causation frameworks for causal attribution (i.e., identifying the mechanisms that explain effects), may have specific utility for conservation. Such approaches look for a connection between causes and effects through identification or confirmation of evidence, drawn from a variety of sources, consistent with causal processes or 'chains', rather than inferring causality on the basis of differences between identical cases (i.e., counterfactual approaches) (Pawson, 2006). Consequently, theory-based approaches first necessitate in-depth analysis of the theoretical basis for how interventions are assumed to work, using methods such as theory of change (Van Belle *et al.*, 2010). Theory-based approaches can provide evidence of whether an intervention contributed to an effect, explain why, and generate strongly generalizable results, but at the expense of internal validity (e.g., such approaches are particularly subject to biases, such as selection bias) and the ability to estimate the extent of an effect (Stern *et al.*, 2012).

Specific theory-based methods of particular relevance for impact evaluation of SMART include contribution analysis and dose-response analysis (Stern *et al.*, 2012; Rogers, 2014). Contribution analysis is a structured approach for assessing whether an intervention contributed to observed impacts by verifying a theory of change with empirical evidence (Mayne, 2012). Similarly, dose-response analysis involves examining associations between the intensity at which an intervention is applied (the dose) and its outcomes (the response), in conjunction with a well-developed theory of change (Rogers, 2014). The approach is designed to provide confirmatory evidence of causal effects rather than to infer causality (Hill, 1965). One benefit to conservation of these approaches is that they can either confirm the theory of change or suggest refinements based on in-depth understanding of impact mechanisms and analysis of evidence (Rogers, 2014).

Finally, whilst heterogenous treatments and complex environments pose less of a challenge for non-counterfactual approaches to impact evaluation, this complexity must still be accounted for to ensure correct interpretation of conservation outcomes (Ferraro and Pressey, 2015). For SMART, this issue may be acute, as adaptive management is often aspired to in conservation but rarely achieved in practice (Nichols and Williams, 2006). Treatment heterogeneity is hitherto poorly appreciated in conservation, but methods from other policy arenas could usefully be exploited within the field of conservation. For example, implementation evaluation – examining the implementation, functioning and setting of an intervention – is a common practice in fields where practitioners are trying to effect difficult behaviour change in highly variable conditions (e.g., education (Humphrey *et al.*, 2016), public health (Escoffery *et al.*, 2016), and behavioural medical interventions (Oakley *et al.*, 2006)).

1.3 Aims and objectives

This thesis has two primary aims. Firstly, I aim to improve the evidence base underpinning the use of ranger-led law enforcement patrols to reduce illegal activity in protected areas. In so doing I aim to contribute to the debate over whether patrolling, as currently practiced, represents an effective response to wildlife crime, and to provide practical recommendations for how effectiveness could be improved. Secondly, I aim to evaluate a popular intervention designed to improve patrol effectiveness, SMART, drawing on flexible, formative approaches from policy arenas outside conservation. In so doing I aim to advance methods for rigorous evaluation in conservation and to provide evidence which can be used to judge whether SMART represents a wise allocation of resources, to improve the intervention, and to inform decisions about future deployment.

To achieve these aims my main objectives are:

1. to assemble monitoring data on patrol activity and wildlife crime from a broad sample of protected areas globally;
2. to investigate how patrolling is practiced in protected areas, at present;
3. to explore whether and in what contexts patrols effectively deter wildlife crime;
4. to illustrate and interrogate how SMART aims to reduce illegal activity;
5. to develop a framework for evaluating treatment heterogeneity in conservation interventions, and apply the framework to SMART;

6. to evaluate whether SMART contributed to improved patrol effectiveness; and,
7. to provide practical advice to stakeholders for improving patrolling and patrol management, based upon the research outcomes.

1.4 Thesis outline

The thesis is structured as follows:

Chapter 2 uses a theory of change approach to describe SMART and the causal pathways through which it acts to reduce illegal activity via patrolling, and reviews extant evidence and relevant literature to interrogate assumptions underlying these pathways. This understanding is used to frame evaluation of SMART and to identify priorities for research into patrolling (Fig. 1.1). I find that counterfactual methods for establishing whether SMART contributed to a reduction in illegal activity will be unsuitable. One promising alternative, contribution analysis, involves constructing and verifying a theory of change with empirical evidence. In subsequent chapters (3-5), I develop evidence for verifying whether successive aspects of the theory held true in practice and investigate priority areas of research, including how SMART was implemented (chapter 3), how patrolling is practiced and whether it is improved by SMART (chapter 4), and whether and how patrols act to reduce illegal activity (chapter 5).

Chapter 3 presents a novel framework for describing heterogeneity among implemented conservation interventions and demonstrates the framework's utility by applying it to SMART. The framework considers three critical aspects of implementation: Activities, Inputs, and Moderators. I examine how faithfully SMART was implemented in practice (Activities), how implementation was achieved (Inputs), and whether the contexts in which it was implemented were conducive to success (Moderators). I find that SMART implementation varied between participating protected areas and from programme designs. I use this understanding to interpret outcomes in subsequent chapters and to make recommendations for future deployment of the intervention.

Chapter 4 investigates how patrolling is practiced in protected areas. I exploit recent, widespread deployment of patrol monitoring, via SMART, to provide the first global analysis of spatiotemporal patrol presence in protected areas. I estimate spatial and temporal coverage provided by ranger patrols within and across sites, evaluate results with respect to industry benchmarks, and estimate change over time. I also assess factors influencing patrol presence, including dose-response analysis of the effects of SMART management mechanisms on rates of change. I find that patrols typically provided very low spatial coverage of protected areas, frequently fell short of industry targets, and were constrained by limited budgets for patrolling. In some contexts, coverage was improving, but these changes may be unrelated to implementation of SMART. My findings suggest that levels of patrolling globally may be insufficient to monitor or reduce illegal activity, and increased financial support is urgently required.

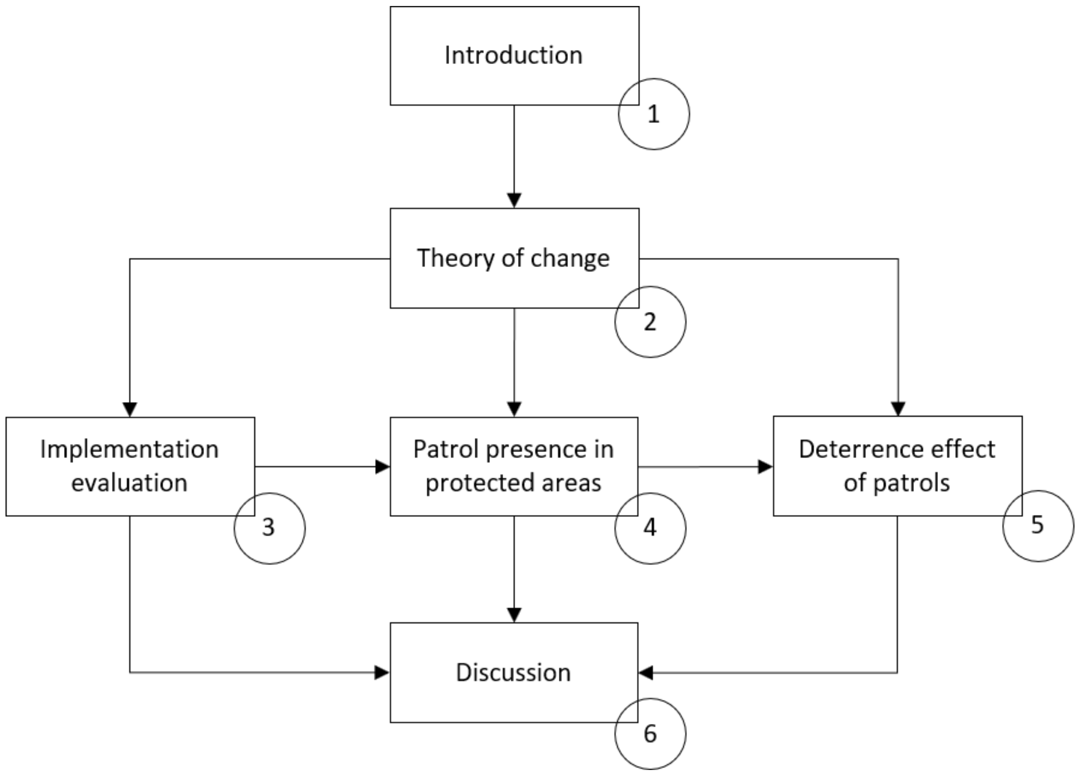


Figure 1.1. Thesis structure by numbered chapter, showing logical flow between chapters. Chapter 2 frames evaluation of SMART and identifies priorities for research intro patrolling (Fig. 1). Subsequent chapters develop evidence for verifying successive aspects of SMART's theory of change (3-5) and investigate priority areas of patrol research (4-5). Chapter 6 discusses findings and synthesises results of SMART's evaluation.

<u>**Chapter 5**</u> explores whether and in what contexts patrols effectively deter wildlife crime, by applying a novel metric of deterrence, that had reliably identified deterrence using simulated data, to real patrol data for the first time. Using monitoring data collected in four diverse protected areas around the world, I assess the effect of changes in patrol presence on changes in illegal activity. I also examine whether differences in deterrence can be explained by site-level characteristics, such as habitat type, and explore methods for applying the deterrence metric to real data, including the effect of temporal scales. I find that patrol presence may have reduced illegal activity, but the relationships were weak, inconsistent and context-dependent. The absence of consistent evidence of deterrence could indicate that patrols do not

reliably deter crime or that my application of the metric was not sufficiently sensitive. Questions remain for future applications, including appropriate spatial scales.

Chapter 6 discusses the thesis's key findings and conclusions and synthesises the different elements of SMART's evaluation. I evaluate whether SMART contributed to reduced illegal activity by verifying whether evidence developed in the previous chapters confirmed the theory of change. I provide recommendations for improving the effectiveness of patrolling and SMART, respectively, and suggest avenues for future research.

Chapter 2

Framing evaluation of a tool for improving patrol effectiveness in protected areas: A theory of change for SMART

Abstract

Ranger-led law enforcement patrols are the primary response to illegal natural resource use in protected areas. Until recently, protected area managers lacked adequate monitoring data to manage patrols effectively to address this threat, including where illegal activity is occurring, or how patrols are performing in response. Law Enforcement Monitoring (LEM) tools, which exploit data collection by rangers on patrol, were developed to fill this gap. One LEM tool, SMART, which aims to improve patrol effectiveness through implementation of adaptive management, is increasingly popular in protected areas globally. However, there has been an absence of rigorous evaluation of LEM tools. The starting point for any evaluation is a clear description of what an intervention entails, and an understanding of how it is assumed to act to produce change. This knowledge is necessary to frame outcome evaluation and interpret results, and is an essential prerequisite for theory-based impact evaluation approaches, such as contribution analysis. An understanding of where evidence for effects is weak or lacking is also necessary to prioritise research efforts. Here, using a participatory theory of change approach, I describe SMART, and illustrate the causal pathways through which it acts to reduce illegal activity. I also interrogate assumptions underlying these pathways, by assessing the strength of extant evidence. SMART's theory of change provides a clear description of what implementation of the tool involves, including inputs and activities, and the chain of intermediate outcomes necessary for implementation to achieve impact via changes to the behaviour of patrol managers, rangers and perpetrators of illegal activity. This understanding can be used to guide and interpret rigorous evaluation, such as whether the intervention was implemented as described, and whether the chain of results occurred. Interrogation of the theory of change also highlighted priorities for evaluation. I found limited understanding of whether patrols reliably deter illegal activity, and weak evidence that data collection by rangers on patrol is a robust method for monitoring illegal natural resource use. It is essential to address these weaknesses, in combination with rigorous evaluation framed by the theory developed here, to ensure patrols are deployed effectively to combat threats to biodiversity in protected areas worldwide.

2.1 Introduction

Illegal use of natural resources is a major threat to biodiversity in protected areas, particularly in developing countries (Schulze *et al.*, 2018). Poaching, for example, has driven defaunation in sites throughout Southeast Asia (Gray *et al.*, 2017) and contributed to declines in large mammals in African protected areas (Craigie *et al.*, 2010). To address this threat, effective enforcement of laws is crucial (Rowcliffe *et al.*, 2004; Tranquilli *et al.*, 2014). Ranger-led patrols, which aim to detect and deter illegal activity, are the principal mechanism through which protected area managers seek to enforce laws (Henson *et al.*, 2016). Indeed, levels of enforcement, such as guard density and number of patrols, are strongly associated with reduced threats and higher protected area effectiveness (Bruner *et al.*, 2001; Hilborn *et al.*, 2006; Moore *et al.*, 2018). However, law enforcement in protected areas is expensive (Jachmann, 2008) and resources available for patrolling are often limited (Nolte, 2016). Consequently, effective management and efficient deployment of ranger patrols is essential (e.g., to target patrols towards illegal activity 'hotspots' (Critchlow *et al.*, 2017)).

Until recently, patrol managers lacked data necessary to effectively address illegal activity in protected areas (Stokes, 2010). Effective management of natural resources is contingent upon adequate monitoring, which provides information about the state of the system under management, and how it responds to management activities (Lyons *et al.*, 2008). In protected areas, managers need information on aspects such as the state of biodiversity, the extent and distribution of exploitation of biodiversity (both legal and illegal), and enforcement activities undertaken in response to illegal exploitation (Gray and Kalpers, 2005). However, in protected areas, which are often large, inaccessible and poorly funded, monitoring is notoriously challenging, particularly of illicit and frequently unreported illegal activity (Gavin *et al.*, 2010). Even monitoring of relatively straightforward features, such as the activity of rangers whilst on patrol, has been largely inadequate.

To address these shortcomings, conservationists have developed tools to improve monitoring and management of enforcement in protected areas, collectively termed Law Enforcement Monitoring (LEM) tools (Hötte *et al.*, 2016). LEM tools, such as MIST (Management Information SysTem) and SMART (Spatial Monitoring And Reporting Tool), exploit ranger-based monitoring – the collection of data by rangers whilst on patrol – including observations of illegal activity, when encountered, and evidence of rangers' own activities, such as patrol routes. LEM tools facilitate standardised collection, storage, analysis, mapping and reporting of ranger-collected data, which generates information that can be used to inform patrol management (Stokes, 2010). The SMART system takes LEM tools further, by providing training and standards to place monitoring information within an adaptive management cycle, which aims to iteratively improve the effectiveness of patrols in reducing illegal activity (SMART Partnership, 2017).

Use of LEM tools in protected areas is now ubiquitous; however, despite substantial resources expended in development and implementation, there has been an absence of rigorous evaluation. Increases in patrol activity have been recorded in sites in which LEM tools have been implemented (Gray and Kalpers, 2005; Jachmann, 2008) but without causal attribution. SMART, for example, has been implemented in >600 sites across >55 countries globally since launching in 2012, with 11 countries adopting the tool nation-wide (SMART Partnership, 2018). Positive outcomes were apparent in four sites using SMART in Far East Russia (Hötte *et al.*, 2016), but it is unclear whether these results are common throughout SMART sites or if the tool was responsible for the change observed. Rigorous evaluation, which indicates whether, how and why LEM tools contribute to positive change, is essential to determine whether they represent a wise investment, and to understand how LEM tools can be improved if they are failing (Baylis *et al.*, 2016).

The starting point for any evaluation is a clear description how an intervention is assumed to act to produce change (Margoluis *et al.*, 2009a; Moore *et al.*, 2015; Baylis *et al.*, 2016). This knowledge is necessary to frame evaluation and interpret results (De Silva *et al.*, 2014). In practice, the theory underlying how interventions are assumed to work is rarely made explicit (Moore *et al.*, 2015). Decision-makers implement solutions to proximate problems (e.g., limited resources for patrolling) aiming to achieve ultimate goals (e.g., reduced illegal activity), without a clear understanding of how the two are connected. Consequently, an essential first step in evaluation is to map out the 'missing middle' between what an intervention does and what it aims to achieve (Biggs *et al.*, 2017). Moreover, formative evaluation, which develops evidence which can be used to improve interventions and predict whether effects can be expected if the intervention is replicated, also requires a clear description of what implementation involves and the contexts within which it occurs (Montgomery *et al.*, 2013; Moore *et al.*, 2015).

An understanding of the theory and assumptions underlying interventions is also a pre-requisite for theory-based approaches to impact evaluation, which may be particularly relevant for LEM tools. Sites without law enforcement monitoring will, by definition, lack comparable monitoring data required to assess outcomes in a comparison group, and data independent of LEM tools in protected areas are rare. Consequently, counterfactual designs, the gold standard for post-hoc casual attribution, are impracticable. One promising alternative approach is theory-based evaluation: identifying patterns of evidence consistent with understanding of causal relationships (Stern *et al.*, 2012; Rogers, 2014). Contribution analysis, for example, involves interrogating assumptions underlying theory and verifying weak assumptions with empirical evidence (Mayne, 2012). Such approaches could usefully be applied to LEM tools, but necessitate understanding of the theoretical basis for how interventions are assumed to work (Van Belle *et al.*, 2010).

Various diagrammatic methods have been developed to model the relationship between programs and impacts (Schwartz *et al.*, 2012). One method – theory of change – is increasingly common as a tool for framing design, monitoring and evaluation of interventions (e.g., in development (Vogel, 2012)). Theory of change approaches require explicit illustration of causal links between intervention activities, immediate outputs, intermediate outcomes and ultimate impacts, and allow for feedbacks and multiple levels of interaction between programme components (Biggs *et al.*, 2017). Theory of change also requires explicit articulation of assumptions underlying each step in the causal chain, providing an opportunity for these assumptions to be interrogated (White, 2009). Reviewing assumptions with reference to existing evidence can highlight weakness in design, if evidence is weak, or identify priorities for evaluation, if evidence is lacking. Finally, theory of change approaches also emphasise working with stakeholders to produce a shared understanding of how the intervention is thought to work (De Silva *et al.*, 2014). The approach is less common in conservation but has been used to guide design (Morrison, 2015; Fleishman *et al.*, 2016; Biggs *et al.*, 2017) and evaluation (Johnson *et al.*, 2016; Stebbings *et al.*, 2016).

Here, I illustrate and interrogate how a major LEM tool – SMART – acts to produce change. First, I use a participatory theory of change approach, in collaboration with SMART's developers and implementers, to describe what implementation entails and the causal pathways through which the intervention acts to reduce illegal activity via patrol monitoring and management. This understanding can be used to guide rigorous evaluation of SMART and interpret outcomes, including the expected chain of results. Second, I articulate assumptions underlying SMART's theory of change, and assess the strength of extant evidence supporting these assumptions, drawing on the literature, published best practice and expert knowledge. In so doing I identify gaps in understanding and weaknesses in design to highlight priorities for evaluation.

2.2 Methods

2.2.1 A theory of change for SMART

To develop a retrospective theory of change for SMART, I followed a multistep, iterative process of specification and validation, in collaboration with stakeholders. An existing but unverified model predicted that implementation of SMART within an adaptive management cycle would lead to an increase in tiger abundance (the focal species), via improvements in patrol performance and a reduction in poaching of the focal species' prey (Hötte *et al.*, 2016). Initially, I drew on this model, SMART documentation and relevant scientific literature, to specify a formal logic model. In April 2016, I convened a workshop with individuals responsible for SMART's development and implementation, representing all the non-governmental organizations collaborating on the programme at that time, and presented the logic model to the participants. During and after the meeting I sort multiple iterations of feedback to produce and validate a full theory of change that accurately reflected the participants' understanding of what SMART implementation entails and the sequence of changes it attempts to make to achieve its goals. The finished model also includes causal linkages amongst programme components and contextual factors external to implementation likely to influence success. Identification of factors external to implementation which may influence successful implementation can aid assessment of whether necessary conditions are in place (Biggs *et al.*, 2017) and recognition of confounding factors during evaluation.

2.2.2 Strength of assumptions underlying SMART's theory of change

I articulated and sort feedback on assumptions underlying causal links in the theory of change. Assumptions describe the logic of why and how components are expected to effect change on each other (White, 2009) (e.g., attempting to improve ranger motivation by tying patrol performance to incentives assumes that ranger motivation is poor or can be improved, and that incentives are an effective means to improve motivation). I assessed the strength of extant evidence supporting assumptions underlying casual links in SMART's theory of change by reviewing evidence from the conservation and human resource management scientific literature and documented best practice for law enforcement in protected areas. I also obtained the expert input of SMART's developers, who have extensive experience of implementation, collected during unstructured interviews over the course of the study. To highlight evaluation priorities and design weakness, I assigned each assumption a confidence score consistent with the strength and availability of supporting evidence, according to specified criteria (Table 2.1; Simkin, 2016), and colour coded causal links in SMART's theory of change accordingly (high confidence=green, medium confidence=orange, low confidence=red).

Table 2.1. Confidence scores and associated criteria for judgement of evidence supporting SMART's casual assumptions.

Confidence score	Criteria
High	Strong evidence from the literature, documented best practice in same contexts, or the expert knowledge of programme developers and implementers.
Medium	Mixed evidence from the literature or best practice, or evidence from different contexts which may not apply.
Low	Weak evidence or unknown.

2.3 Results

2.3.1 A theory of change for SMART

SMART's theory of change (Fig. 2.1) describes: a) the ultimate 'Aim' SMART implementation attempts to achieve, via changes to immediate 'Outputs', intermediate 'Outcomes', and long-term 'Impacts', and causal linkages amongst them; b) 'Inputs' and 'Activities' comprising implementation of the tool; and c) 'Enabling conditions' likely to influence success.

a) Outputs, Outcomes, Impacts and Goal

The ultimate aim of SMART implementation, which involves improvements to monitoring and management of law enforcement in protected areas (outlined below), is to improve conservation of threatening biodiversity in protected areas, either of the target species or habitats directly or via supporting species (e.g., prey) or habitats. The primary pathway by which implementation aims to achieve this goal is by motivating rangers, improving the quantity, quality and timeliness of information on threats and enforcement responses, and optimizing patrol deployment. In theory, these outputs improve aspects of ranger and patrol performance, specifically: (1) the amount of work undertaken by rangers whilst on patrol, 'ranger performance' (e.g., increased distance patrolled, or time spent on patrol); and (2) effective and efficient deployment of patrols in space and time, 'patrol performance' (e.g., increased coverage of sites, targeting of high-risk areas, or unpredictability of patrol presence).

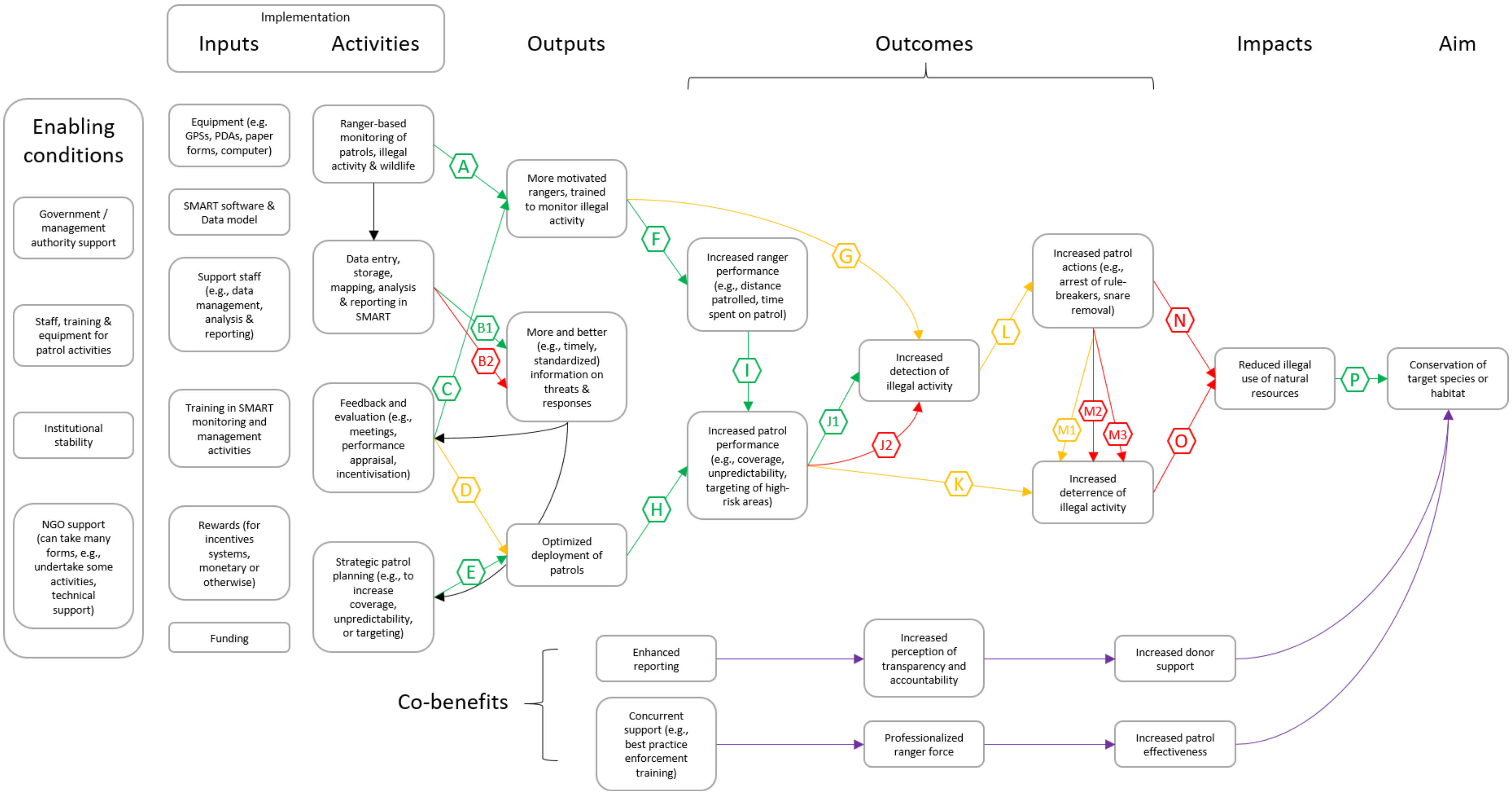


Figure 2.1. Theory of Change for implementation of SMART to improve protection of species threatened by illegal activity in protected areas. Implementation involves *Inputs* (e.g., training) which support monitoring and management *Activities* (e.g., ranger-based monitoring, and ranger performance evaluations), to produce near-term *Outputs* (e.g., more motivated rangers), which effect change in a sequence of short- and medium-term *Outcomes* (e.g., increased patrol coverage and increased deterrence of illegal activity), for long-term *Impact* (reduced illegal activity). Black arrows indicate information flows between activities. Coloured arrows indicate causal links between components, and confidence in assumptions: high confidence (green), medium confidence (orange), low confidence (red). See Table 2.2 for associated assumptions coded by letter. Success is supported by *Enabling conditions* (e.g., government support) and can be influenced by *Co-benefits* of implementation (purple arrows).

Increased, targeted and/or unpredictable presence of patrols in space and time enhances both detection and deterrence of illegal activity, aided by rangers' enhanced capacity to identify illegal activity signs. Enhanced detection permits more patrol actions (e.g., arrests of rule-breakers or snare removal), which reduces the number and capacity of rule-breakers and contributes to deterrence. Fewer potential rule-breakers and a higher proportion deterred from undertaking illegal activity results, ultimately, in a reduction in illegal killing.

SMART implementation also has the potential to benefit area-based species conservation through alternative, indirect pathways: (1) by enhancing reporting of conservation activities, implementation can improve perceptions of transparency and accountability, leading to increased donor support; and (2) rangers trained in use of SMART often receive concurrent training in best practice enforcement techniques, resulting in a more professional ranger force and more effective patrols.

b) Inputs and Activities

Implementation of SMART involves a combination of ongoing law enforcement monitoring and management activities. Rangers using SMART collect time-stamped position records whilst on patrol, in conjunction with recording of (1) metadata, such as transport type and no. of personnel, (2) opportunistic observations of signs of illegal activity and wildlife, when encountered, and (3) patrol actions (e.g., arrests). On return from patrol, these data are entered into specialised GIS-based software (called SMART), which facilitates storage, mapping and analysis, leading to standardised reporting of trends in patrol activities, threats and wildlife (e.g., patrol routes, ranger performance summaries, or maps of threats). Subsequently, this information is intended to be used to inform enforcement management activities as part of an adaptive cycle of monitoring and assessment, including: (1) strategic planning of future patrols (i.e., targeted deployment to areas of high threat, to systematically increase coverage in space or time, or to increase unpredictability); and (2) mechanisms for feeding results back to frontline staff (e.g., via meetings), evaluating ranger or team performance, and using evaluations to inform incentives schemes. SMART is unique amongst LEM tools in that management activities are intended to be part of implementation, rather than a potential beneficiary of the process (Fig. 2.2).

Monitoring and management activities are supported by inputs, including: SMART software, which is open-source and includes a customisable monitoring data model; equipment (e.g., GPS receivers and paper forms or digital devices for rangers to record observations, and computers for running software and storing data); support staff (e.g., to manage and analyse data, and produce reports); training (e.g., of rangers in data collection, support staff in software, and managers in use of information to inform management); rewards for incentivisation schemes; and, underlying all inputs, funding. Not all inputs are necessarily required to implement activities. At minimum, implementation is feasible solely using software, free training materials, and existing staff and equipment. Implementing sites may also already undertake some or all activities, in which case SMART is intended to improve these processes. Such sites may have previously used simpler tools (e.g., spreadsheets) or software similar to SMART (e.g., MIST).

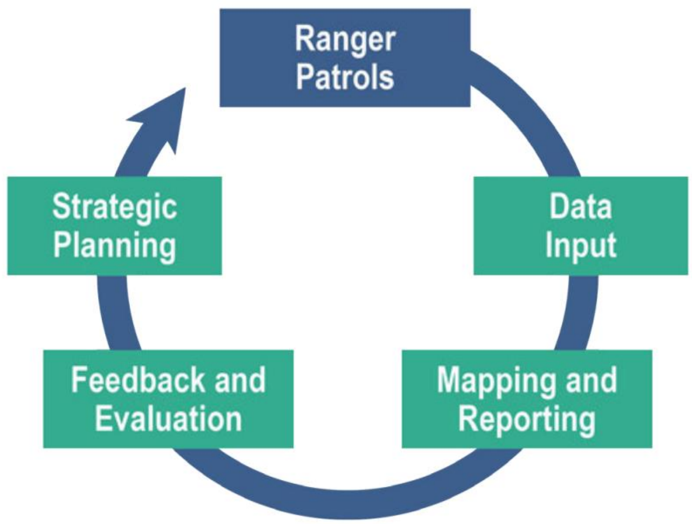


Figure 2.2. Graphic from SMART training materials. SMART's developers intend for the tool to provide a framework for adaptive management of patrol activities, through a continuous iterative cycle of monitoring and assessment (SMART Partnership, 2017).

c) Enabling conditions

Stokes (2010) identified three factors likely to influence the success of LEM implementation: (1) institutional support from relevant government or protected area authorities, (2) adequate resources for enforcement, including rangers, support staff, training and equipment, and (3) site-level institutional stability (e.g., low turnover of involved staff, and sustained support from management). In addition, whilst support from a SMART partner is not a pre-requisite, in practice implementation often involves substantial NGO support, in addition to funding. This support will vary in strength between sites and take various forms, from directly undertaking activities, such as monitoring, data input or reporting, to providing permanent on-site technical support.

2.3.2 Strength of assumptions underlying SMART's theory of change

In summary, there is strong evidence that SMART can improve patrol performance by motivating rangers and optimizing patrol deployment, assuming monitoring and management activities are implemented as intended (Table 2.2). However, evidence that implementation can reliably increase detection and deterrence of illegal activity is mixed and context-dependent.

Improvements in patrol performance have previously been ascribed to LEM tools (Gray and Kalpers, 2005; Jachmann, 2008; Hötte *et al.*, 2016). These studies did not examine the mechanisms responsible for the effect, but evidence from other sources suggests that LEM tools act by increasing ranger motivation or by optimizing the efficiency with which patrols are deployed. Human resource management theory suggests that the increased responsibility, training, evaluation and oversight of employees required by SMART, such as GPS tracking of patrols, should engender greater motivation (Mayo, 1933; Herzberg, 1959), which is strongly related to employee performance (Cascio and Aguinis, 2011). Incentive mechanisms, such as bonuses, can also influence ranger performance (Henson *et al.*, 2016). Likewise, LEM tools can be used to optimise deployment of patrols, by providing managers with information necessary to improve patrol allocation, such as accurate maps of patrol coverage (Stokes 2010). For example, application of methods for improving patrol allocation, which made use of ranger-based monitoring data, increased illegal activity detections by up to 250% (Critchlow *et al.*, 2017).

Improvements in patrol performance will only be achieved if SMART is used for these purposes. In practice, complex interventions are rarely implemented as intended, and frequently undergo modifications to suit local contexts (Craig *et al.*, 2008). For SMART, which is a tool for managers to use or ignore, this issue is potentially acute. Which activities managers implement, and the extent to which activities are implemented as intended may vary according to the needs, capacity and context of each site, and through time following implementation. For example, whilst implementation may always involve ranger-based monitoring, how sites undertake this activity can vary (e.g., in types of data collected, frequency and accuracy of records, and number and extent of patrols). Whether managers will make use of information to inform adaptive management is even less clear. In general, adaptive management of natural resources is rarely achieved in real-world settings (Nichols and Williams, 2006). Anecdotally, the number and fidelity of management activities implemented varies between sites. The process also assumes information flows occur with sufficient celerity to be useful and are delivered to the right people.

Evidence for whether patrol presence can effectively reduce illegal activity is mixed. Multiple studies relate patrol activity to negative trends in illegal activity (e.g., Jachmann (2008)), suggesting patrols are important in addressing crime. However, more recent studies find variable effects on poaching at fine-scales (Barichievy *et al.*, 2017; Beale *et al.*, 2018). Theory suggests rule-breakers will be increasingly deterred from undertaking illegal activity as patrol presence rises (Pratt and Cullen, 2005). However, deterrence theory relies on rule breakers' often inaccurate perception of benefits and risks and supporting evidence for this model is from the global north (Moreto and Gau, 2017). In practice, rangers may not consistently arrest rule-breakers or take appropriate action when encountering illegal activity. Rangers in some sites are incentivised according to frequency of arrests (Henson *et al.*, 2016) or other actions (e.g., snare removal (Becker *et al.*, 2013)), but elsewhere they may lack the capacity or motivation necessary to arrest rule-breakers (Ogunjinmi *et al.*, 2009), avoid potentially lethal confrontations (Moreto, 2016), or even tolerate illegal activity (Lescuyer, cited in Curran *et al.* (2009)). Even if rangers take action, it is unclear whether this acts as a reliable deterrent. Indirect actions such as snare removal have garnered mixed results (Wiafe and Amoah, 2012; Gray *et al.*, 2017). In the global south some justice systems do not have a vested interest in punishing arrested offenders sufficiently severely to deter crime (Moreto and Gau, 2017). Any effect may also be undermined in contexts where authorities are complicit in illegal activity (Kahler and Gore, 2012). Indeed, in countries with high poverty and weak governance approaches relying on deterrence may deteriorate political legitimacy and, perversely, increase illegal killing (Redpath *et al.*, 2017).

Finally, there is weak evidence that SMART can provide a reliable source of information on threats. The primary objective of rangers on patrol is to seek out and deter illegal activity, with observations collected opportunistically, generating biases in data for which, at present, it is difficult to account (Keane *et al.*, 2011). For example, as patrols follow accessible routes and target rule-breaking their activity is patchy and non-random. Consequently, estimates of illegal activity risk becoming a function of biases in sampling effort (Moreto *et al.*, 2014). If rule-breakers' behaviour changes in response, as intended, using these data to target high-pressure areas may result in an increasing number of infractions going undetected (Lemieux *et al.*, 2014). Reducing bias is possible (e.g., by ensuring consistent and broad or near-random patrol coverage), but potentially at the expense of deterrence (Keane *et al.*, 2011). Finally, whilst rangers receive training in observing illegal activity, some may resent the additional work or resist adopting the new technology (Sintov *et al.*, 2018).

Table 2.2. Causal assumptions underlying links in SMART's theory of change, associated confidence scores and key supporting evidence.

Code	Assumption	Confidence	Supporting evidence
A	Training and responsibility for monitoring, and knowledge of increased oversight of patrol activities by management, motivates or empowers rangers.	High	Mayo (1933); Herzberg (1959)
B1	Analyses of patrol activity data generated by ranger-based monitoring are accurate.	High	Expert input
B2	Analyses of observations of illegal activity and wildlife generated by ranger-based monitoring are accurate.	Low	Keane *et al.* (2011)
C	Performance evaluations motivate personnel, particularly when tied to incentives.	High	Henson *et al.* (2016)
D	Feedback mechanisms with frontline staff (e.g., meetings) facilitate information flows (e.g., meetings are regular, and staff are encouraged to provide feedback).	Medium	Expert input
E	Lack of information or tools limits managers' capacity to optimise patrol deployment.	High	Stokes (2010)
F	Lack of motivation limits ranger performance.	High	Ogunjinmi *et al.* (2009); Cascio and Aguinis (2011)
G	Rangers trained to use SMART are more likely to observe and record illegal activity whilst on patrol.	Medium	Sintov *et al.* (2018)
H	Inadequate patrol activity can be improved through optimised patrol deployment.	High	Critchlow *et al.* (2017)
I	Poor patrol effort limits patrol effectiveness.	High	Expert input
J1	Inadequate patrol coverage and unpredictability allow rule-breakers to avoid detection (if managers' aim is to increase coverage or unpredictability).	High	Plumptre *et al.* (2014)
J2	Analyses of ranger-based monitoring data reliably predict future illegal activity (if managers' aim is to increase targeting of high-risk areas).	Low	Keane *et al.* (2011)
K	Patrol presence increases deterrence of illegal activity in PAs.	Medium	Jachmann (2008); Kahler and Gore (2012); Barichievy *et al.* (2017)
L	Rangers arrest rule-breakers, when encountered, or take appropriate action when encountering other signs (e.g., removal of snares).	Medium	Becker *et al.* (2013); Henson *et al.* (2016)
M1	Rule-breakers are rational and mindful of the costs and benefits of illegal activity.	Medium	Pratt and Cullen (2005); Moreto and Gau (2017)
M2	Actions such as ground snare removal or burning of poaching camps significantly increase the costs associated with illegal activity borne by rule-breakers.	Low	Gray *et al.* (2017)
M3	Arrests are translated into successful prosecutions, which carry sufficient penalties.	Low	Moreto and Gau (2017)
N	Successful prosecutions remove rule-breakers or their capacity to commit wildlife crime.	Low	Unknown
O	Deterrence effect outweighs other drivers of illegal behaviour.	Low	Cooney *et al.* (2017); Redpath *et al.* (2017)
P	Natural resource use is a significant threat to target species.	High	Schulze *et al.* (2018)

2.4 Discussion

To frame rigorous evaluation of SMART I constructed a theory of change for the tool and assessed the strength of the theory's underlying assumptions. The theory of change provides a clear description of what the intervention entails, which can be used to shape investigation of programme components and inform interpretation of outcomes (Moore *et al.*, 2015). SMART's theory of change revealed that implementation is a complex of inputs (e.g., training) and activities (e.g., patrol planning), and successful implementation may be influenced by multiple external factors (e.g., government support). Interrogation of assumptions underlying the theory of change suggested that these components may vary strongly amongst participants. As a result, outcomes may vary because the intervention differed from what was intended, rather than due to weaknesses in design. Consequently, evaluation should include detailed examination of how and in what contexts SMART is implemented in practice, framed by the understanding developed here. Implementation evaluation in conservation is rare but methods from other policy arenas could usefully be adopted (e.g., from public health (Moore *et al.*, 2015)). In addition, the number of uncontrolled factors suggests that ex post causal attribution may be particularly challenging, so outcome evaluation with less stringent requirements will be essential (e.g., nonexperimental quantitative designs and qualitative sampling (Margoluis *et al.*, 2009a)), or consider drawing on the theory developed here to create a fuzzy predictive model (Game *et al.*, 2018). Importantly, the theory of change also highlighted factors which contribute to success, such as adequate resources for enforcement, which may have a stronger influence on outcomes than implementation of SMART.

The theory of change also revealed that SMART acts via a series of changes to the behaviour of rangers, managers, and perpetrators to achieve its goals. In addition, SMART can influence conservation success through alternative pathways. Consequently, illegal activity may respond subtly to implementation, because of the many barriers to behaviour change (Veríssimo, 2013), and effects may be hard to discern and even harder to attribute to SMART using rigid counterfactual approaches to impact evaluation. Theory-based approaches to impact evaluation, which place less emphasis on estimating the extent of effects, but instead look for confirmatory evidence consistent with causal chains, may be more applicable to SMART (Pawson, 2006). Contribution analysis, for example, which involves verifying a theory of change with empirical evidence, including whether the chain of results occurred, may be particularly relevant (Befani and Stedman-Bryce, 2017). SMART's theory of change elucidates this causal chain, providing a framework for evaluation. Specifically, evaluators should consider successive aspects such as: (1) effects of implementation on ranger motivation and the quantity and quality of information flows, in the short-term; (2) effects of outputs on ranger and patrol performance (e.g., distance patrolled and patrol coverage, respectively) in the medium-term; and (3) effects of patrols on detection and deterrence of illegal activity, and patrol actions, in the long-term. However, results should be interpreted with care, as causal chains in conservation frequently exhibit nonlinearity and can be strongly influenced by heterogeneity and scale (Qiu *et al.*, 2018).

Supporting evidence provided high confidence that SMART can, in theory, act as an effective means to improve patrol performance. However, the mechanisms responsible are unclear (i.e., whether changes to ranger motivation or patrol deployment are more important for outcomes). A better understanding of these mechanisms could enable improvements to the intervention and should be a high priority for evaluation. Nevertheless, rangers in protected areas are subject to multiple stresses (Moreto, 2016) and typically poorly motivated, due to inadequate funding, equipment, staffing, salaries, incentives and promotions (Ogunjinmi *et al.*, 2009). SMART's improved oversight and performance management of staff cannot address all of these issues. Moreover, whilst imposition of patrol tracking via GPS limits potential for misreporting, falsification is possible, and as rangers' primary duty is not data collection errors can occur (Pantel, 2007). Incentivization systems also need to be sustainably and equitably managed to ensure positive impacts (Henson *et al.*, 2016). Equally, whilst optimised deployment of limited resources is essential, outcomes can only be expected if information generated by SMART is used within an adaptive cycle to improve enforcement (SMART Partnership, 2017). Consequently, evaluation should focus on whether and how adaptive management can be achieved.

Supporting evidence provided medium confidence that SMART can reliably deter illegal activity. Our understanding of the relationship between patrol presence and illegal activity in protected areas is generally poor, suggesting that this step in SMART's theory of change is a high priority for evaluation. Future research should focus on establishing whether deterrence operates in practice in relevant contexts and levels of patrol activity necessary to have a real-world effect on illegal natural resource use. The process also suggested factors likely to maximise successful outcomes, which can be used to inform future implementation, such as (1) providing resources, capacity and incentives necessary for rangers to take appropriate action, and (2) focusing deployment on areas with strong judiciary and governance or combining deployment with effective activities to increase successful prosecutions and enhance legitimacy. Finally, the review provided low confidence that ranger-based monitoring tools, such as SMART, can provide reliable information on threats in protected areas, especially if patrols provide inconsistent or patchy coverage. Consequently, research should focus on assessing levels of coverage provided by patrols in SMART sites and whether these levels are sufficient to provide reliable data.

It is possible that my application of the theory of change approach may not be an accurate representation of how SMART is assumed to act, or that my review of existing evidence was biased. I collaborated with SMART's developers, with individuals drawn from across the range of organisations partnering on the tool and with experience of implementation in protected areas around the world. Yet these individuals were generally senior representatives of those organisations and not based within sites. A more representative theory of change might have been produced which included the views of those actively working to implement the tool at site-level. Similarly, whilst I reviewed a broad range of literature, the review was neither systematic nor exhaustive. Future applications should consider employing systematic protocols for reducing biases and uncertainties when reviewing literature (e.g., Pullin & Stewart (2006)).

In conclusion, a theory of change approach has produced understanding necessary to frame and interpret rigorous, formative evaluation of SMART. By mapping the 'missing middle' between SMART's activities and its intended impacts (Biggs *et al.*, 2017) I have also identified gaps in understanding and weaknesses in design which can be used to guide evaluation activities. Unsustainable hunting is the most commonly reported threat to biodiversity in terrestrial protected areas globally (Schulze *et al.*, 2018). The knowledge established here is essential to develop effective tools for combatting this threat. Finally, the issues identified are common to many conservation interventions (e.g., variable treatments and subtle effects (Margoluis *et al.*, 2009a)), suggesting theory of change would benefit conservation evaluation in general.

Chapter 3

A framework for implementation evaluation of conservation interventions

Abstract

Rigorous evaluation of conservation interventions is essential, but rare, because the context poses specific challenges. One underappreciated challenge is heterogeneity among implemented interventions, which frequently differ: from programme designs, between participants, and with context. Variation in implementation can influence outcomes, so must be accounted for to ensure correct interpretation during impact evaluation, and understanding heterogeneity is essential to improve interventions and predict whether effects will be replicated. Consequently, evaluation of implementation is common in other policy arenas, but rare and usually informally applied in conservation. Here, drawing on guidance for complex medical interventions, I develop a novel framework for evaluating implementation of conservation actions, which considers three critical aspects: Activities, Inputs, and Moderators. I demonstrate the framework's utility by applying it to a popular intervention for monitoring and management of law enforcement in protected areas globally, which is lacking rigorous evaluation: SMART. I examine how faithfully SMART was implemented in practice (Activities), how implementation was achieved (Inputs), and whether the contexts in which it was implemented were conducive to success (Moderators). Results indicate that SMART was commonly implemented as a tool for monitoring, but less frequently and faithfully to inform management. Inputs supporting activities (e.g., technical support) were comparatively consistent, although training was often weak. Lastly, SMART was frequently implemented in supportive contexts (e.g., where management were committed to implementation), although enforcement levels were often inadequate. That implementation fails to achieve adaptive management consistently may limit the extent to which outcomes are achieved. Implementers should strengthen managers' capacity to use monitoring information (e.g., through enhanced training). Finally, outcome evaluation must account for heterogeneity in implementation, and should employ flexible, mixed-methods approaches (e.g., dose-response analysis). My results demonstrate that thorough analysis of implementation is essential for formative, rigorous impact evaluation in conservation. The framework developed here provides clear, accessible guidance in this regard.

3.1 Introduction

Empirical evaluation of conservation interventions is essential – to help practitioners and policy-makers identify which approaches represent a wise investment and to improve those that are failing (Baylis *et al.*, 2016). In part, this requires application of rigorous methods for inferring effectiveness – the degree to which interventions contributed to conservation outcomes – such as quasi-experimental designs for estimating the counterfactual, or what would have happened in the absence of an intervention (Ferraro, 2009; Bottrill and Pressey, 2012; Jones and Lewis, 2015). The conservation community has been slow to adopt such methods, because the context poses specific challenges (Fisher *et al.*, 2014). For example, relevant outcomes data are rarely available or expensive to obtain, time lags between intervention and impact are long, and causal relationships are poorly understood (Margoluis *et al.*, 2009a). Conservation also usually acts on units several steps removed from biodiversity, such as the behaviour of individuals or governments, generating subtle effects which are difficult to quantify and even harder to attribute to interventions (Howe and Milner-Gulland, 2012).

Whilst rigorous methods for inferring effects are essential, another significant but underappreciated challenge is heterogeneity in treatments: variation in how and in what contexts interventions are implemented in practice (Ferraro and Pressey, 2015). Evaluation methods often treat interventions as a consistent treatment variable (e.g., protected vs. non-protected areas), or as a factor that varies systematically between participants. In practice, conservation interventions are rarely homogenous (Pfaff *et al.*, 2015). Ostensibly similar interventions can vary broadly in objective (e.g., protected areas represent a range of management approaches (Dudley, 2008)). More problematically, interventions implemented in practice frequently differ from the designs of programme developers (Durlak and DuPre, 2008). Developers may plan for an intervention to be applied consistently, but delivered interventions often vary from these intentions. Complex interventions, such as those found in conservation, typically undergo modifications and tailoring to suit real-world conditions. (Craig *et al.*, 2008). For example, at one extreme, protected areas may be 'paper parks', which provide protection only in name (Barnes *et al.*, 2018). Importantly, these differences in implementation may vary between participants, and through time (e.g., teething problems) (Moore *et al.*, 2015). Furthermore, interventions often take place across widely different contexts, which shape implementation and moderate the mechanisms through which they effect change (Marchal *et al.*, 2013).

Heterogeneity in implementation can have serious implications for outcomes, which must be accounted for during evaluation. Calls for application of more rigorous methods tend to focus on effects (e.g., Ferraro and Pattanayak (2006)) which, whilst crucial, are only part of the cause-effect equation. Critically, variation in how interventions are implemented can have implications for the extent to which outcomes are achieved (Steckler and Linnan, 2002). For example, the capacity of and resources available for management predict positive species population trends across protected areas (Geldmann *et al.*, 2018). Consequently, understanding heterogeneity in implementation is essential for correct interpretation of outcomes. For example, poor outcomes may reflect implementation failure rather than weaknesses in design or underlying theory (Montgomery *et al.*, 2013). Similarly, by altering interventions' mechanisms of impact, contextual factors external to implementation can moderate outcomes. For example, protected areas in Costa Rica were more likely to result in positive outcomes if tourism was also present (Ferraro and Hanauer, 2014). Understanding how implementation varies with context is thus also essential to predict whether similar effects can be replicated if an intervention is repeated with new participants (Moore *et al.*, 2015).

Consequently, examining the implementation, functioning and setting of an intervention is a common practice in policy arenas other than conservation, and is variously called implementation science, implementation evaluation or process evaluation (Humphrey *et al.*, 2016) (henceforth, implementation evaluation). In contrast with outcome evaluations, implementation evaluations are generally formative, with the aim of strengthening and refining interventions (Rossi *et al.*, 2004). In combination with outcomes, implementation evaluations can provide insights into how and why an intervention resulted in an effect, or why implementation failed, which can be used to inform improvements (Montgomery *et al.*, 2013). For example, process evaluation of conservation education in Belize and Costa Rica enabled implementers to improve their activities (Jacobson, 1991). Implementation evaluation is customary where practitioners are trying to effect difficult behaviour change in highly variable conditions (e.g., in education (Humphrey *et al.*, 2016), public health (Escoffery *et al.*, 2016), and behavioural medical interventions (Oakley *et al.*, 2006)). For example, in public health, the risk of unduly dismissing interventions which have limited effects due to implementation failure has long been recognised (Basch *et al.*, 1985). However, despite operating under similar constraints, implementation evaluation is rarely explicitly reported in the conservation literature. Whilst analyses of the effects of factors internal and external to interventions on outcomes are frequently reported, and considerations of context are commonplace (e.g., in the political ecology literature), there are few formal, accessible reported methods for evaluating implementation, hindering rigorous impact evaluation.

SMART, a popular intervention which aims to improve the effectiveness of protected area patrolling, and which is currently implemented in >600 protected areas across >55 countries around the world (SMART Partnership, 2018), is a case in point. Protected areas globally are failing to stem illegal killing threatening biodiversity, leading to declines in species populations across the tropics (Craigie *et al.*, 2010; Gray *et al.*, 2017). SMART aims to reduce illegal killing of wildlife by improving monitoring and management of law enforcement (Stokes, 2010). However, despite the criticality of these objectives and substantial resources invested in development and deployment of SMART, there has been no rigorous, empirical evaluation of its implementation or impact. Positive outcomes have been reported in sites in which SMART has been implemented (Hötte *et al.*, 2016) but without causal attribution. One challenge to attribution is that SMART implementation may be strongly heterogeneous: implementation is a complex of inputs and activities which are not strictly regimented and may vary according to the needs and capacity of participating sites, and multiple factors external to implementation will influence success (chapter 2). For example, some of the causal chains through which SMART is expected to work assume that managers use data generated through ranger-based monitoring to inform adaptive enforcement management activities (e.g., patrol planning). However, in conservation settings, adaptive management is rarely achieved in practice (Nichols and Williams, 2006). Consequently, outcomes may vary because of implementation heterogeneity rather than weaknesses in design. Understanding how implementation varies from that intended by developers and between sites is essential for correct interpretation of SMART's outcomes, but has not been empirically evaluated, exacerbated by the lack of clear, accessible methods.

Here, I draw on guidance for complex interventions in other policy arenas to develop a novel framework for evaluating implementation of conservation interventions, structured around three critical aspects: Activities, Inputs, and Moderators. I then demonstrate the framework's application and utility for informing rigorous, formative impact evaluation, by using it to evaluate implementation of SMART. The analysis is conducted with close reference to SMART's theory of change, which was developed explicitly to guide evaluation activities (Fig. 2.1). Specifically, I examine how faithfully the intervention was implemented in practice (Activities), how implementation was achieved (Inputs), and whether the contexts in which it was implemented were conducive to success (Moderators). In so doing, I demonstrate that rigorous evaluation of implementation is essential, and that the framework can facilitate that process. I also use results to suggest ways in which SMART can be implemented to maximise the chances of success and to frame and inform outcome evaluation, to improve the effectiveness of enforcement in protected areas globally.

3.2 Methods

3.2.1 Framework for implementation evaluation of a conservation intervention

I developed a framework for evaluating implementation of conservation interventions by drawing on guidance for process evaluation of complex behavioural medical and education interventions (Moore *et al.*, 2015; Humphrey *et al.*, 2016). Complex behavioural interventions are typically targeted at multiple groups or organisational levels and involve numerous programme components interacting to produce difficult behaviour change (Craig *et al.*, 2008). These characteristics are common to conservation in general (Margoluis *et al.*, 2009a) and implementation of SMART specifically (chapter 2). Complex interventions also often exhibit a high degree of variation in implementation (Craig *et al.*, 2008). Implementation evaluations aim to understand this heterogeneity by examining aspects of (1) implementation and (2) context (Moore *et al.*, 2015). Typically, examination of these aspects is framed by a clear description of the intervention, and prioritised through appreciation of the intervention's underlying causal assumptions (Craig *et al.*, 2008).

Implementation (1) is typically concerned with two factors: (a) the processes and resources by which an intervention is put into practice (*how* delivery of the intervention is achieved) (Lendrum and Humphrey, 2012), and (b) the quantity and quality of *what* is actually delivered. The latter (b) primarily relates to issues of implementation 'fidelity', or the extent to which the sequence and structure of intervention activities formulated by programme developers are implemented as planned (Breitenstein *et al.*, 2010). Deviations from these plans will have implications for whether and how outcomes are achieved (Steckler and Linnan, 2002). Additional factors may also be considered, such as the amount of an intervention to which participants are exposed ('dosage'), changes made during implementation ('adaptations'), the extent to which an intervention reached its intended audience ('reach'), and participant responses to an intervention ('mechanisms of impact'). Context (2) concerns pre-existing factors external to the intervention which may influence implementation or moderate mechanisms by which activities achieve intended outcomes, such as implementers' and participants' capacity and motivation for change, resource availability, and support for implementation (Humphrey *et al.*, 2016).

Drawing on these methods, I developed a novel framework for conservation implementation evaluation (Table 3.1), which considers three critical aspects: (1) Activities (What was delivered in practice?), (2) Inputs (How was delivery achieved?) and (3) Moderators (In what contexts?). To facilitate framing and interpretation of outcomes, (1) and (2) closely align with programme components commonly identified in conceptual models, such as logic models, results chains and theories of change, which are frequently used to guide conservation impact evaluation (Margoluis *et al.*, 2009b). Activities are those actions carried out by programme actors to achieve some defined conservation end (e.g., protected area managers deploy ranger patrols to detect and deter illegal activity). Inputs are resources invested by implementers to support those activities (e.g., in the previous example, rangers, enforcement training, boots or funding).

Table 3.1. Framework for evaluating implementation of conservation interventions, and SMART-specific evaluation questions and associated data collection methods.

Aspect of evaluation	Broad question addressed	Implementation characteristic/s	SMART-specific evaluation questions	SMART-specific method/s
Activities	What was delivered in practice?	Implementation fidelity, dosage, adaptations and reach, and participant responses to implementation	Which SMART monitoring and management activities were implemented in practice, to what extent were those activities implemented as intended, and how did fidelity vary amongst participants?	Semi-structured interviews with key informants
Inputs	How was delivery achieved?	Implementation process (e.g., resources, support structures, training, etc.)	What processes and resources supported SMART implementation (e.g., training, equipment, funding, staff), and how did inputs vary amongst participants?	Participant questionnaires
Moderators	In what contexts was the intervention delivered?	Contextual factors which influence implementation or moderate impact mechanisms	How do contextual factors that may affect success of SMART implementation vary amongst participants?	Semi-structured interviews with key informants; Participant questionnaires

The framework also considers the influence of external context (Moderators), focused on factors likely to contribute to success, to emphasise implementation evaluation's formative role. For example, protected area effectiveness is influenced by extent of political support (Watson *et al.*, 2014). As many conservation evaluations are undertaken ex post on interventions which were implemented across heterogeneous circumstances (Ferraro and Pattanayak, 2006) a strong understanding of confounding factors is essential.

3.2.2 Applying the framework to SMART

I applied the framework to SMART, with close reference to the intervention's theory of change (Fig. 2.1). I assessed implementation in a sample of participating protected areas, including Activities, Inputs and Moderators, and how these aspects varied in practice between subjects, drawing on interviews with key informants and a survey of site managers (Table 3.1). To summarise results, I also characterised strength of implementation, by aspect and site, with respect to the intentions of programme developers.

3.2.2.1 Site selection

The study sample comprised 27 diverse protected areas in 15 countries across Asia, Africa and Latin America. The 27 sites were drawn from a list of 134 sites globally in which SMART had been implemented at that time (November 2015). Random sampling was not possible as, for practical or political reasons, a proportion of sites would be unable to participate (e.g., lack of resources available to commit to the study or law enforcement too politically sensitive). Consequently, a subset of 41 sites was identified which were likely to be able to engage, in conjunction with SMART's developers. Partner NGOs supporting implementation at each site nominated a key informant who had direct knowledge of site-level implementation, but who was independent of management (to reduce bias in responses). Generally, this meant technical advisors employed by partners to support site-level implementation and patrol operations. One informant was identified per site, although a few individuals were responsible for more than one site. 20 informants, representing 27 sites, responded to requests to participate. The 27 sites varied in factors other than location, including: area (mean=2,949 km^2 ± 4,276 SD), IUCN protected area management category (mostly National Parks (category II), with some Protected landscapes (V), Habitat/species management areas (IV) and Protected areas with sustainable use (VI), and one Strict nature reserve (Ia)); governance type (majority state managed, and a few under shared state/private governance); and primary threat (majority commercial or subsistence poaching, but encroachment, logging, human-wildlife conflict and mining also reported). To encourage participation, the sites' identities were anonymised by removing identifying features (e.g., name, location and area) from all outputs and assigning randomly-generated number IDs.

3.2.2.2 Aspects of implementation

I evaluated implementation of factors broadly corresponding to programme components identified in SMART's theory of change (Fig. 2.1). Specifically, to answer what was delivered in practice (Activities), I assessed fidelity of implementation of five broad categories of activity, with respect to the intentions of the intervention's developers. I captured which activities had been implemented, and to what extent those activities were implemented as intended, by exploring characteristics considered essential for effective implementation. Monitoring activities included: (1) ranger-based monitoring of wildlife, threats and patrol activity using SMART protocols, and subsequent entry and storage of data in a SMART database; and (2) mapping, analysis and reporting of data facilitated by SMART software. Management activities, informed by SMART reports, included: (3) strategic planning of patrols; (4) meetings with frontline staff to feedback results; and (5) ranger or team performance evaluations and associated incentive mechanisms. Within activities, characteristics included: (1) specific activities taking place (e.g., whether individual ranger and/or team performance was evaluated using SMART information); (2) timeliness (e.g., how frequently reports were produced); (3) participants (e.g., staff responsible for activity, or attendees at feedback meetings); and (4) monitoring information involved (e.g., whether reports included maps of illegal activity, or whether such maps informed patrol planning). To answer how delivery was achieved I measured inputs supporting SMART activities in practice, including: (1) funding; (2) technical support from a partner NGO; (3) staff; (4) training; and (5) equipment. Finally, to answer in what contexts the intervention was delivered, I assessed factors external to implementation that may influence success, as identified in the theory of change, including: (1) resources available for enforcement; (2) institutional stability; and (3) whether implementation was supported by the government and/or management authority (whichever was more relevant).

3.2.2.3 Data gathering

Methods used varied by aspect of implementation evaluation (Table 3.1): to capture Activities, I conducted semi-structured interviews with the 20 key informants; to capture Inputs, data questionnaires were distributed to managers of the 27 sites, for completion under the supervision of informants; and Moderators were captured using both methods. Interviews followed a pre-defined structure of closed-ended questions, administered verbally, providing flexibility to ask open-ended follow-up or supplemental questions, as needed (Newing *et al.*, 2011). I piloted the approach with an individual holding the same position as interviewees (i.e., a site-based SMART technical advisor), who was not part of the study sample, and refined questions, as necessary. Interviews lasted for approx. 1-2 hours, and were conducted remotely over voice-only Skype, with one exception conducted in person. Closed-end responses were noted as they were given, but interviews were also audio recorded for later review. Ideally, multiple informants would be interviewed for each site to triangulate responses and reduce bias associated with interviews (Bernard, 2011). In this instance, only one informant per site was practicable as implementation is generally associated with individual technical advisors. Thus, to reduce errors, I employed interview tactics as recommended by Newing *et al.* (2011). For example, to reduce non-directional errors (e.g., lack of knowledge) I explained that 'I don't know' was an acceptable answer, and to reduce directional biases I emphasised that results would be anonymous and remained neutral about responses. Non-numerical responses to closed-ended questions were coded prior to interviews and questionnaires with numerical values, to enable quantitative descriptive analysis of all closed-ended results by aspect (e.g., proportion of sites producing reports, proportion of sites in which activities were supported by full-time support staff, and proportion of sites in which management were committed to implementation).

3.2.2.4 Characterising strength of implementation

I characterised the strength of implementation of each aspect, drawing on discussions with developers and SMART training literature. How I characterised strength varied by aspect.

<u>Activities evaluation</u>: For Activities, I assessed how many of the five monitoring and management activities were implemented and the fidelity of their implementation. To construct an index that summarised fidelity of implementation, I aggregated response values. Responses were coded on a scale from 0 to 1: 1 representing 'perfect' implementation for that variable, and 0 the obverse (e.g., 'Are reports produced using SMART?' had two possible responses: 'Yes' (1) or 'No' (0). Questions with >2 responses were coded on a scale from 0 to 1 in equal increments (e.g., 'How quickly are patrol data entered into SMART?' had five responses '<1 week', '1 week-1 month', '1-3 months', '3-6 months', and '6 months+', coded 1, 0.8, 0.6, 0.4, 0.2 and 0, respectively. Number of questions varied between activity and characteristics. Consequently, when constructing the index, response values were weighted according to number of questions, to grant each characteristic equal weight within activities, and each activity equal weight within the index. Final index scores were calculated as a proportion of maximum possible scores. I categorised 'strong' implementation as implementation of all five activities with fidelity ≥ the sample's second quartile; 'moderate' as at least four activities with fidelity ≥ the first quartile but < the second quartile; and 'weak' as less than four activities or fidelity < the first quartile.

Inputs evaluation: To assess strength of implementation of Inputs, I considered what proportion of inputs were adequate. I classified adequacy for staff, technical support, training and equipment, as: at least one full-time support staff, at least quarterly on-site technical support, at least three activities trained, and at least an on-site computer, respectively. I categorised 'strong' implementation as ≥75% of inputs were adequate, 'moderate' as ≥50%, and 'weak' as <50%.

Moderators evaluation: Finally, to assess strength of Moderators, I considered what proportion of contexts were conducive to success, calculated as the number of affirmative responses to context questions as a percentage of all context questions. I categorised 'strong' implementation as ≥75% of contexts were conducive, 'moderate' or 'neutral' as ≥50%, and 'weak' as <50%. To limit excessive interview duration the number of contexts questions asked varied by site. Only sites where ≥50% context questions were asked were included in the strength analysis. Open-ended responses to follow-up questions were also reported using quotes if illustrative of or opposed to numerical results.

3.3 Results

3.3.1 Activities evaluation: What was delivered in practice?

I found that SMART was commonly implemented as intended as a tool for monitoring, but less frequently and less faithfully to inform management (Fig. 3.1). All sites implemented monitoring activities (i.e., data collection & entry, and analysis & reporting), but whether sites subsequently made use of monitoring information to inform implementation of management activities (as intended by developers) was more variable. All but one site used results derived from SMART to inform patrol planning and most used results to inform staff performance evaluations (81% of sites) or fed results back to frontline staff via meetings (81%). However, fewer sites undertook all management activities (70%). Additionally, whilst the majority of sites implemented all activities, the fidelity of individual activities varied by activity and between participants.

Monitoring activities were generally implemented across all sites with strong fidelity. In all but one site, ≥95% of patrols collected monitoring data, which was generally entered into a database within one month of collection (78%) by dedicated data entry staff (70%) who were based on-site (85%). All sites also produced reports, which in all but one site were standardised (i.e., to a template), by dedicated member/s of staff (81%) who were based on site (74%). All but one site produced reports at least quarterly (96%) and sometimes at least monthly (52%), predominantly to a schedule (89%). Reports in all sites were sent to site managers, generally sent to section heads/patrol leaders (81%), but less commonly to rangers (52%). Types of information contained in reports also varied. Patrol effort, team performance, patrol routes, and maps of illegal activity or wildlife were commonly reported (≥85%), while information such as individual ranger performance, patrol activities (e.g., seizures, arrests), patrol plans and problems encountered were less common (56-70%), and intelligence, recommendations for follow-up action, and temporal trends in patrol performance, threats and wildlife, were rare (26-44%).

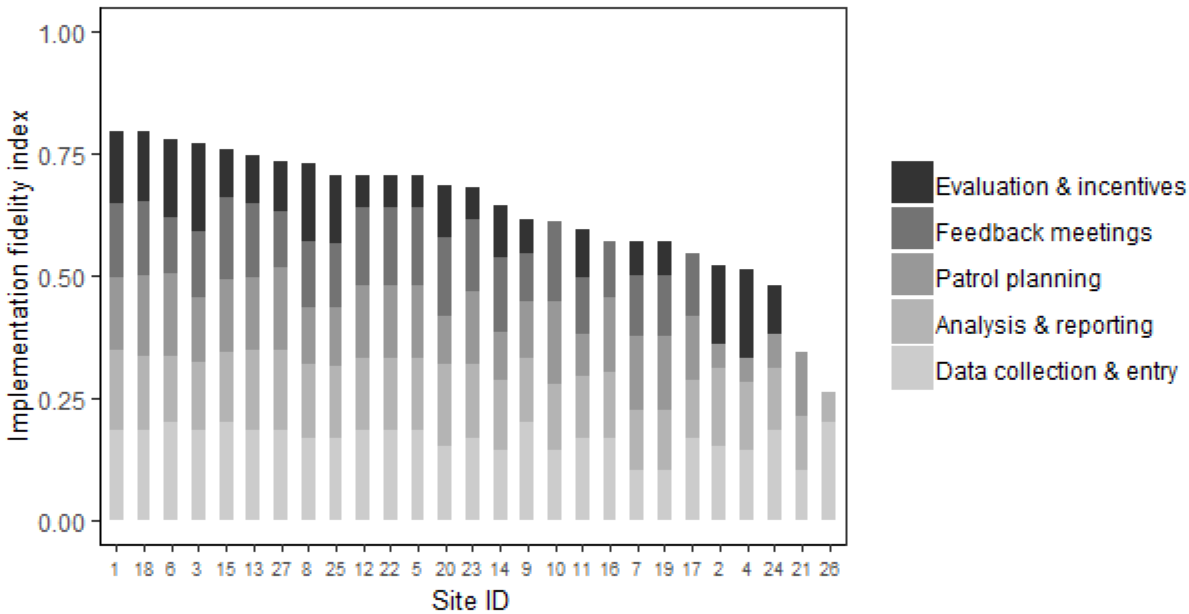


Figure 3.1. Which SMART activities were implemented and extent to which activities were implemented as intended (fidelity) across sites, ranked by index score. Presence of shaded bars indicates whether activities were implemented, and length of bar indicates fidelity. SMART was commonly implemented as intended as a tool for monitoring (Data collection & entry, and Analysis & reporting), but less frequently and less faithfully to inform management (Patrol planning, Feedback meetings, Evaluation & incentives).

Management activities were implemented less faithfully and more variably between sites. Patrol planning was ubiquitous, but fewer participants (67%) set specific targets for the subsequent reporting period. The activity was conducted at least quarterly, and often at least monthly (67%). Results used in patrol planning tended towards patrol effort and patrol routes (81% and 89%, respectively), but use of maps of illegal activity or wildlife was rarer (56% and 44%, respectively). Similarly, formal meetings were held in the majority of sites, and meetings were held on-site (78%), generally to a schedule (70%), and involved patrol team leaders (78%) who were generally invited to comment on results (67%). However, relatively few (52%) involved rangers in those meetings, and fewer still invited rangers to comment

(37%). The activity was conducted at least quarterly, but less commonly at least monthly (52%). Finally, whilst performance evaluation was implemented in the majority of sites, at the team or individual ranger level (67% and 59%, respectively), relatively few tied evaluations to formal incentives (37%) (mainly to bonuses (30%), occasionally to salaries (15%) and rarely to promotions (4%)). Moreover, only half of sites employing incentives consulted rangers about the mechanism prior to implementation (19% of all sites). Failure to implement management activities with consistent fidelity was reflected in open-ended responses:

"The main challenge is for mid-management level to take responsibility for the full cycle of implementation."

"The feedback system has been problematic at points."

3.3.2 Inputs evaluation: How was delivery achieved?

Inputs supporting delivery of activities varied between sites and across input types (Fig. 3.2). In summary, most sites received good support in terms of technical assistance, staffing and equipment, but training was often weak. Most sites received permanent on-site technical support for implementation from a partner NGO (63%). However, a minority received technical support semi-annually (11%) or less frequently (15%). Similarly, whilst many sites employed at least one full-time member of staff to undertake data entry, analysis and reporting (78%), a few employed only part-time staff (17%), and one employed no support staff. In most cases, these staff were funded by a supporting NGO (59%), but sometimes by the state (26%) or through a joint NGO-state arrangement (15%). Equipment available for implementation also varied but was generally strong. All but 2 sites had on-site computers, either with (74%) or without internet access (19%). Rangers generally used paper forms to record observations in conjunction with a GPS device (70%), but a few sites used digital recording devices (11%) or were transitioning to digital (19%). Training was less supportive. Rangers and support staff received formal training in data collection, entry, analysis and reporting in all sites, but training of managers in processes for using results to inform patrol planning or performance evaluation and incentives was rare (33% and 19%, respectively). Moreover, whilst all rangers had received training, it was unclear if this was provided upon initial implementation or as part of an ongoing process, which was identified as important by respondents:

"Refresher training [in data collection] is definitely required at least once a year."

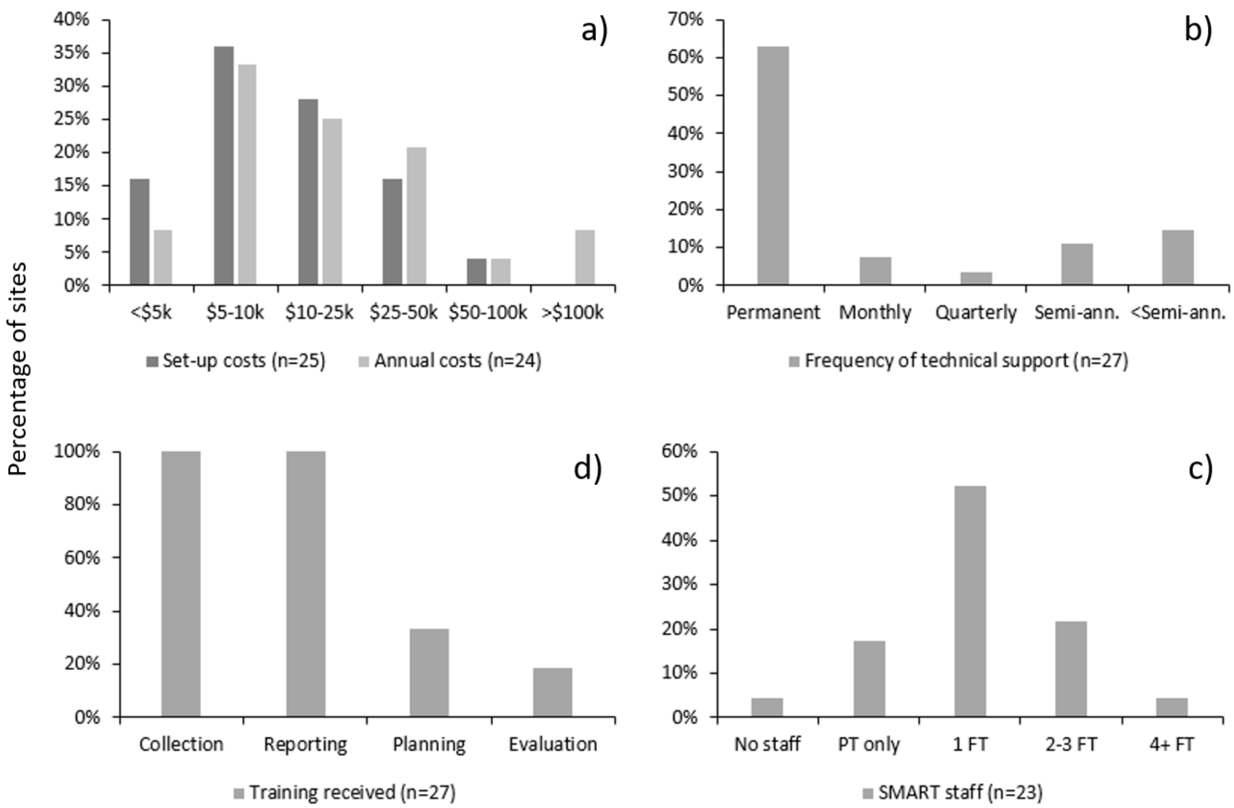


Figure 3.2. Inputs supporting delivery of SMART. Clockwise from top left: a) funding (initial and ongoing); b) technical support from a partner NGO; c) support staff (Part Time (PT) or Full Time (FT)); and d) training (in data collection & entry, analysis & reporting, patrol planning, and performance evaluation & incentives, respectively).

Funding supporting these inputs, in terms of initial set-up cost and per annum ongoing costs, fell in the $5,000 to $10,000 range for most sites (36% and 33%, respectively), but exceeded $50,000 in a few instances (4% and 12%, respectively). These costs were unrelated to either a site's area or the number of rangers it employed. Accordingly, respondents' perceptions of whether funding for implementation was sufficient also diverged:

> *"For sure [resources were sufficient], given that it was a priority site where we had some money for it."*

> *"We need more money to improve SMART."*

3.3.3 Moderators evaluation: In what contexts was the intervention delivered?

Finally, SMART was frequently implemented in contexts conducive to success, but resources for enforcement may have been inadequate in some circumstances (Fig. 3.3). In general, implementation was supported by both relevant local management authorities (80%) and site-level management (87%). However, the strength of leadership provided by site management was less constant, and open-ended responses indicated that management authority support may have been variable or only apparent once positive results had been demonstrated:

“[The management authority] are excited about the way SMART is working.”

“At first, [the management authority] didn't see the point in a new system.”

"It's very hard to convince them [the management authority] that this is a good tool to monitor their work [law enforcement]"

Resources available for non-SMART law enforcement activities were less frequently conducive to success. Only half of sites reported that resources were adequate, which was corroborated by open-ended responses, including shortages in staff, training and equipment:

“Staffing is the biggest problem . . . we don't have enough people to carry out patrolling.”

“We still need more rangers.”

“Rangers don’t have special enforcement training . . . they are just local people . . . they attend some training but before they are hired, they have no training."

“It would be better for antipoaching activities [if rangers were armed].”

However, in some cases, implementation may also have involved concurrent support for non-SMART activities:

“Because we implemented SMART [our technical partner] also gave us lots of other law enforcement support.”

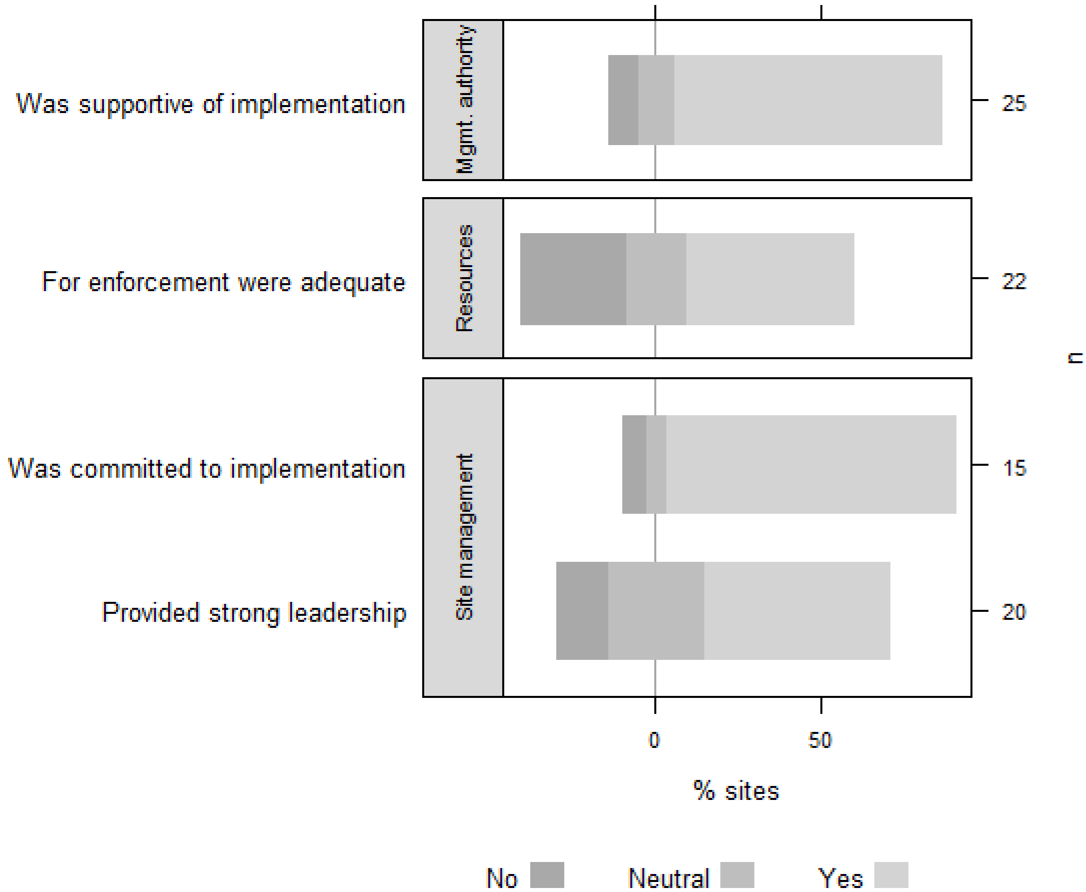


Figure 3.3. Respondents perceptions of contextual factors external to SMART implementation that may influence success.

3.3.4 Strength of implementation across Activities, Inputs and Moderators

Implementation of SMART varied by site and between the three aspects of implementation but was generally strong (Fig. 3.4). Implementation was most consistently strong in terms of the contexts in which SMART was delivered (Moderators), slightly more variable in terms of how delivery was achieved (Inputs), and most variable in terms of what was delivered in practice (Activities). All Activities were implemented with strong fidelity in half of sites (52%), but for a similar number one or more activities were not implemented, or activities were implemented with moderate or weak fidelity (26% and 22%, respectively). For the majority of sites (67%), delivery of these activities was supported by most or all inputs, but in a third of cases inputs provided only moderate or weak support (26% and 7%, respectively). Lastly, in a high proportion of sites (74%), SMART was implemented in contexts which were generally conducive to success, but for a few sites contextual factors were mixed (11%) or, in one case, unfavourable.

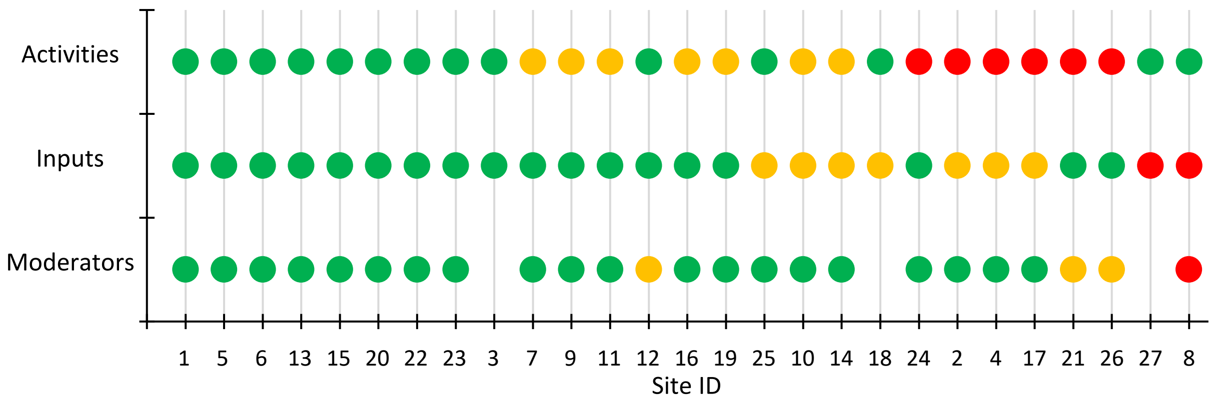


Figure 3.4. Strength of SMART implementation by site and aspect of implementation, ranked by overall strength of implementation and site ID. Green = strong implementation, amber = moderate/neutral, red = weak, blank = insufficient data.

3.4 Discussion

Rigorous evaluation of conservation is hindered by a paucity of clear methods for understanding heterogeneity in treatments and contexts. Here, to address this shortcoming, I developed a novel framework for evaluating implementation of conservation interventions, drawing on guidance from other policy arenas. I demonstrated the framework's application and utility by using it to guide implementation evaluation of a popular intervention for monitoring and management of law enforcement in protected areas globally, SMART. Evaluation of the effect of SMART on outcomes is essential but is limited by poor understanding of how implementation varies in practice (chapter 2). Application of the framework revealed that SMART implementation was generally strong but heterogenous, varying from the intentions of developers, and between participants and aspects of implementation. The framework helped to explain this variation, by identifying important aspects of implementation and providing research questions to co-ordinate and prioritise research effort. The heterogeneity in SMART implementation observed is consistent with expectations (chapter 2), and our understanding of variation in implementation of complex interventions in general (Craig *et al.*, 2008). This finding is probably also common amongst conservation interventions specifically, although direct comparisons are difficult as implementation is rarely explicitly evaluated in the conservation literature, even though poor appreciation of heterogeneity is recognised as an inhibitor to rigorous evaluation (Ferraro and Pressey, 2015).

Implementation was most variable in what was delivered (Activities), more consistent in how delivery was achieved (Inputs) and most consistent in terms of contexts which were conducive to success (Moderators). That SMART was commonly implemented as a tool for monitoring, but less frequently and faithfully to inform management is consistent with natural resource management in general. Use of information to inform adaptive management is often stated as the primary goal of biodiversity monitoring, but is difficult to achieve in complex systems (Nichols and Williams, 2006; Game *et al.*, 2014). Few sites received formal training in management activities, which may have influenced participants capacity to implement these activities faithfully. It is also possible that achieving faithful implementation of activities takes time, and more complex management activities which rely on monitoring data would naturally be implemented later in this process. That inputs supporting activities other than training were comparatively consistent, and that SMART was frequently implemented in contexts conducive to success, could be explained by the fact that, in general, site-level implementation of SMART is supported and initiated by a partner NGO (chapter 2). Consequently, partner NGOs are likely to preferentially select sites where contextual factors are supportive and to ensure adequate inputs are in place. Nevertheless, respondents in only half of sites reported that resources available for law enforcement were sufficient for implementation, which is a common challenge in protected areas (Nolte, 2016).

Variation in implementation will have implications for whether outcomes are achieved. Significantly, results suggest that outcomes may vary because of how SMART was implemented in practice, particularly regarding aspects of activities and context, rather than weaknesses in design. In theory, SMART acts to reduce illegal activity through multiple causal pathways (Fig. 2.1). One pathway predicts that SMART can improve ranger motivation through monitoring activities alone, by providing rangers with increased job satisfaction and the perception of increased oversight. However, these perceptions will be short-lived if oversight is not acted upon, and SMART is primarily thought to work by using monitoring information to inform recurring management mechanisms (e.g., performance evaluation and patrol planning). Implementing such mechanisms, which link information captured by monitoring to decision-making, is a crucial step in achieving successful adaptive management (Keith *et al.*, 2011). Consequently, whether management activities are being implemented faithfully, will influence the extent to which outcomes (i.e., increased ranger and patrol performance) are achieved. The mechanisms through which interventions operate are also moderated by context (Marchal *et al.*, 2013). This may be a particular issue for SMART sites in which resources for enforcement were inadequate. SMART may successfully improve ranger and patrol performance, but if there are too few trained, equipped rangers to provide broad and consistent patrol coverage then information generated by rangers on trends in illegal activity and wildlife may be misleading (Keane *et al.*, 2011), and illegal activity may not be deterred (Plumptre *et al.*, 2014).

As SMART outcomes may vary because of implementation or design, evaluation must take heterogeneity in treatments and moderators into account (Steckler and Linnan, 2002). The results produced here can be used to interpret why outcomes vary. However, the diversity and extent of heterogeneity observed suggests methods for inferring effects which have restrictive requirements in terms of treated and control groups and which are commonly advocated in the literature (e.g., quasi-experimental matching methods (Ferraro, 2009; Jones and Lewis, 2015)), will be unsuitable. Consequently, evaluation should adopt flexible, mixed-methods designs, incorporating approaches such as nonexperimental or qualitative sampling, process tracing, contribution analysis and dose-response analysis, whilst appreciating that such methods have less power to detect causal relationships (Margoluis *et al.*, 2009a; Rogers, 2014; Befani and Stedman-Bryce, 2017). Strong heterogeneity in treatments and moderators may limit the extent to which we can reliably predict relationships between interventions and impact, for SMART specifically and conservation in general. However, evaluation methods which have a stronger focus on process and less on results can play an important formative role (Rossi *et al.*, 2004). Accordingly, the results of this evaluation suggest ways in which SMART implementation might be improved to maximise success. Implementers should strengthen managers' capacity to faithfully implement management activities (e.g., through enhanced training), and direct resources to sites with high overall enforcement levels or provide concurrent support to increase enforcement.

Limitations in the methods used to assess implementation of SMART may constrain the generalizability of findings and their utility for interpreting outcomes and improving the intervention, in terms of the study sample and the methods used to gather data and characterise results. Firstly, the sample only comprised sites which were willing and able to share sensitive enforcement information and devote time to the study. Consequently, results may not be representative of all sites implementing SMART. Obtaining sensitive information from sites outside of this sample may be an intractable problem. A comparison of results in sites where information was easily obtained vs. challenging to obtain might elucidate whether there is a difference and its direction. Secondly, all results drew on the perceptions and self-reported behaviour of key informants and site managers, and only involved individual informants for each site. The risk of directional and non-directional biases in such data is high (Bernard, 2011). Future research should aim to elicit a variety of perspectives for each site (e.g., including managers, rangers and other stakeholders) and triangulate responses to reduce bias (Newing *et al.*, 2011). Thirdly, I treated all sites equally in terms of inputs. However, there may be good reasons why some inputs varied between sites (e.g., sites with more patrol staff may generate more monitoring data and so need more support staff to handle entry and analysis). For the purposes of characterization of strength of inputs this should have a limited effect on results as I compared inputs against minimum adequate levels (e.g., at least one full-time support staff), rather than comparisons between sites. Outcome evaluation which attempts to account for heterogeneity in inputs should also account for varying need (e.g., use number of support staff per ranger as a comparison metric).

Nevertheless, the framework developed here provides clear, systematic guidance for empirical evaluation of implementation of conservation interventions. Application of the framework to SMART revealed that implementation varied between sites and from programme designs. The results of this application can be used to evaluate and improve the intervention, to enable more effective enforcement in protected areas globally. These results also demonstrate that rigorous analysis of implementation is essential. Evaluation in conservation tends to focus on short-term outputs, which may have weak relationships with impact (e.g., protected area coverage (Pressey *et al.*, 2015) or number of sites in which SMART has been implemented). Moving beyond outputs to inferring effects on outcomes is crucial (Baylis *et al.*, 2016). A stronger understanding of heterogeneity in treatments and moderators will be central to achieving this goal.

Chapter 4

State and trends in spatiotemporal presence of ranger-led enforcement patrols in protected areas globally, an analysis using SMART monitoring data

Abstract

Illegal use of natural resources is a major threat to endangered species in protected areas globally, which is countered, primarily, by ranger-led law enforcement patrols. Rangers travel around parks, aiming to find rule-breakers, remove passive hunting devices, and deter illegal activity. Typically, managers attempt to maximise detection and deterrence of illegal activity by maximising presence of patrols in space and time. Yet, despite concerns that presence may be generally low and potentially inadequate, understanding of spatiotemporal patrol activity is poor. Recent, widespread deployment of a tool for standardised patrol monitoring (SMART) provides an unprecedented opportunity to estimate patrol activity at broad scales. Here, I take advantage of SMART's ubiquity to provide the first global analysis of spatiotemporal patrol presence in terrestrial protected areas. Using data assembled from 21 diverse sites across Africa and Asia, I estimate spatial and temporal coverage provided by ranger patrols within and across sites, estimate trends over time in coverage following SMART implementation, and evaluate coverage with respect to industry benchmarks. I also assess factors influencing coverage, focussing on whether differences between sites can be explained by resources available for patrolling, and whether SMART management mechanisms designed to improve performance influenced positive change over time. Results indicate that patrols typically provided very low spatial coverage of protected areas at monthly scales, and low coverage at annual scales, with large proportions of many sites rarely patrolled and the majority of sites falling short of industry targets. Results also indicate that shortfalls in funding may be an important constraint to greater spatial coverage. In general, coverage improved following SMART implementation, but to what extent SMART contributed to change is unclear. My findings suggest that levels of patrolling globally may be insufficient to monitor and reduce illegal activity, and increased financial support is urgently required to protect threatened species.

4.1 Introduction

Illegal use of natural resources is a major threat to endangered species in protected areas globally. Illegal logging, fishing and hunting, for example, are common throughout protected areas in the Brazilian Amazon (Kauano *et al.*, 2017). Poaching, in particular, has driven widescale defaunation in reserves across the tropics (Harrison, 2011). Law enforcement – monitoring and punishment of crime – is essential to address this threat (Keane *et al.*, 2008). The importance of enforcement for discouraging illegal activity and encouraging compliance with laws is supported by theoretical models (e.g., Rowcliffe et al., (2004)) and empirical, correlative analyses. For example, the effectiveness of tropical parks correlates with enforcement aspects of management (Bruner *et al.*, 2001), and presence of enforcement is the best predictor of great ape survival across African protected areas (Tranquilli *et al.*, 2012). Similarly, reduced investment in enforcement correlates with increases in crime and negative impacts on biodiversity (Jachmann and Billiouw, 1997; Hilborn *et al.*, 2006; de Merode *et al.*, 2007). However, whilst evidence suggests enforcement is important for effective protection, illegal killing is still common in sites where enforcement is present (Nolte, 2016). Determining why illegal activity persists in ostensibly protected areas is essential to address continued declines in global biodiversity.

Ranger-led patrols are the principle means by which protected area managers enforce laws at site-level (Henson *et al.*, 2016). Teams of rangers are deployed to patrol protected areas, either on foot or by vehicle. Rangers are assumed to reduce illegal activity via two mechanisms: detection of crime that has already occurred (e.g., rangers find and arrest rule-breakers or find and remove passive hunting devices); and deterrence of crime yet to occur (ranger presence discourages rule-breaking by increasing the perceived risk of arrest and punishment) (Nagin, 2013a). Protected areas are often large and inaccessible (Joppa *et al.*, 2008) and resources for enforcement are limited, so establishing a permanent ranger presence throughout at-risk areas is unattainable. Consequently, managers typically attempt to maximise detection and deterrence of illegal activity by deploying patrols to provide as much presence as possible in space and time (Plumptre *et al.*, 2014). For example, enforcement best practice guidelines suggest managers should aim to achieve a minimum of 50% temporal coverage and 75% spatial coverage of protected areas by ranger patrols per month (Singh *et al.*, 2015). However, whether levels of coverage provided by patrols achieve these targets or patrol presence is sufficient to reduce illegal activity is poorly understood, because monitoring of patrols was, until recently, largely inadequate.

Historically, monitoring of patrol activity in protected areas was relatively crude, as managers lacked resources, capacity and incentives necessary to implement more sophisticated monitoring systems. For example, until recently, patrol presence was simply recorded as total number of days for which rangers were on patrol in a site (e.g., Leader-Williams *et al.* (1990); Jachmann and Billiouw (1997)), providing no information on variation in activity within sites and over time during patrols. Consequently, understanding of spatiotemporal patrol activity in most sites is limited and levels of coverage provided by patrols are largely unknown. Yet information on the spatial and temporal presence of patrols is essential to evaluate whether and how patrols deter crime and to establish whether levels of patrolling are sufficient to effectively reduce illegal activity (Dobson *et al.*, 2018). Detailed data on patrol effort is also essential to implement management mechanisms to improve coverage and effectiveness (Stokes, 2010). Moreover, indirect evidence suggest levels of patrolling in many protected areas may be low. For example, patrols in the Greater Virunga Landscape in central Africa provided low spatial coverage of protected sites, and that area was concentrated close to patrol posts (Plumptre *et al.*, 2014). Outside of this context, patrol activity is poorly understood, but presence is likely to be inadequate as protected areas frequently receive limited financial and political support (Watson *et al.*, 2014).

The advent of cheap, accessible radio navigation-satellite services (e.g., GPS) and systems for managing geographic data (e.g., GIS) have provided opportunities for monitoring of patrols at fine scales. In recent years, Law Enforcement Monitoring (LEM) tools, such as MIST (Management Information SysTem) and SMART (Spatial Monitoring And Reporting Tool) have exploited these technologies to enable high-resolution tracking and mapping of patrol activity in space and time (Hötte *et al.*, 2016). LEM tools are designed to be relatively inexpensive to implement, as rangers are responsible for data gathering, and to facilitate standardised collection and analysis of data on patrol movements, illegal activity and wildlife (Stokes, 2010). LEM tools are also explicitly intended to generate information that can be used to inform management. SMART, for example, is intended to be implemented within an adaptive management framework, which uses monitoring data to inform mechanisms for increasing patrol performance (SMART Partnership, 2017). For example, SMART data can be used to assess historical coverage and strategically deploy patrols to fill gaps, or to motivate rangers to increase effort (e.g., distance walked or hours on patrol) through performance evaluation and by tying results to incentives, such as bonuses. Tools such as SMART provide a means to fill the historical gap in information on patrol activity. Consequently, the use of LEM tools in protected areas has grown rapidly. SMART, for example, has been implemented in >600 sites across >55 countries globally since launching in 2012 (SMART Partnership, 2018).

The widespread deployment of patrol monitoring using LEM tools provides an unprecedented opportunity to assess patrol presence at broad scales and to evaluate whether targets are being met. This information is a prerequisite to determine why illegal activity killing continues to threaten wildlife in ostensibly protected areas. Moreover, because monitoring is standardised, LEM data provide an opportunity for cross-site comparisons, which can be used to assess which factors influence patrol presence and to suggest ways in which performance might be improved. For example, comparing trends in patrol coverage across sites in which implementation of performance management mechanisms varies (e.g., via SMART), could provide insights into whether such interventions are effective (chapter 3). Here, I take advantage of SMART's ubiquity to provide the first global analysis of state and trends in spatiotemporal patrol presence in terrestrial protected areas around the world. Using patrol data assembled from 21 diverse sites situated in 13 countries across Africa and Asia, I estimate spatial and temporal patrol coverage within and across sites, estimate trends in coverage over time following implementation of SMART, and evaluate coverage with respect to common industry benchmarks. I also assess factors influencing variation in coverage, focusing on (1) whether differences between sites can be explained by resources available for patrolling, and (2) whether implementation of SMART management mechanisms influenced improvements in coverage over time, employing a 'dose-response approach for confirming causal relationships (Rogers, 2014).

4.2 Methods

4.2.1 Study sites and patrol data

I assembled patrol data from a sample of protected areas around the world in which ranger-led law enforcement patrols were deployed to detect and deter illegal activity and in which SMART was used to monitor and manage patrolling. Potential study sites were selected from a list, maintained by SMART's developers, of all 134 protected areas globally in which the tool had been implemented at that time. I targeted a subset of terrestrial sites for participation. I focused exclusively on sites in which ≥95% of patrols had been monitored for at least one year and where monitoring data had been entered into a SMART database, and thus which could provide ≥12 months of representative patrol data. I also focused on sites which were likely to be able and willing to share politically-sensitive law enforcement data, drawing on conversations with NGOs supporting site-level patrol monitoring. I excluded relatively small sites by area (i.e., those <25 km^2), where measures of spatial coverage were less relevant, and sites which did not meet the broad definition of a protected area (Dudley, 2008). To encourage participation, the sites' identities were anonymised by removing identifying features (e.g., name, location and area) from all outputs and assigning randomly-generated number IDs.

21 sites agreed to participate and shared patrol monitoring data, including nine sites in seven countries across sub-Saharan Africa, and 12 sites in six countries across Asia (Fig. 4.1). The sites varied in factors other than geographic location, including: management objective (ten National Parks (IUCN PA management category II), two Protected Landscapes (V), two Protected areas with sustainable use of natural resources (VI), one Strict Nature Reserve (Ia), and six without a reported IUCN category); governance (majority state managed, with the remainder under shared governance with private organisations (e.g., NGOs)); primary threat (mostly commercial poaching, with subsistence poaching and encroachment (e.g., village expansion or clearance for agriculture) the next most common primary threats); area (mean=2,789 km^2 ± 2,301 SD); and ranger density (mean=0.035 rangers/km^2 ± 0.045 SD).

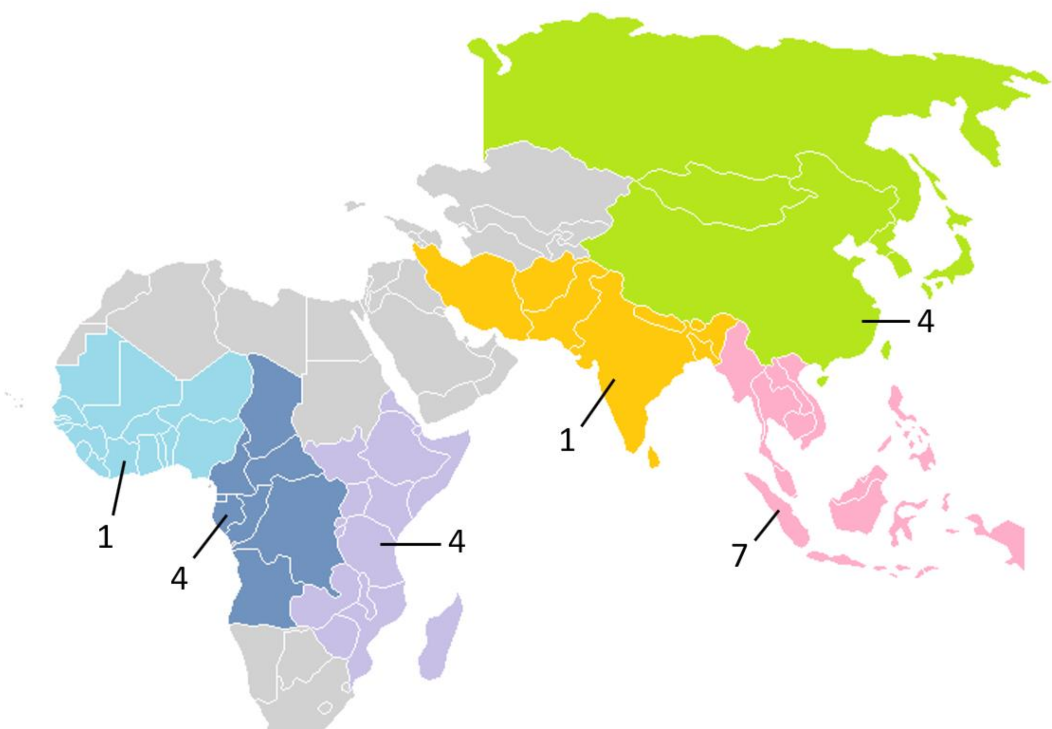


Figure 4.1. Number of study sites by subregion across Africa and Asia.

Rangers in SMART sites used handheld GPS devices to create time-stamped location records (waypoints) at the beginning and end of patrols, at intervals in-between (e.g., every 30 min, although frequency and regularity of waypoints can vary), and when they observe signs of illegal activity or wildlife. The combined dataset from the 21 study sites comprised >874,000 waypoints created during >53,000 patrols throughout >101,000 days of patrolling. All patrol data were collected between January 2012 and November 2017, although the start and duration of monitoring period varied by site (mean=3.79 years ± 0.97 SD), beginning with SMART implementation. The combined dataset is not included here, as the data are sensitive and would identify participating sites, so cannot be shared.

4.2.2 Analysis

Using patrol data, I calculated spatiotemporal patrol presence within and across sites, estimated trends in presence over time following SMART implementation, and evaluated results with respect to industry benchmarks. Finally, I assessed factors influencing differences in presence across sites and over time.

4.2.2.1 Estimating patrol presence within and across sites

I applied patrol presence measures which were commonly used by protected area managers for patrol monitoring and evaluation, and which were relatable to industry benchmarks (Singh *et al.*, 2015) (Table 4.1). Specifically, for each site, I estimated (1) spatial patrol coverage per month, measured as percentage of protected area visited by patrols per timestep (see calculation detail below), and (2) temporal coverage per month, measured as number of 24 h days for which one or more ranger teams were on patrol within sites, and calculated all measures for the entirety of each site's monitoring period. I reported mean coverage across all timesteps for each measure, and the mean of means for each measure across all sites.

I included a third measure – (3) spatial patrol coverage per year - for two reasons. Firstly, a metric of percentage area visited only indicates how much of a site is patrolled at each timestep, but not which areas. Consequently, stationary coverage over time can indicate that the same area was patrolled repeatedly, or that different areas of comparable size were patrolled. To circumvent this problem, I calculated both monthly and annual measures and compared outcomes. Sites with patrol activity concentrated in one location throughout the year would display little difference between outcomes. Conversely, sites in which activity location changes would display a difference in outcomes, proportional to the amount of variation in location. Secondly, the two timescales may also capture different management mechanisms. For example, over shorter timescales (e.g., one month), managers may attempt to maximise presence by incentivising rangers to travel further and stay on patrol for longer, whereas over longer timescales (e.g., one year) managers may attempt to maximise presence by varying the locations to which patrols are deployed.

Table 4.1. Patrol presence measures, industry benchmarks and evaluation thresholds.

Measure	Unit	Industry benchmark	'Good' coverage	'Moderate' coverage	'Poor' coverage
Spatial coverage per month	%	≥75% (of readily accessible areas)	≥37.5%	≥18.75% but <37.5%	<18.75%
Spatial coverage per year	%		≥75%	≥37.5% but <75%	<37.5%
Temporal coverage per month	24 h days	15 days/nights (i.e., ≥50%)	≥15 days	≥7.5 days but <15 days	<7.5 days

I included all ground-based transport types in calculations (e.g., foot, car and motorbike patrols), and included records from the entire monitoring period for each site, except for the first 3 months following implementation of SMART and the final timestep. In some sites all patrols were monitored using SMART from the outset and as such patrol data were an accurate representation of coverage. In other sites the proportion of patrols monitored increased through time, as successive ranger teams were trained and equipped. For these sites, coverage would be underestimated until all patrols were monitored, and increase artificially as teams were added. Excluding the first 3 months of data allowed a period for all patrols to be monitored. I also excluded the final timestep for each measure, which might include partial data. All data were processed and analysed within the R software environment (R Core Team, 2018).

To estimate spatial coverage, I constructed a grid of 1 km cells, corresponding to each protected area's boundary, and calculated the percentage of unique cells through which patrol routes passed per timestep, up to a maximum of 4 years following the 3-month burn-in period. Site boundaries were predominately sourced from site management. In a few cases, boundaries were source from the World Database on Protected Areas (IUCN and UNEP-WCMC, 2018), if these matched maps used by management in SMART. I estimated patrol routes by assuming the shortest route between successive position records. The industry benchmark for monthly spatial coverage did not specify at what spatial resolution the standard should be applied. Consequently, I used a 1 km grid cell size, which is the default resolution in SMART software.

To estimate temporal coverage, I calculated the total time each month (in 24 h days) during which one or more ranger teams were on patrol, in accordance with the industry benchmark. As patrols generally started from patrol posts or headquarters outside site boundaries, I included patrol activity starting within a 10 km boundary of each protected area, which managers considered to be time spent on patrol. I calculated totals irrespective of time of day. For example, 15 days' coverage could equate to 30 non-contemporaneous 12-hour patrols during daylight hours, one continuous patrol that endured for 15 days and nights, or a mixture of overlapping patrols at varying times of day or night, which in total provided 15 days (or 360 hours) of presence by one or more patrols.

4.2.2.2 Estimating trends in patrol presence over time

I estimated trends in patrol presence over time following implementation of SMART for each site, using the same three coverage measures. To account for seasonality in monthly data, which was apparent in non-tropical sites, I decomposed coverage time series using Seasonal Trend Decomposition by Loess and subtracted the estimated seasonal component from the original series (Cleveland *et al.*, 1990; Chandler and Scott, 2011). Where necessary (one site), I interpolated missing values in monthly time series using a seasonal Kalman filter (Zeileis and Grothendieck, 2005). I estimated rates of change in coverage by fitting simple linear regression models to the deseasonalised data with coverage measure as the response variable and time as the explanatory variable. For five sites there were insufficient monitoring data to estimate trends over time in annual spatial coverage (i.e., <three years' data). I also calculated average rates of change across sites for each measure.

4.2.2.3 Evaluation and summarisation

I evaluated results for the state of patrol coverage within each site with respect to industry guidelines. I set thresholds for 'Good', 'Moderate' and 'Poor' coverage, drawing on best-practice benchmarks, and assessed each site's results against these thresholds (Table 4.1). The benchmark for temporal coverage (minimum of 15 24 h days per month) was directly applicable. The benchmark for spatial coverage specified that patrols should cover at least 75% of 'readily accessible' areas per month but did not define what makes an area readily accessible. For the evaluation, I assumed that 50% of each site met this criterion, so setting a benchmark of 37.5% of the entire area visited per month. There was no associated benchmark for spatial coverage at the annual scale. However, throughout the course of a year patrols should be aiming for broad coverage of sites, whether areas are easily accessible or not (Plumptre *et al.*, 2014). Consequently, for spatial coverage per year, I set thresholds in line with the monthly industry benchmark.

For the purposes of the evaluation, I also summarised results for trends over time in coverage following implementation of SMART. For each site, for both spatial coverage measures, I classified a rate of change of ≥2.5%/year or ≤-2.5%/year as a clear positive or negative trend, respectively. For monthly temporal coverage, I classified a rate of change of ≥0.75 days/year or ≤-0.75 days/year as a clear positive or negative trend, respectively. I used rate of change rather than absolute change as the length of timeseries differed between sites. I only classified a trend over time as positive or negative if it was significant (at the .05 level) for monthly measures or explained >60% of the variance for the annual measure.

4.2.2.4 Statistical analysis of factors influencing patrol presence

I assessed factors influencing differences in patrol presence, focussing on whether differences in mean coverage between sites could be explained by resources available for patrolling, and whether implementation of SMART management mechanisms influenced improvements in coverage over time. All relationships were assessed by fitting linear mixed effects models to the data (see specific response and explanatory variables for each model below) using restricted maximum likelihood (with significance at the $p<.05$ level), with site nested within country as a random factor, as sites situated within the same country were similar in multiple, potentially-influential aspects (e.g., geography and management). Model assumptions were checked via examination of residuals. All models were generated using the nlme package in R (Pinheiro *et al.*, 2018). Coefficient of determination (R^2) was computed using the Kenward-Roger approach (Edwards et al., 2008).

To assess whether differences in mean coverage between sites could be explained by resources available for patrolling I fitted individual models to each coverage outcome, with annual budget for site-level law enforcement (mean=\$241/km^2 ± 231 SD) and density of rangers employed for enforcement patrolling (mean=0.035 rangers/km^2 ± 0.045 SD) included as continuous explanatory variables. Budget data were only available for 20 out of 21 sites, so I limited the analysis to these sites. To account for temporal pseudoreplication arising from repeated measurements within each site, I also included year of measurement as a continuous random effect for the annual outcome, and month of measurement as a continuous fixed effect with a first-order autoregressive autocorrelation structure for the monthly outcomes.

To assess whether SMART management mechanisms designed to increase performance influenced improvements in coverage over time, I employed a 'dose-response' approach. Examining dose-response patterns between the intensity at which an intervention is applied (the dose) and its outcomes (the response) is an increasingly common approach to causal attribution in programme evaluation, which cannot infer the effect of management, but can provide confirmatory evidence for a relationship when used in conjunction with a well-developed theory of change (Rogers, 2014). The hypothesised causal pathway between SMART's management activities and its intended outcomes has been clearly described (chapter 2). I examined this relationship, using fidelity of implementation of SMART activities as a measure of intensity of application of the intervention. The faithfulness with which SMART monitoring and management activities are implemented varies between sites (chapter 3), which will influence whether outcomes such as increased coverage are achieved (chapter 2). I used an index that captured fidelity of implementation of SMART in each site (chapter 3). The index measures fidelity on a scale from 0 to 1, 1 representing 'perfect' implementation for that variable, and 0 the obverse (mean=0.64 ± 0.13 SD). I fitted mixed effects models to each outcome, with rate of change over time in coverage following SMART implementation as the response variable and the fidelity index as the explanatory variable.

4.3 Results

4.3.1 Patrol presence within and across sites

Results for spatial patrol presence varied by temporal scale (Fig. 4.2). At the monthly scale, spatial coverage was generally very low across sites (mean of means = 13.3% ± 12.9), and extremely low (mean ≤5%) in 43% of sites. At the annual scale, spatial coverage was better but still low overall (mean of means = 43.0% ± 25.7). In 57% of sites, patrols covered less than half the area under protection per year. On average, spatial coverage at the annual scale was multiple times higher than at the monthly scale (4.77 ± 2.00 times higher), and this difference increased over time. In contrast, temporal coverage was generally high across sites (mean of means = 17.3 days ± 9.0 per month) (Fig. 4.3). Mean temporal coverage was ≥15 days per month, the recommended industry minimum, in 62% of sites. Mean temporal coverage was lowest in sites with strong seasonality in patrolling (two sites).

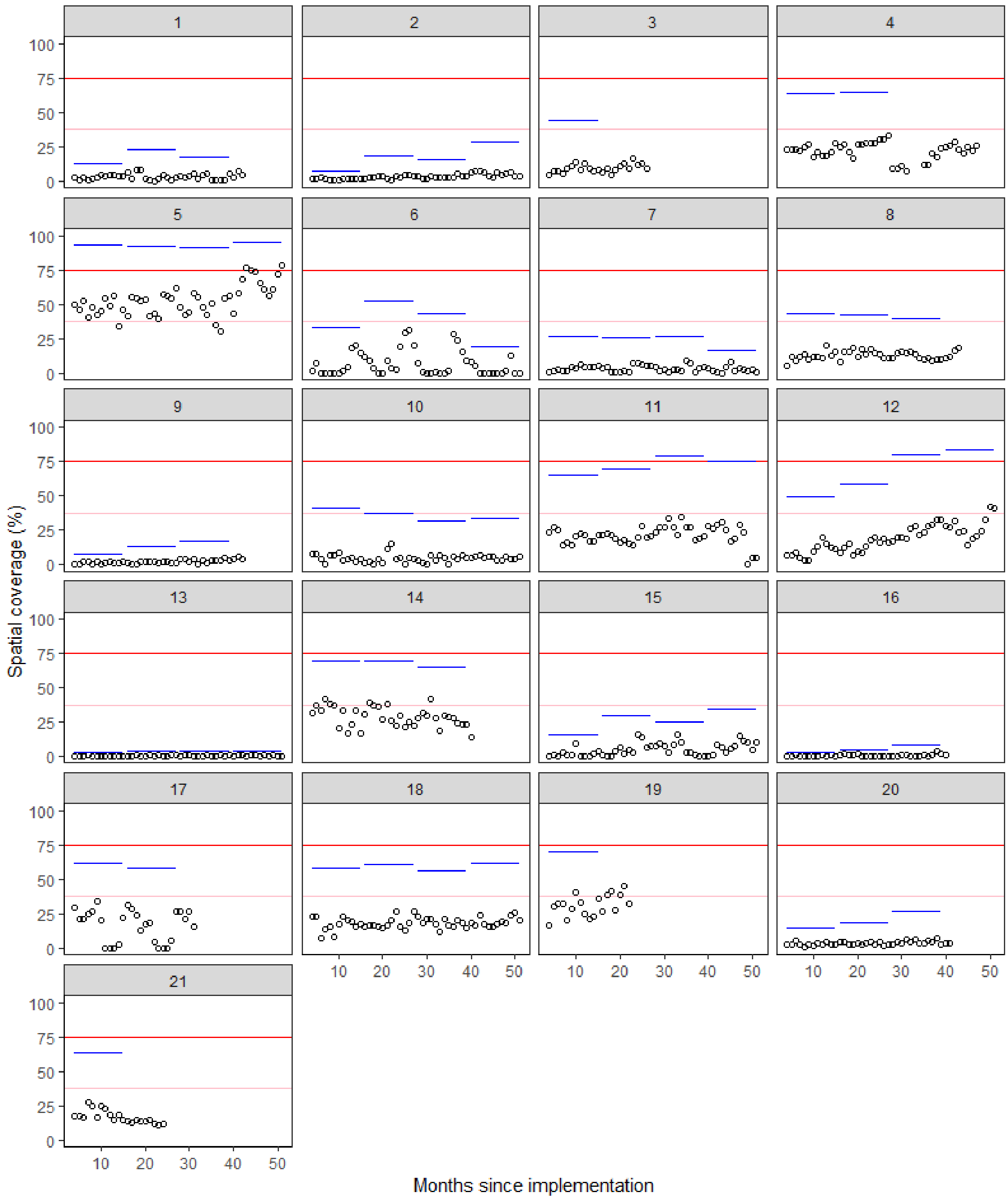


Figure 4.2. Spatial coverage (%) at 1 km resolution of protected areas by patrols per month (open points) and per year (blue lines) over time following SMART implementation, ordered by site ID. Pink and red lines indicate evaluation thresholds for 'Good' coverage per month and per year, respectively. Annual coverage only shown where 12 months of data were available.

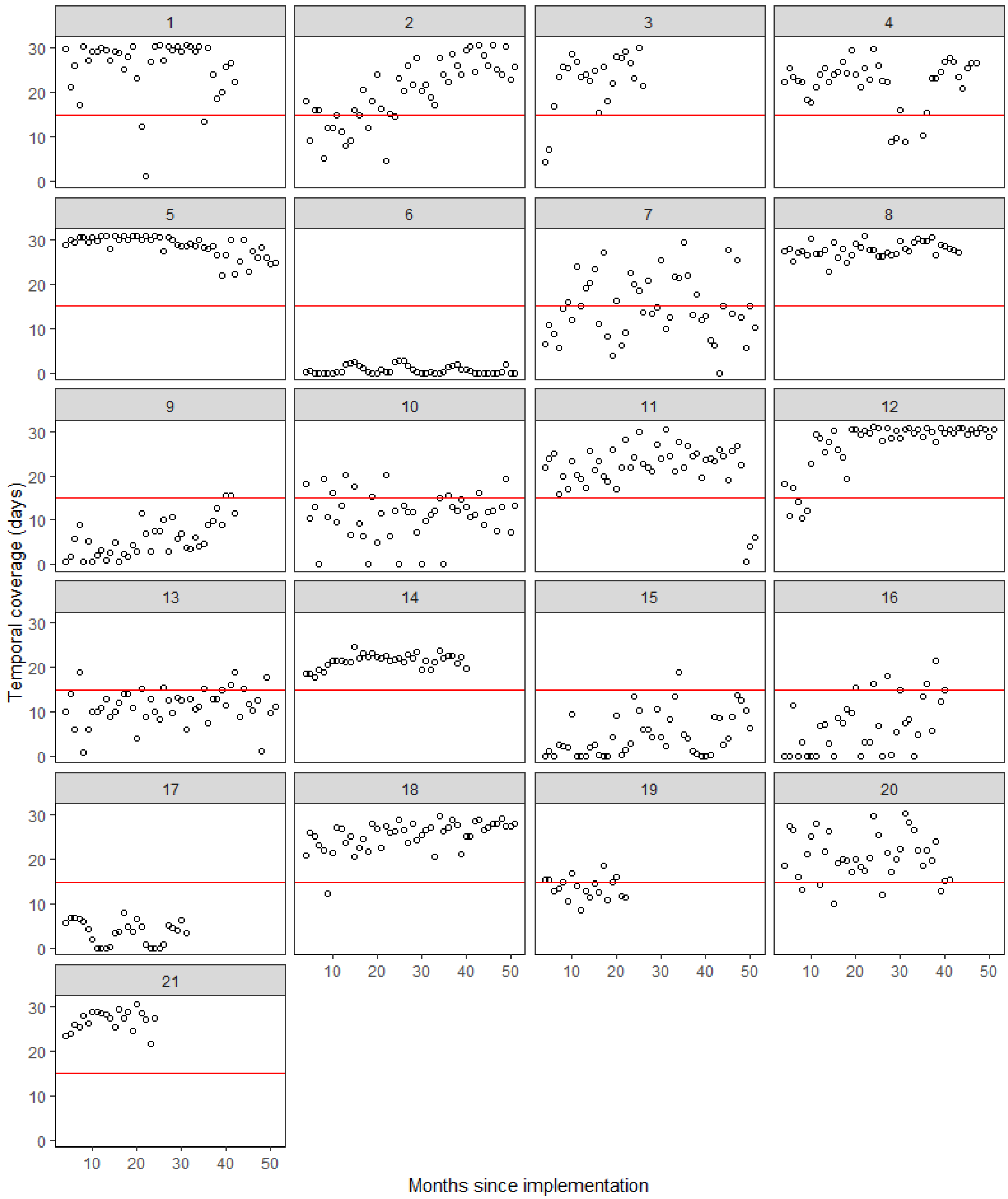


Figure 4.3. Temporal coverage (in 24 h days) of protected areas by patrols per month (open points) over time following SMART implementation, ordered by site ID. Red line indicates industry minimum benchmark (15 days/nights per month) and evaluation threshold for 'Good' coverage.

4.3.2 Trends in patrol presence over time

Rates of change in patrol presence over time following SMART implementation also varied between sites and by outcome (Fig. 4.4). At the monthly scale, change over time in spatial coverage was positive but negligible (mean across sites = 0.06%/month ± 0.26, equivalent to 0.72%/year ± 3.14). However, at the annual scale, change over time in spatial coverage was positive and moderately fast (mean across sites = 1.92%/year ± 4.48), with rates of ≥2.5%/year in 50% of sites. Rates of change over time in temporal coverage were similarly positive (mean across sites = 0.08 days/month ± 0.16, equivalent to 1.00 days/year ± 1.97), with significant positive change of ≥0.75 days/year in 33% of sites.

4.3.3 Evaluation and summary

Spatiotemporal patrol presence in protected areas around the world varied between sites, by outcome and changed through time following SMART implemementation, but performance – with reference to industry benchmarks – was predominately poor (Fig. 4.5).

Across sites, results indicate that patrols often provided good temporal coverage, but spatial coverage was far poorer, at both monthly and annual scales, such that the majority of many sites was rarely or never patrolled. Spatial coverage at the monthly scale was particularly poor, with patrols in 71% of sites providing <18.75% coverage per month, and frequently far lower. At the annual scale, patrols performed slightly better, but spatial coverage was still poor in 48% of sites, and only moderate in the same proportion. In contrast, temporal coverage was good in 57% of sites and only poor in 24%.

While results were predominately poor, there was evidence of positive trends in presence over time following SMART implementation, for two of the three outcomes. Clear, positive trends in spatial coverage at the annual scale were apparent for 44% of sites, and in temporal coverage for 33% of sites (vs. negative change for 13% and 6%, respectively). Moreover, positive trends were particularly apparent for sites with poor coverage (50% of sites with poor spatial coverage, at the annual scale, and 60% with poor temporal coverage). However, trends in spatial coverage at the monthly scale, the worst performing outcome, were negligible for the majority of sites.

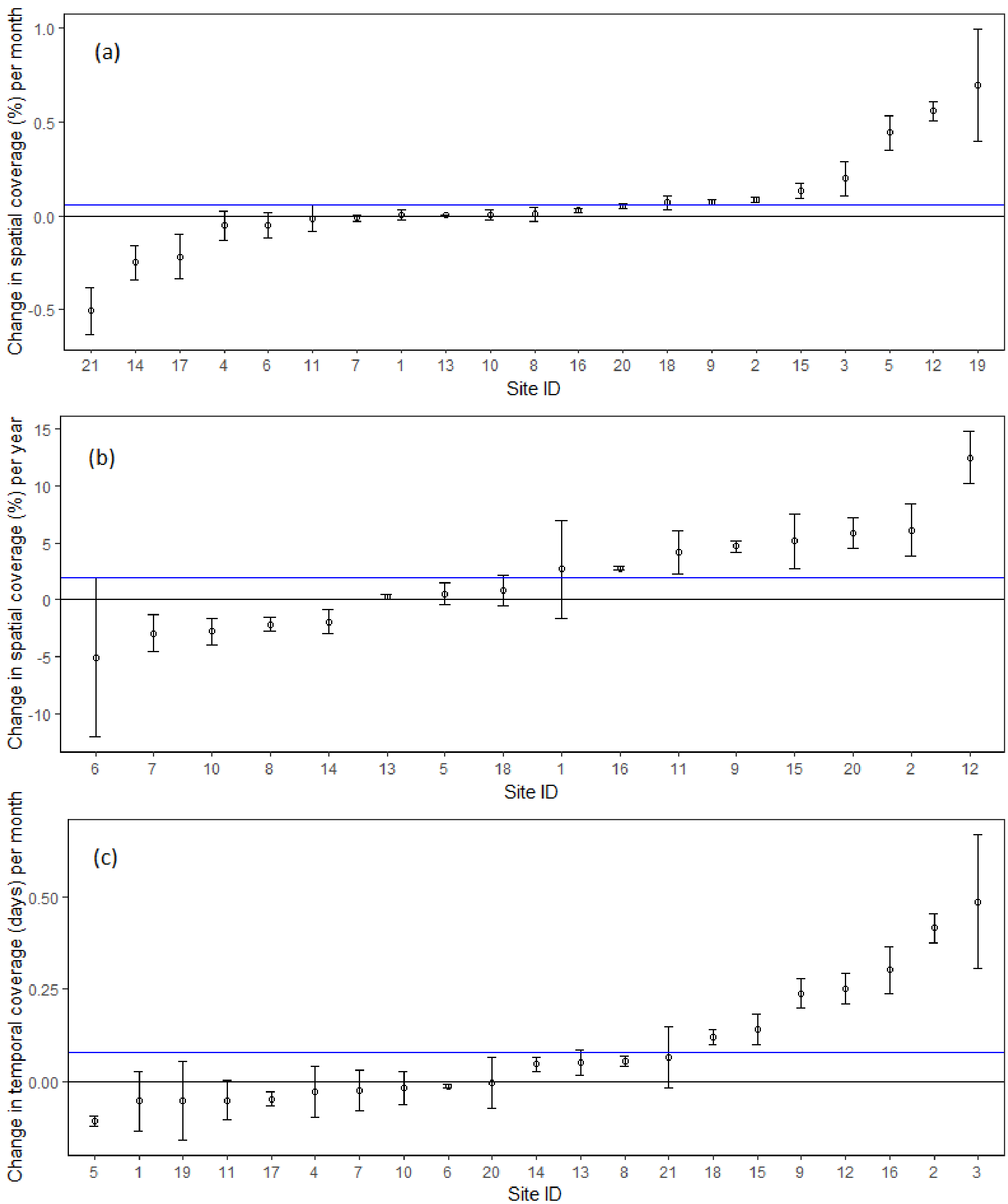


Figure 4.4. Rates of change over time (± SE) following SMART implementation in (a) monthly spatial coverage, (b) annual spatial coverage, and (c) monthly temporal coverage, provided by patrols in protected areas, ranked by rate. Blue line indicates mean rate of change across sites.

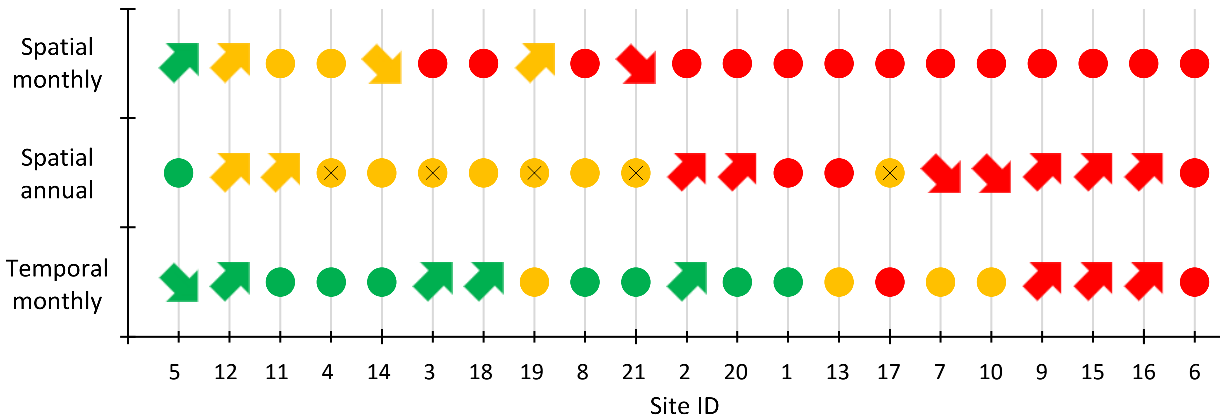


Figure 4.5. State and trends over time following SMART implementation in spatiotemporal patrol presence in protected areas globally, ordered by overall score, measured as spatial monthly (top), spatial annual (middle) and temporal monthly (bottom) patrol coverage. Colour indicates state of patrol coverage (green=good, amber=moderate, red=poor). Arrows indicate clear positive or negative trends over time (no arrow=no trend detected, x=insufficient data to measure trend).

4.3.4 Factors influencing patrol presence

Differences between sites in spatial coverage at the monthly scale was the only outcome that was significantly predicted by resources available for patrolling. Annual budgets for law enforcement, but not ranger density, significantly predicted monthly spatial coverage, but the effect was small (Table 4.2). The fixed effects in the final model explained 56.9% of the variance, with a partial R^2 of 0.566 for budget and 0.047 for month. A similarly positive relationship was also apparent between budgets for law enforcement and annual spatial coverage, but the relationship was non-significant.

Table 4.2. Predictors of differences in monthly spatial patrol coverage across sites. Random effects structure = ~ 1 | Country / Site ID. AR(1) correlation Structure = ~ Month | Country / Site ID.

Fixed effects	Estimate	Standard Error	df	t value	Pr (>|t|)
Initial model					
Intercept	0.0519	0.0513	780	1.0114	.312
Annual enforcement budget ($/km2)	0.0004	0.0001	5	2.6976	.043
Ranger density	-0.6606	0.6655	5	-0.9927	.366
Months since implementation	0.0008	0.0003	780	2.5651	.011
Final model					
Intercept	0.0453	0.0483	780	0.9377	.349
Annual enforcement budget ($/km2)	0.0003	0.0001	6	2.4952	.047
Months since implementation	0.0008	0.0003	780	2.5623	.011

Differences in trends over time following SMART implementation between sites were unrelated to implementation of SMART management mechanisms to improve performance. Fidelity of implementation of management mechanisms had no significant effect on rates of change for any of the three patrol coverage outcomes (Table 4.3).

Table 4.3. Predictors of differences in rates of change over time in three patrol coverage outcomes across sites. Random effects structure = ~ 1 | Country / Site ID.

Outcome Fixed effects	Estimate	Standard Error	df	t value	Pr (>|t|)
Monthly spatial coverage					
Intercept	-0.149	0.315	12	-0.474	.644
SMART fidelity index	0.367	0.474	7	0.774	.464
Annual spatial coverage					
Intercept	2.354	4.153	10	0.566	.583
SMART fidelity index	-0.569	5.807	4	0.774	.464
Monthly temporal coverage					
Intercept	0.396	0.172	12	2.299	.040
SMART fidelity index	-0.484	0.262	7	-1.848	.107

4.4 Discussion

Broad, consistent patrol presence is thought to be key to reducing illegal killing threatening biodiversity in protected areas, but patrol activity is poorly understood, hindered by historically inadequate monitoring. I exploited recent, widespread deployment of patrol monitoring using SMART to conduct a global analysis of state and trends in spatiotemporal patrol presence.

4.4.1 Patrol presence in protected areas

That patrols typically provided very low spatial coverage of protected areas at monthly scales, and low coverage at annual scales, suggests levels of patrolling observed may be insufficient to reduce illegal activity in the majority of protected areas. Industry targets assert that monthly visits to 75% of 'readily accessible' areas are required for effective protection (Singh *et al.*, 2015). Assuming 50% of each site is accessible, patrols in only one out of 21 sites provided such protection, and the remainder fell far short of this target. Importantly, industry targets are not based on empirical research. Indeed, the relationship between patrolling and illegal activity is poorly understood, and activity levels required in reality may be higher still (e.g., broad spatial coverage more frequently than once per month). The mechanisms underlying the deterrence effect of ranger patrols are particularly unclear (Dobson *et al.*, 2018). Previous studies have assumed that monthly visits (to a 1 km grid cell) are necessary for deterrence (Plumptre *et al.*, 2014). In practice, deterrence may operate over spatial scales <1 km (e.g., in dense forests), or when patrols have been present within far shorter timescales of an offender (e.g., within a few hours), or require presence at multiple points over these temporal scales to displace an illegal act outside of a site or to deter it entirely. Even in contexts with far higher activity levels than observed here (e.g., city policing) there is mixed evidence that patrolling deters crime (Nagin, 2013a). Similarly, patrol activity levels may be insufficient for adequate detection of crime. For example, detection rates of passive hunting devices, such as snares, are typically low in protected areas (O'Kelly *et al.*, 2018), suggesting multiple visits may be required to reduce the threat they pose.

The limited coverage observed is also a particular concern with regard to monitoring of crime. Information on the distribution of illegal activity in protected areas is essential for efficient and effective management (e.g., to direct limited resources towards high-risk areas). Ranger-based monitoring, via LEM tools, is increasingly popular for achieving these ends, because it exploits labour and effort which has already been committed (i.e., enforcement patrols) (Gavin *et al.*, 2010), potentially to the exclusion of other monitoring methods. The data rangers collect are used to predict spatial and temporal trends in illegal activity (Critchlow *et al.*, 2015), and thence to target patrols towards predicted crime hotspots (Critchlow *et al.*, 2017). Yet the levels of patrol activity I have found in the majority of sites suggests trends derived from ranger-collected data may be uninformative at best and misleading at worst. Large proportions of many sites were unpatrolled for long periods, meaning that there was essentially no information about illegal activity in these locations, rendering inference of crime in these areas difficult or impossible (Walters, 2003). In general, survey effort should be broad and consistent to be able to draw robust inferences about trends in illegal activity (Keane *et al.*, 2011). The levels of coverage observed were patchy and strongly non-random, and interpreting data derived from this effort will be challenging. Careful analysis may be able to generate reliable trends in contexts where patrol effort is broader and more consistent (e.g., using Bayesian hierarchical models (Critchlow *et al.*, 2015)), but such methods are not the norm, and may be difficult to implement for site managers. My findings are a particular concern given the plethora of Artificial Intelligence (AI) methods currently under development which use patrol data to predict crime in protected areas (Fang *et al.*, 2019). Predictions derived using AI methods can be especially hard to interpret (Wearn *et al.*, 2019).

That monthly spatial coverage was very low, was constrained by budgets, and changed little over time, suggests managers may be unable to broaden regular patrol coverage without additional funding support. Moreover, as annual spatial coverage was generally multiple times higher than monthly coverage, was less constrained by budgets, and improved over time, suggests managers may be employing tactics to achieve broad coverage over long timescales (e.g., by varying patrol deployment throughout the year), and they may be getting better at doing so. This may be a sensible strategy if the goal is to achieve broad but infrequent coverage, but it comes at the expense of providing regular ranger presence in any one area. Moreover, coverage was still demonstrably low at annual scales, and neither approach may be an effective means to reduce or monitor illegal activity.

Few studies have assessed patrol activity using comparable measures, but those that have also reported similar results. Annual spatial coverage of a protected area in the Russian Far East varied between 18-27%, which the authors attributed to the site's large area and remoteness (Hötte *et al.*, 2016). Similarly, an analysis of spatial coverage in 12 sites throughout the Greater Virunga Landscape in central Africa reported that, on average, only 22.9% of sites was patrolled at least monthly (Plumptre *et al.*, 2014). Moreover, only 60% of sites had been patrolled at least once throughout the 5+ year monitoring period. However, the authors attributed much of the unpatrolled area at this temporal scale to one site in which insecurity and the presence of armed rebels made access difficult. This is the first analysis of patrol coverage outside of these two contexts, drawing on data from a broad sample of sites across Africa and Asia and integrating hitherto neglected sub-regions of both continents. My findings suggest that poor spatial patrol coverage is common in sites around the world. No studies report comparable measures of temporal patrol coverage. My results are also consistent with indirect evidence that suggests patrol activity may be inadequate in many sites. For example, a global analysis of protected area management effectiveness found that sites generally scored poorly for indicators of law enforcement adequacy (Leverington *et al.*, 2010). Studies of drivers of patrol activity (e.g., funding) are particularly rare. However, examination of law enforcement in Lao PDR found a strong positive correlation between financial investment in enforcement and annual foot patrol effort (Johnson *et al.*, 2016). Protected area effectiveness, in general, is hampered by major shortfalls in financial support globally (Watson *et al.*, 2014).

Whilst my results are consistent with the few studies that have previously assessed patrol presence in protected areas, there were limitations in the study's design which may influence my findings' generalisability. Firstly, the study only included sites in which patrols were monitored using SMART. As sites are not randomly selected for SMART implementation, my findings may not be representative of protected areas outside of this group. In general, it is likely that sites which have the resources and capacity necessary (e.g., well-trained and equipped staff) are more likely to be selected for implementation of an LEM tool such as SMART. Consequently, sites without patrol monitoring implemented may have lower resources and capacity, which might correlate with poorer patrol activity. Nevertheless, if my results are better than average, this only serves to strengthen the conclusion that patrol presence in protected areas is generally poor. Secondly, the sample only comprised sites which were willing and able to share sensitive patrol monitoring data. Data from sites outside of this sample may be difficult to obtain, but a comparison of results in sites where data was easily obtained vs. challenging might elucidate whether there is an important difference and its direction.

Limitations in patrol monitoring data may also have influenced the accuracy of results. Firstly, spurious waypoints which indicated unlikely velocities were common at low frequencies across sites. Identifying such points is non-trivial as transportation type, which will determine velocity, was not consistently recorded. However, patrols in a few sites also gathered tracklogs – sequences of high-frequency position records (e.g., at 1 second intervals) collected automatically by GPS units, which produce a more accurate depiction of route travelled. This provided an opportunity, for a subset of sites, to compare results derived from tracklogs against results derived from waypoints which included spurious observations. For all sites assessed the difference in results was negligible, and the effect of spurious waypoints was to overestimate spatial coverage. Secondly, I estimated patrol routes as the shortest route between successive waypoints, which is an inexact approximation of human behaviour in protected areas (Papworth *et al.*, 2012). Moreover, the accuracy of route estimates will vary with the frequency and regularity of recording of waypoints by rangers, which varied between sites. Nevertheless, the comparison of results from tracklogs and waypoints described previously indicated a negligible effect on results.

In addition, there are multiple issues with using percentage of protected area visited by patrols per timestep as a measure of patrol presence. Firstly, percentage coverage is entirely scale-dependent (i.e., coverage is proportional to cell size). I assumed a 1 km cell size, which was the default in SMART software. In reality, different scales may be important for evaluating patrol coverage. For example, in closed environments with low visibility (e.g., forests) a cell size <1 km may be relevant for the detection and deterrence of crime. Conversely, in environments with high visibility (e.g., savannahs) the obverse may be true. These relationships are poorly understood (Dobson *et al.*, 2018), but future research should investigate and integrate scale-dependency into analyses, and include variables such as habitat to account for differences. Secondly, percentage coverage is a crude measure of patrol presence, which ignores activity within and across visited cells during each timestep, as demonstrated by differences between monthly and annual spatial coverage results. Even when coverage is relatively high, activity may be skewed towards common areas, such as near patrol posts (Plumptre *et al.*, 2014). A more representative measure of presence, particularly at longer timescales, would also account for the magnitude and evenness of patrolling. Nevertheless, spatial coverage is a recognised industry measure, but researchers and managers should consider development of more precise indicators. Thirdly, in practice, the spatial coverage which managers aim to achieve will be <100%, and potentially far lower, as large areas of sites may not require protection (e.g., because threatened species are not present, or the locations are too water-logged or steep for offenders to access). Moreover, the proportion of area requiring protection will also vary by site. Yet identifying site-specific targets is challenging, as managers, who hold knowledge necessary to delineate priority areas, may have a vested interest in providing biased assessments to inflate results. Future research should aim to find objective methods for delineating priority patrol areas within sites (e.g., by computing distance to established routes and waterways, or slope). Finally, managers may also explicitly plan to deploy patrols to site boundaries and avoid effort within protected areas in order to deter ingress and to reduce the creation of access points or paths. In such contexts, percentage coverage of sites may be an inappropriate measure of performance. Patrols in the majority of sites in this study did not appear to follow this pattern, but delineating priority patrol areas inside and outside boundaries would address the issue in future research. Many of these issues are also apparent for a measure of monthly temporal coverage and could be addressed in similar ways. For example, 15 days coverage could be concentrated during daylights hours, rather than randomly distributed. Conversely, only daylight hours may require protection. Measures and targets should take such factors into account.

4.4.2 Influence of SMART on change in patrol presence over time

On average, across SMART sites, patrol presence improved over time following implementation of the tool for two out of three outcomes measures, and changed fastest where the need for improvement was greatest. Similar studies have reported improvements in patrol performance outcomes in sites implementing SMART (Hötte *et al.*, 2016). Yet whether and to what extent these changes are a result of SMART is still unclear. Importantly, coverage may be improving in in all protected areas, not just SMART sites, or some other characteristic of sites in which SMART is likely to be implemented may correlate with improving outcomes. For example, sites implementing SMART may also receive concurrent support for non-SMART aspects of law enforcement (e.g., funding for ranger salaries or equipment (chapter 2)). I attempted to confirm whether SMART had had an effect by assessing variation in outcomes within treated sites using dose-response analysis. I found no relationship between rates of change and variation in implementation of SMART performance management mechanisms. However, my analysis does not prove that the intervention is not effective or that performance management mechanisms *per se* are not worthwhile. Dose-response analyses of the kind used here are designed to provide confirmatory evidence of causal effects (i.e., the presence of a dose-response relationship strengthens confidence in the capacity to infer causality) (Hill, 1965). Had I found an effect it would also have been necessary to account for confounding factors (e.g., concurrent support for non-SMART aspects of law enforcement). However, the absence of a dose-response association does not discount a causal relationship (Reynolds, 1998). For example, the causal relationship may be non-linear or exhibit threshold effects. Furthermore, limitations in my experimental design may also have reduced power to detect an effect. For example, the fidelity index I used integrated fidelity of all SMART monitoring and management activities, but individual activities may be responsible for effects, or SMART may have had an effect but other factors (e.g., changing resources available for patrolling) exerted a stronger influence, which it was not possible to discern using a sample of this size. Future analyses should explore the effect of different management activities individually and include a larger sample of sites, if possible.

4.4.3 Conclusion & Recommendations

Illegal natural resource use, such as poaching, is a major threat to biodiversity in ostensibly protected areas around the world (Harrison, 2011). Ranger-led law enforcement patrols are the primary response to this threat (Henson *et al.*, 2016). My findings suggest that patrolling, as currently practiced, may be inadequate to reduce illegal activity in protected areas, because of insufficient funding. For patrols to be able to provide broad, regular presence in protected areas, which is sufficient to generate adequate monitoring data, and which has the potential to reduce illegal activity, much greater financial support is urgently required for patrol activities, alongside management mechanisms to improve patrol performance, such as via SMART. Concurrently, given the low coverage observed, there is a similarly pressing need to recognise the limitations of ranger-based monitoring via LEM tools such as SMART, to analyse and interpret patrol data with care, and to continue to employ alternative, rigorous methods for monitoring until coverage is improved.

Chapter 5

Do ranger-led enforcement patrols deter illegal activity in protected areas? Evidence from application of a differenced-CPUE metric to ranger-collected data

Abstract

Ranger patrols are the primary means by which protected area managers aim to deter illegal killing of wildlife within park boundaries. However, evidence for deterrence operating in practice is lacking, because most available data on illegal activity, which are collected by rangers on patrol, are vulnerable to bias. Common metrics of deterrence that use patrol data, such as plots of illegal activities detected per unit of patrol effort (CPUE) over patrol effort, do not adequately account for biases. 'Differenced plots' of temporal change in CPUE over change in effort have been proposed as an alternative, robust metric, which reliably identified deterrence using simulated data. This method, although promising, has only been applied to synthetic data, and there are remaining questions for its application in practice. To assess whether I find evidence that patrols deter illegal activity I apply differenced plots to real patrol data collected in four diverse protected areas around the world, using two indicators of illegal activity: snares and direct observations of people. I examine whether differences in deterrence between sites can be explained by differences in habitat type. I also explore methods for applying differenced plots in practice, including plausible time intervals and lags and approaches for aggregating catch data. My results suggest that deterrence may have been operating in one of four sites for snares, and two of four sites for people, but the relationships were weak and only apparent at some timesteps, the temporal resolution of which varied amongst sites. The absence of consistent evidence of deterrence could indicate that patrols do not reliably deter illegal activity or that my application of differenced plots was not sufficiently sensitive. My findings also indicate appropriate methods for aggregating catch data and suggest that exploration and selection of appropriate time intervals is crucial when using differenced plots. Questions remain for future applications, including appropriate spatial scales.

5.1 Introduction

Illegal killing of wildlife threatens biodiversity in protected areas around the world (Laurance *et al.*, 2012; Schulze *et al.*, 2018). Poaching, in particular, is driving declines in ostensibly protected bird and mammal populations throughout the tropics (Tranquilli *et al.*, 2014; Benítez-López *et al.*, 2017; Gray *et al.*, 2017). For example, the illegal killing of African elephants has resulted in precipitous declines in certain populations (Wittemyer *et al.*, 2014). The primary means by which protected area managers attempt to address this threat is through investment in ranger-led law enforcement patrols (Henson *et al.*, 2016). Across sites, investments in site-level law enforcement, such as ranger density, are associated with positive conservation outcomes (Bruner *et al.*, 2001; Chhatre and Agrawal, 2008; Tranquilli *et al.*, 2012). Conversely, wildlife is most threatened by illegal killing where local enforcement is stretched (e.g., African elephants (CITES, 2010)). However, understanding of the processes underlying the relationship between enforcement and illegal activity within sites is poor.

Deterrence, the process of discouraging rule-breakers from committing offences through fear of apprehension and punishment (Pratt and Cullen, 2005), is the primary mechanism by which ranger patrols are assumed to reduce illegal activity. In theory, patrols could also reduce crime through incarceration of offenders or reduce its impacts by eliminating passive hunting devices (e.g., snares) (Keane *et al.*, 2008). Yet detection rates in protected areas are generally low (e.g., of snares (O'Kelly *et al.*, 2018)), so deterrence is the principal mechanism through which patrols are thought to act. However, evidence for deterrence operating in protected areas in practice is lacking. In general, the deterrence effect of policing on crime is difficult to identify, even in contexts with ample data where crime is regularly reported (Paternoster, 2010). In protected areas, the issue is acute, as most information on illegal activity is collected by rangers, who record observations of crime encountered opportunistically whilst on patrol (e.g., Critchlow *et al.* (2015); Johnson *et al.* (2016)). Identifying evidence of deterrence in ranger-collected data is particularly challenging. Consequently, few rigorous, empirical studies have demonstrated deterrence in protected areas (e.g., Beale *et al*. (2018)).

Ranger-collected data are vulnerable to multiple sources of bias and thus require careful post-hoc analysis, but common metrics of deterrence that use patrol data do not adequately account for biases and can be misleading (Keane *et al.*, 2011). Opportunistic encounter data in general, such as voluntary biological surveys, are subject to biases arising from uneven sampling effort and detectability, which generate significant noise (Isaac *et al.*, 2014). Analysing data collected by rangers on patrol can be particularly problematic, as rangers' primary focus is law enforcement, not monitoring (Gray and

Kalpers, 2005). Consequently, patrol effort is typically directed towards locations where and periods when illegal activity is expected to occur (Hötte *et al.*, 2016). To account for variation in effort, managers generally use simple Catch Per Unit Effort (CPUE) measures of the number of illegal activities detected per unit of patrol effort (e.g., snares encountered per patrol day (Stokes, 2010)). These measures can be difficult to interpret but, nevertheless, are often used to identify deterrence. For example, common deterrence metrics, such as plots of CPUE over time (Hötte *et al.*, 2016) or CPUE over patrol effort (Hilborn *et al.*, 2006), fail to account for temporal changes in illegal activity that are unrelated to patrolling or ignore temporal autocorrelation in patrol data. Such measures are straightforward to apply, which is essential for managers, but can be misleading.

Differenced plots of change in CPUE over change in patrol effort (henceforth, 'differenced plots') have recently been proposed as an alternative metric to diagnose whether patrols deter illegal activity using ranger-collected data, which is designed to be more robust to confounding temporal effects and autocorrelation but remains relatively simple to apply (Dobson *et al.*, 2018). A negative correlation between change in CPUE and change in patrol effort, indicating that the rate of appearance of illegal activities decreased as patrol effort increased, would suggest that deterrence is operating. Applied to patrol data simulated using a simple mechanistic model of poaching and patrolling, differenced plots reliably identified deterrence, regardless of reductions in poaching unrelated to patrolling (Dobson *et al.*, 2018).

Differenced plots, although promising, have only been applied to synthetic data, and their wider application is unknown. In addition, there are remaining questions for application of the approach to diagnose deterrence in practice (i.e., area of analysis, scale of spatial and temporal aggregation, effort and catch measures, time lags between effort and catch) (Dobson *et al.*, 2018). For example, it is unclear at what temporal scale deterrence operates. Confining analyses to one scale risks missing effects operating elsewhere (e.g., revaluating arrest and crime data from Oklahoma City at monthly, quarterly, and semi-annual scales found deterrence effects which had formerly been missed at the annual level (Chamlin *et al.*, 1992)). Conversely, searching for deterrence at multiple scales risks generating spurious correlations by chance. Deterrence may also vary with context. For example, deterrence may vary between protected areas, according to site-level characteristics such as habitat type, and between illegal activity types, which differ in terms of their persistence in the landscape and detectability by rangers (Dobson *et al.*, 2018). Consequently, if differenced plots are to become a practical tool, the effects of different analytical decisions on the performance of this metric need to be explored using real patrol datasets that have been collected in a variety of different contexts. Historically, aggregating patrol data from multiple protected areas was difficult as sites used inconsistent protocols and systems. Recent, broad-scale deployment of tools for standardised patrol monitoring, such as MIST (Management Information SysTem) and SMART (Spatial Monitoring And Reporting Tool), provide a unique opportunity for cross-site comparisons (Stokes, 2010; Hötte *et al.*, 2016).

Here, I apply differenced plots to real patrol data for the first time to assess whether I find evidence that presence of ranger-led enforcement patrols deters illegal activity. I assemble SMART patrol data from four diverse protected areas in which biodiversity is threatened by poaching, situated in four countries across tropical Africa and Southeast Asia. I assess evidence of deterrence in each site using two different indicators of illegal activity: snares and direct observations of people. These activities were commonly recorded by rangers in all four sites and occupy opposite ends of the persistence spectrum, with snares remaining detectable in the landscape for weeks or months (Coad, 2007), while direct encounters with people are relatively ephemeral. I also examine whether any differences in the apparent level of deterrence between sites can be explained by differences in habitat type. Finally, I explore the consequences of data processing decisions that must be taken when applying differenced plots in practice, including the effect of a small set of plausible time intervals and lags on the strength of the association between patrol effort and illegal activity, and the effect of using two alternative ways to aggregate observations of illegal activity, either using a simple count technique or using a grid cell occupancy approach.

5.2 Methods

5.2.1 Site selection & Patrol data

I selected four terrestrial protected areas in which: (1) rangers on law enforcement patrols recorded their patrol routes and observations of illegal activity using standardised SMART monitoring protocols; (2) poaching had been identified as a significant threat to biodiversity; and (3) records of snaring and direct observations of people occurred throughout the time period. I also chose sites with relatively high levels of patrol effort (Table 5.1), in which deterrence might be more likely to operate, and where higher survey effort should increase the likelihood that there are sufficient data to detect it. I selected two forest-dominated sites and two savannah-dominated sites. I hypothesised that I would find greater evidence of deterrence in sites dominated by open habitat, such as savannah, than in sites dominated by closed habitat, such as forest, as open habitats might increase the visibility of rangers, or provide fewer refuges for rule-breakers.

Table 5.1. Patrol effort statistics for the four study sites. Rangers were on patrols throughout the study period at each site (i.e., rangers were on patrol for a portion of every timestep) but spatial coverage of patrols per timestep varied (coverage measured as percentage of unique 1 km cells visited per timestep).

	Spatial patrol coverage (%)								
Site	14-day timestep			28-day timestep			42-day timestep		
	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
Forest-dominated									
1	14.7	16.4	7.9	29.4	28.2	7.6	38.1	36.4	8.9
2	19.7	20.2	5.4	29.8	30.5	5.7	37.2	37.6	5.9
Savannah-dominated									
3	19.7	19.7	11.7	31.2	30.5	15.9	39.2	37.7	17.5
4	9.2	10.1	3.6	15.6	17.1	5.1	20.4	22.2	6.4

I assembled patrol data from the four study sites. The combined dataset consisted of 200,035 position records from 7,082 ground-based ranger patrols. SMART-enabled rangers used handheld GPS units to record the time and location at the beginning and end of patrols, when they observed signs of illegal activity or wildlife, and at regular intervals in-between. All patrol data were collected between June 2013 and October 2017, although the start and duration of time period varied by site, beginning with the implementation of SMART (Fig. 5.1). Two sites were located in Eastern Africa, one was in Western Africa, and one was in South-eastern Asia. The four sites also varied in terms of other factors, such as area (mean=2,671 km^2 ± 2,005 SD), ranger density (mean=5.7 rangers/100 km^2 ± 7.6) and management objectives (e.g., National Park vs. Protected area with sustainable use of natural resources). To encourage participation, the sites' identities were anonymised by removing identifying features (e.g., name, location and area) from all outputs and assigning randomly-generated number IDs.

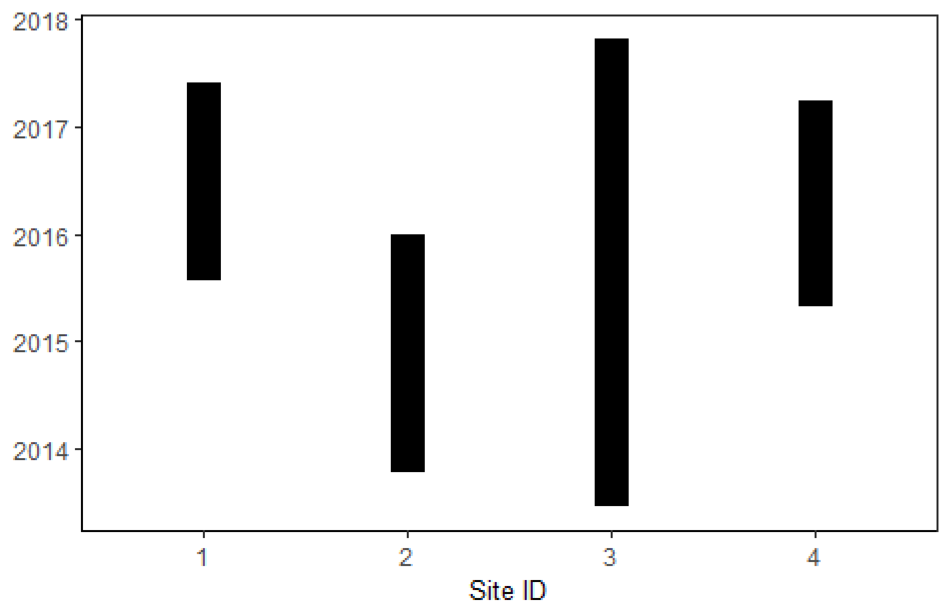


Figure 5.1. Time period of monitoring data by study site, following implementation of SMART.

5.2.2 Differenced plots, Illegal activity indicators & Patrol effort measure

To determine whether the presence of ranger-led enforcement patrols deterred illegal activity I used patrol data to generate differenced plots of change in illegal activity detected per unit of patrol effort over change in patrol effort (ΔCPUE-ΔE), for each study site, and inspected the plots for signals of deterrence. I fitted linear regression models to the data to assess the significance of the relationship between the two variables and to obtain a measure of model fit (r^2), and plotted regression lines with 95% confidence intervals. In the presence of deterrence I expected differenced plots to display a significant negative correlation, with higher r^2-values indicating a clearer relationship (Dobson *et al.*, 2018). To examine the effect of habitat type I compared the apparent level of deterrence in the two forest-dominated sites against that in the two savannah-dominated sites.

I assessed the effects of patrol effort on two different indicators of illegal activity. Rangers using SMART assign observations of illegal activity to descriptive categories according to a defined data model. For example, wire snares are generally assigned to a top-level category, such as 'Weapons', with 'Weapon type' set to 'Wire snare'. SMART allows database managers to define their own data model, so configurations frequently vary between sites. However, some categories are sufficiently consistent to permit cross-site comparisons. I analysed observations of different activity types separately. Different observations types persist in the landscape for varying durations and thus differ in terms of their detectability by rangers (e.g., a snare may persist for months following placement (Coad, 2007), whilst a gunshot is only detectable for a few seconds after the event). Aggregating across types may thus render differenced plots difficult to interpret (Dobson *et al.*, 2018). Moreover, the occurrence of different illegal activity types may be influenced by different deterrence processes. Consequently, I focused on two specific observation types, which were sufficiently common across and within sites, and only aggregated within type. I chose types at either extreme of the persistence spectrum: snares, and direct observations of people. I expected difference plots for the illegal activity type with longer persistence (snares) to display weaker negative relationships, as changes in effort may have impacts beyond the consecutive time-step (Dobson *et al.*, 2018).

In three sites, rangers recorded whether a trap was observed and the type of trap. Here, I confined the analysis to traps classified as wire snares, which comprised all snare records and nearly all trap records. I also confined the analysis to "New" traps in the one site in which trap age was recorded (either as "Old" or "New"). Rangers in the fourth site recorded only whether a snare of any type was observed, so here I included all snare records. Rangers in all sites recorded direct observations of people within PA boundaries. In most cases, observations were associated with an action taken by the patrol, which predominately involved either an arrest, a verbal warning or an unsuccessful pursuit, or, occasionally, no action. Whether action was taken or not observations were generally associated with either a threat and/or infraction type (e.g., gun hunting, trapping, farming, or illegal entry). Action, threat and infraction categories varied widely between sites and were not completed by rangers consistently within sites, so I aggregated all direct observations of people.

I explored the effects of two different methods for aggregating catch data: (1) a count of unique 1 km grid cells in which any observations of the activity type were recorded within the analysis area (henceforth, occupancy method); and (2) a count of all records of observations of the activity type within the analysis area (henceforth, count method). For the patrol effort measure, I used percentage of core patrol area covered by ground-based ranger patrols per timestep (henceforth, spatial coverage). I defined core patrol area as the intersection between (1) a minimum convex polygon surrounding 99% of patrol position records closest to the point's centroid and (2) each site's protected area boundary (i.e., I excluded 1% of points farthest from the centroid and those outside the boundary). I constructed a spatial grid of 1 km cells corresponding to the core patrol area and calculated spatial coverage as the percentage of unique cells through which patrol routes passed. I estimated patrol routes by assuming the shortest route between successive position records within unique patrol legs.

5.2.3 Analysis

I treated each site as one unit, aggregating data from across the entirety of a site's core patrol area. I used core patrol area as my area of analysis, rather than the broader protected area boundary, as in some contexts rangers only routinely patrol part of a site, providing little or no consistent monitoring outside this area. I included records from the entire time period of available monitoring data, except for the first 3 months following implementation of SMART. Excluding the first 3 months of data allowed a period to ensure all patrols were monitored using SMART, as it is common for patrol teams to be trained and equipped successively during implementation. I also removed the final timestep in each timeseries, which might only include partial data.

I explored the effects of three plausible time intervals over which to aggregate data: 14, 28 and 42 days. Studies examining patrol deterrence at annual scales have garnered mixed results (Beale *et al.*, 2018; Moore *et al.*, 2018). No analyses have assessed patrol deterrence over alternative intervals, yet in non-protected area contexts shorter timescales have been more effective for identifying deterrence effects (Chamlin *et al.*, 1992). Aggregating data over intervals of <14 days generated a high proportion of timesteps with zero observations. I generated 3 differenced plots for each time interval for each site with varying degrees of lag between patrol effort and CPUE response. For the default analysis, I assumed a time lag between cause and effect corresponding to one timestep (henceforth, '(t-1)' plots). That is, I assumed rule-breaker behaviour changed in response to change in patrol effort over the preceding timestep, and generated plots of differenced CPUE ($CPUE_t$-$CPUE_{t-1}$) against differenced effort ($effort_{t-1}$-$effort_{t-2}$). However, rule-breakers might make decisions based on recent information on patrol presence or based on accumulation of information from multiple prior time periods. Consequently, for the sensitivity analysis, I repeated the differenced plots, but compared change in CPUE to change in mean patrol effort calculated over a moving window of two and three timesteps starting with the same lag (henceforth, 'MA2' and 'MA3' plots, respectively, for 'Moving Average'). All data were processed and analysed within the R software environment (R Core Team, 2018).

5.3 Results

5.3.1 Summary

Results indicate that deterrence may have been operating in one of four sites for snares, and two of four sites for people. However, the relationships were weak, and were only apparent at some timesteps, the temporal resolution of which varied amongst sites. Two of the sites in which deterrence was apparent were savannah-dominated and one was forest-dominated. Evidence for deterrence was only apparent over the default time lag (t-1) and broke down over longer moving averages. Lastly, correlations were either not apparent or were weaker when using the count method for catch data instead of occupancy.

5.3.2 Snares

Differenced (t-1) plots of snare occupancy returned a significant negative correlation for one of four sites, over one of three timesteps (Table 5.2; Fig. 5.2). The single significant correlation was weak (r^2=0.14) and occurred over the medium timestep (28 days) in a savannah-dominated site (site three). The correlation was also apparent but weaker when using the count method for catch data instead of occupancy (r^2=0.12; Table 5.3, Fig. 5.3), and broke down when using lags with patrol effort calculated over a moving average of two and three timesteps (see supplementary material).

Table 5.2. Regression output for differenced (t-1) plots of snare occupancy, over timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites.

Site	Step	Slope	SE	*F*	DF	*P*	r^2
Forest-dominated sites							
	14	0.000	0.003	0.000	(1,36)		0.00
1	28	0.004	0.005	0.639	(1,17)		0.04
	42	-0.005	0.006	0.812	(1,10)		0.08
	14	0.010	0.005	3.147	(1,47)	.	0.06
2	28	-0.014	0.010	1.931	(1,21)		0.08
	42	-0.038	0.018	4.637	(1,13)	.	0.26
Savannah-dominated sites							
	14	0.000	0.004	0.005	(1,101)		0.00
3	28	-0.010	0.003	7.869	(1,49)	**	0.14
	42	0.006	0.004	2.663	(1,31)		0.08
	14	-0.006	0.023	0.075	(1,39)		0.00
4	28	-0.007	0.015	0.185	(1,17)		0.01
	42	0.018	0.018	1.005	(1,10)		0.09

Table 5.3. Regression output for differenced (t-1) plots of snare counts, over timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites.

Site	Step	Slope	SE	*F*	DF	*P*	r^2
Forest-dominated sites							
	14	-0.013	0.019	0.476	(1,36)		0.01
1	28	-0.016	0.043	0.142	(1,17)		0.01
	42	0.024	0.036	0.446	(1,10)		0.04
	14	0.022	0.013	2.948	(1,47)	.	0.06
2	28	-0.036	0.032	1.250	(1,21)		0.06
	42	-0.108	0.054	3.901	(1,13)	.	0.23
Savannah-dominated sites							
	14	0.002	0.004	0.184	(1,101)		0.00
3	28	-0.010	0.004	6.931	(1,49)	*	0.12
	42	0.008	0.004	3.095	(1,31)	.	0.09
	14	0.029	0.060	0.232	(1,39)		0.01
4	28	-0.017	0.052	0.104	(1,17)		0.01
	42	0.086	0.057	2.253	(1,10)		0.18

5.3.3 Direct observations of people

Differenced (t-1) plots of people occupancy returned significant negative correlations for two of four sites (one forest-dominated site and one savannah-dominated site), over one of three timesteps in both instances (Table 5.4; Fig. 5.4). The significant correlation was over the long timestep (42 days) in the forest-dominated site (site two) and the short timestep (14 days) in the savannah-dominated site (site four). Both relationships were weak (r^2=0.30 and 0.15, respectively) and broke down when using the count method for catch data instead of occupancy (Table 5.5, Fig. 5.5). Both correlations were also only apparent over the default time lag (t-1) and broke down when using lags with patrol effort calculated over a moving average of two and three timesteps (see supplementary material).

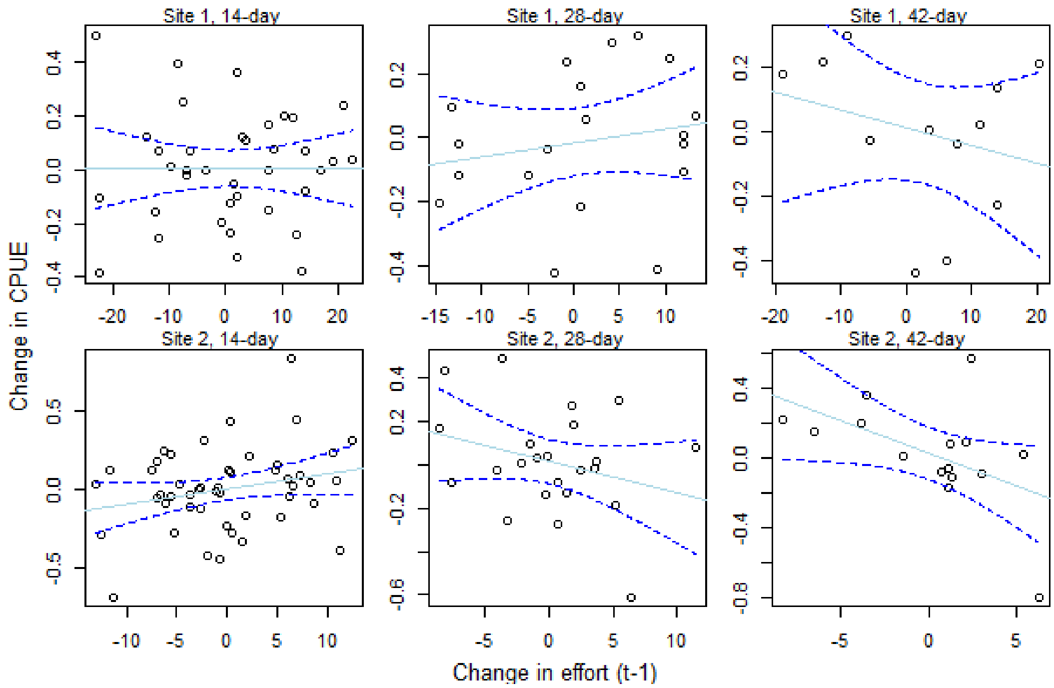
Table 5.4. Regression output for differenced (t-1) plots of people occupancy, with timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites.

Site	Step	Slope	SE	*F*	DF	*P*	r^2
Forest-dominated sites							
	14	0.001	0.002	0.372	(1,36)		0.01
1	28	0.001	0.002	0.061	(1,17)		0.00
	42	0.001	0.003	0.046	(1,10)		0.00
	14	-0.005	0.008	0.461	(1,47)		0.01
2	28	-0.008	0.020	0.142	(1,21)		0.01
	42	-0.064	0.027	5.683	(1,13)	*	0.30
Savannah-dominated sites							
	14	0.001	0.004	0.148	(1,101)		0.00
3	28	0.002	0.002	0.499	(1,49)		0.01
	42	0.003	0.003	0.715	(1,31)		0.02
	14	-0.039	0.015	6.897	(1,39)	*	0.15
4	28	-0.003	0.009	0.086	(1,17)		0.01
	42	-0.001	0.011	0.012	(1,10)		0.00

Table 5.5. Regression output for differenced (t-1) plots of people counts, with timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites.

Site	Step	Slope	SE	*F*	DF	*P*	r^2
Forest-dominated sites							
	14	0.001	0.003	0.246	(1,36)		0.01
1	28	0.000	0.003	0.005	(1,17)		0.00
	42	0.000	0.004	0.008	(1,10)		0.00
	14	-0.009	0.014	0.379	(1,47)		0.01
2	28	-0.028	0.033	0.732	(1,21)		0.03
	42	-0.122	0.059	4.273	(1,13)	.	0.25
Savannah-dominated sites							
	14	0.000	0.005	0.004	(1,101)		0.00
3	28	0.004	0.002	2.016	(1,49)		0.04
	42	0.005	0.004	2.112	(1,31)		0.06
	14	-0.034	0.022	2.411	(1,39)		0.06
4	28	0.003	0.015	0.046	(1,17)		0.00
	42	-0.003	0.016	0.032	(1,10)		0.00

Forest-dominated sites



Savannah-dominated sites

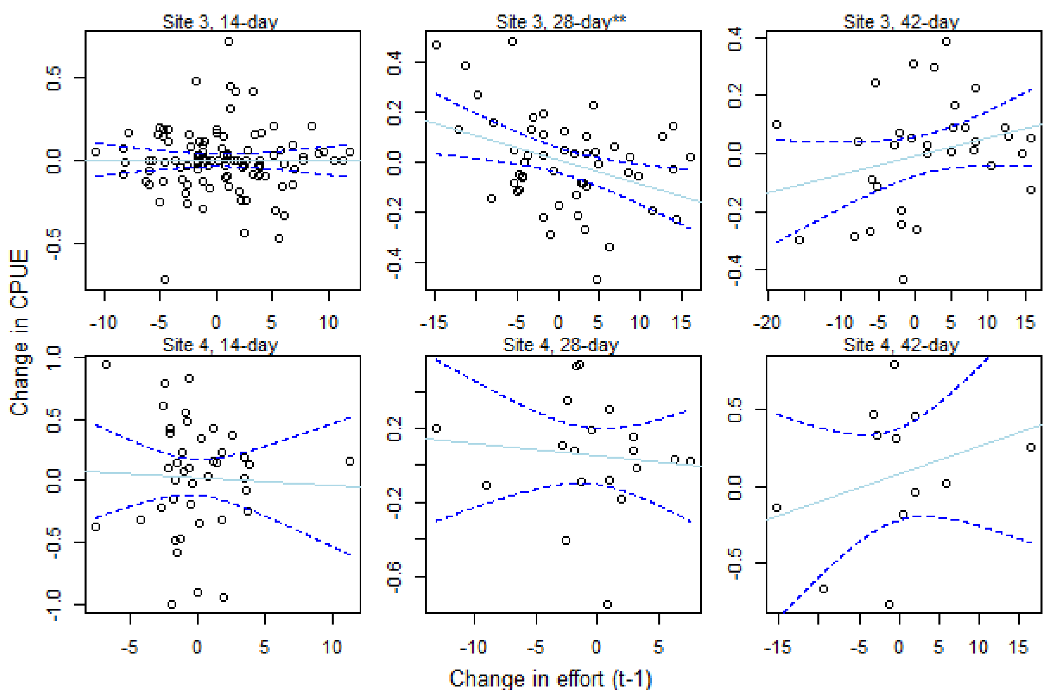
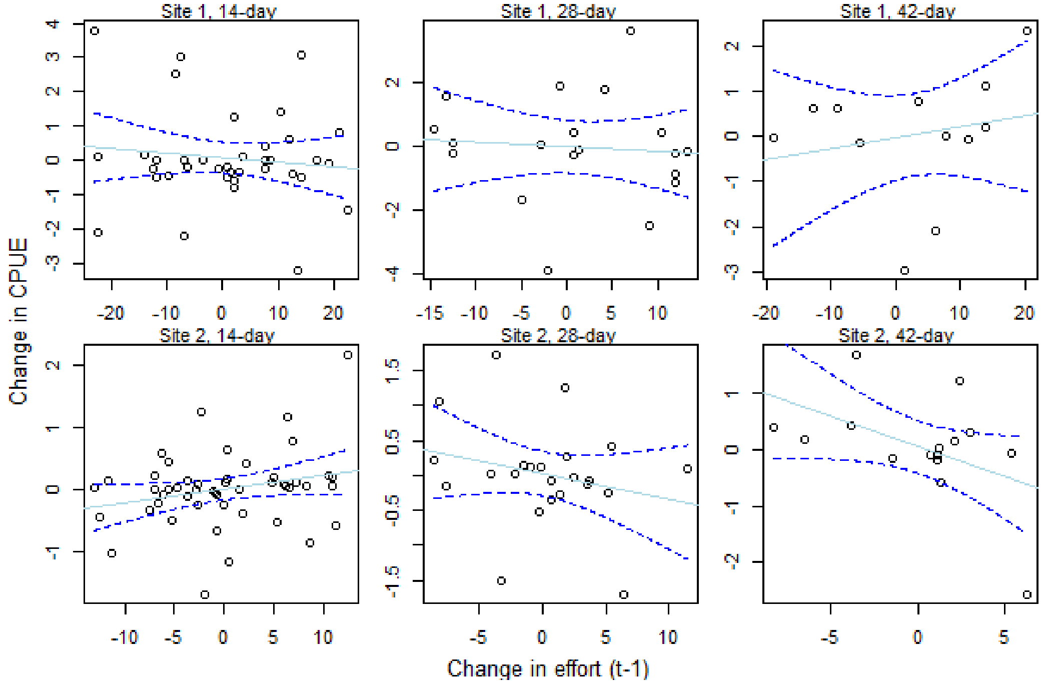


Figure 5.2. Differenced (t-1) plots of snare occupancy, over timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites. Significance at .01 level = ** and .05 level = *. Light blue line = linear regression line. Dotted blue lines = 95% confidence intervals.

Forest-dominated sites



Savannah-dominated sites

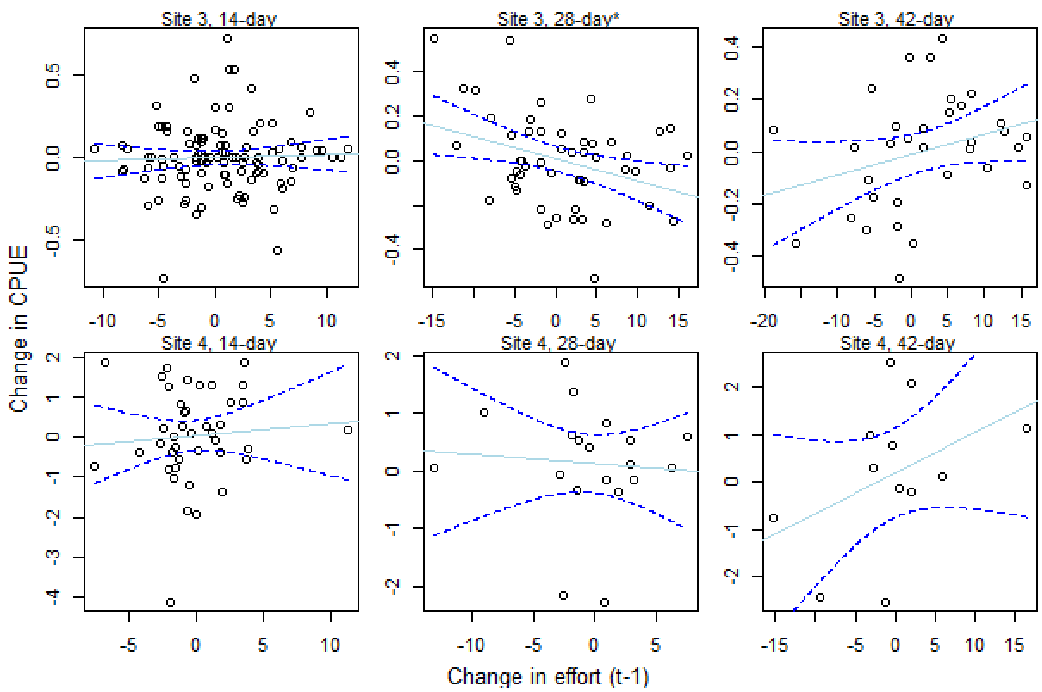
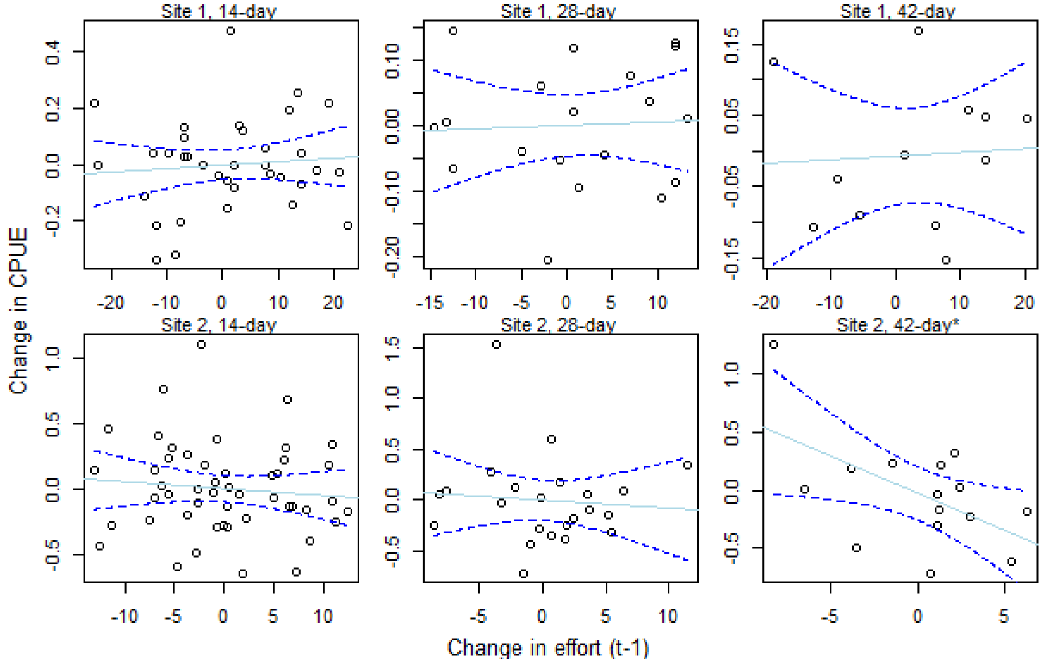


Figure 5.3. Differenced (t-1) plots of snare counts, over timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites. Significance at .01 level = ** and .05 level = *. Light blue line = linear regression line. Dotted blue lines = 95% confidence intervals.

Forest-dominated sites



Savannah-dominated sites

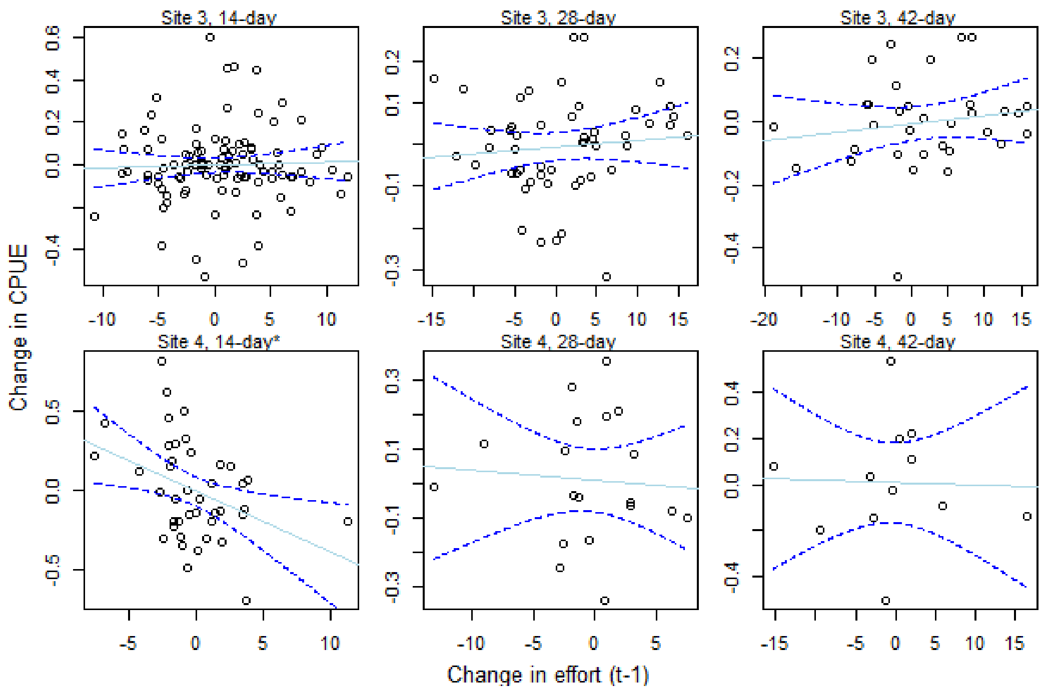
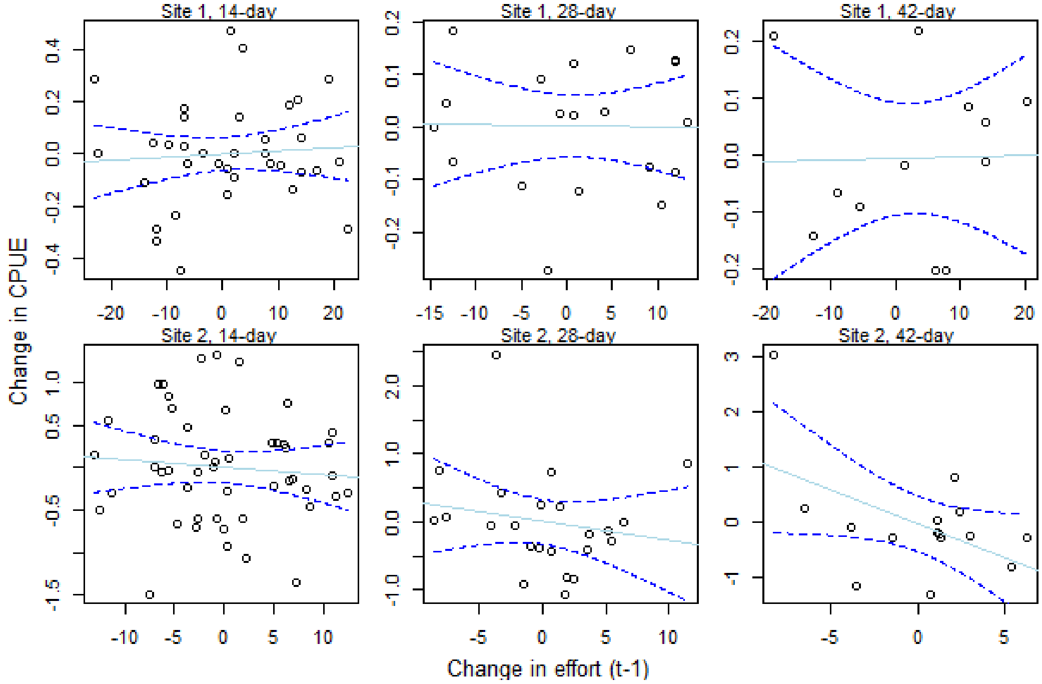


Figure 5.4. Differenced (t-1) plots of people occupancy, over timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites. Significance at .01 level = ** and .05 level = *. Light blue line = linear regression line. Dotted blue lines = 95% confidence intervals.

Forest-dominated sites



Savannah-dominated sites

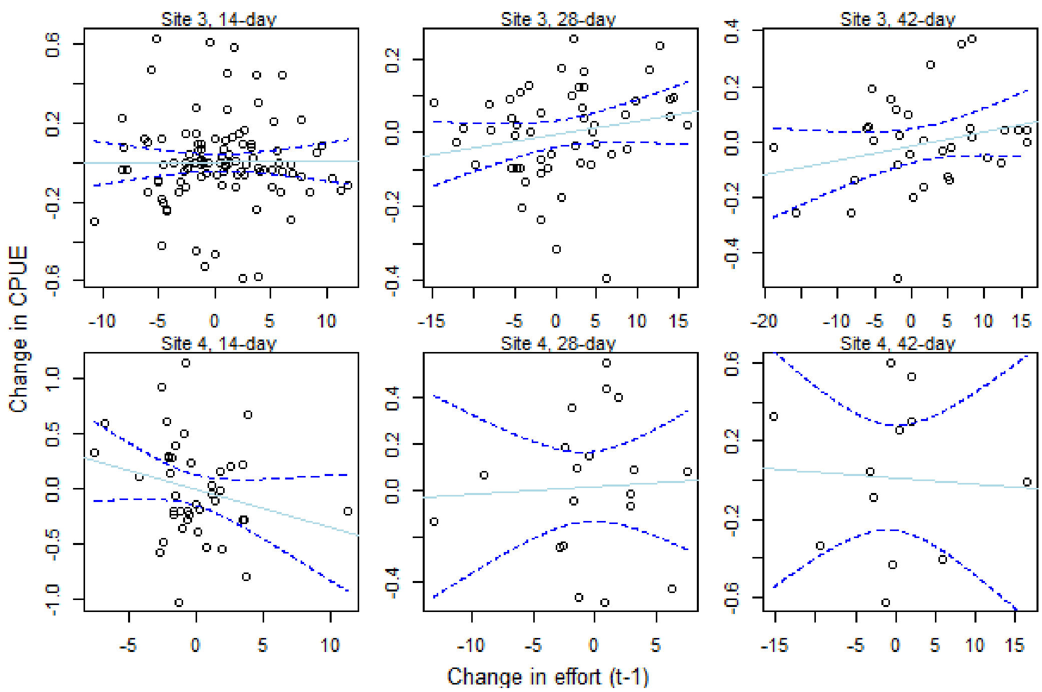


Figure 5.5. Differenced (t-1) plots of people counts, over timesteps of 14, 28 and 42 days, for two forest-dominated and two savannah-dominated sites. Significance at .01 level = ** and .05 level = *. Light blue line = linear regression line. Dotted blue lines = 95% confidence intervals.

5.4 Discussion

5.4.1 Evidence for ranger-led enforcement patrols deterring illegal activity in protected areas

Ranger-led law enforcement patrols are the primary means by which managers aim to deter illegal activity in protected areas, but evidence of deterrence operating in practice is lacking. Here, I applied differenced plots, a recent advancement in methods for identifying deterrence, to real patrol data for the first time. My results suggest that deterrence may have been operating in one of four sites for snares, and two of four sites for people, but the relationships were weak and were only apparent at some timesteps, the temporal resolution of which varied amongst sites. Very few empirical studies have demonstrated deterrence in protected areas rigorously, with Moore *et al.* (2018) a notable exception. However, similarly sophisticated analyses have failed to find an effect of patrol presence on poaching (Beale *et al.*, 2018). My results are in line with existing research: deterrence may be operating in protected areas, but the evidence remains weak. The absence of strong evidence of deterrence in sites with relatively high patrol effort implies that the deterrence effect may be weaker and harder to identify in protected areas in general. In contrast, studies of the effects of increasing police presence on crime have found evidence of deterrence in a variety of contexts (Nagin, 2013a).

The absence of consistent evidence of deterrence across sites precluded a comprehensive analysis of factors associated with stronger or weaker deterrence from patrols. However, these results are in line with my hypothesis that deterrence associated with illegal activity types which persist in the landscape, such as snaring, will be harder to detect from patrol data than those which are ephemeral, such as people, because changes in effort may have impacts beyond the consecutive time-step (Dobson *et al.*, 2018). Results also suggest that deterrence may have been operating in one savannah-dominated site for snaring and in one forest- and one savannah-dominated site for people. These findings are broadly in line with my hypothesis that I would find greater evidence of deterrence in sites dominated by open habitat, suggesting patrols may be more effectively deployed to deter illegal activity in protected areas such as these.

5.4.2 Implications for deterrence

In general, I did not find consistent evidence of deterrence. Lack of evidence could indicate that (1) patrol presence does not reliably deter illegal activity, or that the effect is often too weak to detect, or (2) patrol presence does deter illegal activity, but the methods I have used are not able to identify the effect. Deterrence could, indeed, be rare. In theory, risk of punishment following detection by rangers should inhibit the criminal behaviour of potential rule-breakers (Pratt and Cullen, 2005). In practice, these risks may be too low to achieve deterrence in many protected areas. We selected SMART-monitored sites with relatively high patrol effort, but patrols still provided little presence in space and time, providing ample opportunity for rule-breakers to go undetected. In contexts with much higher patrol effort (e.g., police presence in cities in the global north) deterrence can be difficult to demonstrate (Ratcliffe *et al.*, 2011). Even when individuals are detected, risk of punishment may still be low. Rangers may lack incentives required to arrest rule-breakers (Ogunjinmi, Umunna and Ogunjinmi, 2009) or, where arrests occur, justice systems may not punish offenders sufficiently severely (Moreto and Gau, 2017). Compliance with rules may also depend on individuals' perceptions of legitimacy and fairness, which may be undermined in countries where authorities are complicit in illegal activity (Kahler and Gore, 2012). If deterrence is rare or weak then ranger patrols may not represent a wise investment for many protected areas, and resources might be more effectively deployed elsewhere (e.g., away from preventative patrols and towards intelligence-led policing (Moreto, 2015)).

Alternatively, it is also possible that ranger patrols do deter illegal activity, but the approach I used was not sufficiently sensitive. I chose to aggregate effort and catch data at broad spatial scales (i.e., across the entire core patrol area) and explored relatively short temporal scales (14-, 28- and 42-day intervals). In reality, deterrence may operate over different time intervals. For example, potential rule-breakers may consider levels of patrolling across the previous year and make decisions about their actions in the present year accordingly. However, studies examining deterrence at annual scales have garnered mixed results (Beale *et al.*, 2018; Moore *et al.*, 2018). Alternatively, deterrence may operate at very short temporal scales, with decisions about immediate actions made based upon recent events (e.g., within hours or a few days). Deterrence could also operate over multiple temporal scales simultaneously. I also assume similar timescales for cause and effect, although this may not be true (e.g., the effects of a long-term increase in patrol effort may only persist for a short time). Similarly, it is also possible that a form of deterrence takes place at spatial scales finer than I have assessed. My application of differenced plots can only detect deterrence if patrol presence leads to an overall reduction in illegal activity throughout a monitored area. Yet deterrence could be operating at finer scales, but with displacement within that area, such that the overall frequency of activities remains unchanged. The ultimate aim of patrolling is to deter illegal activity within a protected area and aggregating across broad scales assesses whether this goal has been achieved. Nevertheless, if deterrence with displacement is operating, this may have implications for how patrols should be resourced and deployed. For example, patrolling may be warranted but only if broad consistent presence, which precludes internal displacement, is achievable.

Regardless of questions of scale, it is possible that the metric I used, how I applied it, or the data I applied it to, are not able to reliably detect deterrence. Firstly, lack of a clear negative correlation in differenced plots does not confirm lack of deterrence. Deterrence may not operate in a continuous fashion, with a greater reduction in illegal activity depending on the magnitude of change in effort, but binarily, being present where there is any quantity of patrolling and absent where there is none, which would result in a flat plot (Dobson *et al.*, 2018). Secondly, my measure of patrol effort might be a poor proxy for aspects of patrol presence which are important for deterrence. I focused exclusively on patrol effort inside site boundaries. Any effort immediately outside the boundary would not be included but could deter entry and thereby reduce illegal activity. Nearly all patrol effort in the study sites occurred within boundaries, but in other protected areas patrol effort may be directed to the outskirts of parks to deter ingress. Moreover, my measure of effort was relatively crude, capturing only the percentage of unique grid cells visited and ignoring aspects such as levels of activity within and across visited cells. For example, a unique visited cell could signify one or multiple visits, and either a brief visit or extended patrolling throughout the cell. In theory, these aspects of patrol activity might be important for deterrence but are ignored in this analysis. In practice, the majority of patrol activity is often heavily skewed towards common areas, such as near patrol posts (Plumptre *et al.*, 2014). Consequently, and as I was considering relatively short time periods, I considered spatial coverage appropriate, but research at finer scales should consider within cell activity patterns (e.g., distance travelled).

My indicators of illegal activity might also be a poor proxy for the abundance of crimes. The number of snares or people recorded by rangers (or unique grid cells in which either occurred) may be only weakly related to the illegal behaviour of potential rule-breakers. Persistence is a particular issue for snares, which may have been laid down at any point over several months prior to detection (Coad, 2007). Direct observations of people, whilst not subject to issues of persistence, may not necessarily be associated with a crime. The IUCN categories of the assessed sites suggest park entry was not strictly illegal, so individuals encountered may not have intended to commit a crime. Nevertheless, observations of people were often associated with either a patrol action or an infraction categorisation across sites, suggesting patrols aimed to discourage their presence. Recording of whether direct observations of people were associated with actions or infractions was too inconsistent in the sites assessed to delineate only those people whose presence was clearly illegal, but future deterrence studies should take this factor into account. In general, illegal activity data collected by rangers on patrol and derived CPUE indices are subject to numerous uncontrolled biases, including non-random distribution of sampling effort, imperfect detection, inaccurate reporting, variable catchability and nonlinear relationships between abundance and CPUE (Keane *et al.*, 2011). Consequently, apparent changes in the level of illegal activity indicated by CPUE measures may reflect sampling biases rather than underlying trends. I selected sites with relatively high patrol coverage to attempt to reduce issues associated with patchy or inconsistent sampling effort, but other biases are less easily controlled. Methods have recently been developed to account for biases in surveillance effort using hierarchical models and applied to patrol data (Beale *et al.*, 2014; Critchlow *et al.*, 2015) but their application is non-trivial, particularly for protected area managers who may lack the necessary capacity.

5.4.3 Implications for application of differenced plots to diagnose deterrence in practice

Application of differenced plots to real patrol data necessitates multiple decisions regarding the spatial and temporal scales over which we assume deterrence operates, and selection of appropriate measures for summarizing patrol effort and illegal activity data (Dobson *et al.*, 2018). However, at present, understanding of the mechanisms underlying deterrence is too poor to provide clear guidance. Here, I explored methods for applying differenced plots in practice, including plausible time intervals and lags, and approaches for aggregating catch data. The occurrence of evidence for deterrence over only specific choices of temporal resolution within sites suggests that selection of an appropriate time interval is crucial for identification of deterrence. Moreover, as no one resolution was favoured across sites, results suggest that the temporal scale of deterrence may be context dependent, suggesting that exploration of a small set of plausible timesteps is strongly advisable. Deterrence studies in the world of policing have come to similar conclusions (Chamlin *et al.*, 1992). Conversely, all correlations were only apparent over the default time lag (t-1) and broke down over longer moving averages, suggesting future studies should focus on the former. Deterrence could still operate by rule-breakers making decisions based on accumulation of information from multiple prior time periods, but the relationships may be too weak to withstand the effects of averaging patrol effort. Results also suggest that aggregating illegal activity catch data using a unique grid cell occupancy approach may be more effective for identifying deterrence than a simple observation count method, for analyses conducted at broad spatial scales using spatial coverage as an effort measure. Very high observation counts for a few timesteps, perhaps representing strong clustering of observations in the landscape, may have blurred any relationships when using the count method. Direct counts may be more effective at finer scales and when using an effort measure which captures variation in effort within cells.

5.4.4 *Conclusions & Recommendations*

My results lend some weak support to the conjecture that presence of ranger patrols may deter illegal activity in certain contexts, suggesting that investment in patrolling may be warranted. However, these contexts may not be common, the effect was weak and inconsistent, and it is still unclear under what conditions patrols are effective. Moreover, even if deterrence is operating in the study sites, pressure on biodiversity may not be reduced if illegal activity is displaced into habitat surrounding the protected area or if rule-breakers switch hunting methods (e.g., from snaring to gun hunting). Understanding whether and how conservations activities are effective is essential to ensure limited resources are efficiently deployed (Ferraro and Pattanayak, 2006). To this end, differenced plots are a useful metric for diagnosing deterrence in patrol data but should be applied and interpreted with care. Importantly, understanding of the spatial and temporal scales over which deterrence operates is still too poor to definitively conclude that lack of evidence constitutes lack of deterrence. My findings provide guidance to inform future application, but questions remain for how patrol data should be aggregated, and uncontrolled biases in patrol data may still exist. More fundamental research is needed to determine whether and how patrols do deter illegal activity, the associated costs and benefits, and to establish how to make reliable use of patrol data. Ultimately, this will require in-depth study of systems in which deterrence is likely, with independent data on illegal activity, and in which some confounding factors are accounted for.

Chapter 6

Discussion

6.1 Background

Law enforcement – policing, prosecution, and punishment of crime – is the dominant paradigm for addressing the threat of illegal use of natural resources in protected areas globally (Moreto and Gau, 2017). In the literature, debates proliferate over whether a plurality of approaches may be necessary to reduce illegal activity (Cooney *et al.*, 2017). However, in practice, law enforcement, operationalised via ranger-led patrolling, remains the dominant mode, and with good reason: in contexts where it has been evaluated, protected areas with enforcement present display better conservation outcomes (Bruner *et al.*, 2001; Tranquilli *et al.*, 2012). Yet illegal activity is still common in sites where rangers are present (e.g., in Brazil's protected area network (Kauano *et al.*, 2017)), and driving declines in biodiversity.

There is an urgent need to develop a better understanding of how ranger patrolling is implemented and operates to reduce crime, in order to understand why it is failing and how it can be improved. Yet the theory and practice of patrolling has hitherto received relatively little research attention. For example, the extent of patrolling undertaken in protected areas is largely unknown, despite fears that enforcement activity may be generally inadequate (Leverington *et al.*, 2010; Plumptre *et al.*, 2014). Moreover, there is limited understanding of the theoretical mechanisms through which patrols operate to reduce illegal activity (Dobson *et al.*, 2018). In this thesis, I set out to address this paucity of knowledge. In so doing I have improved the evidence base underpinning the use of ranger-led law enforcement patrols in protected areas. I discuss these contributions, limitations in methods employed, and avenues for future research in section 6.2.

Concurrently, there is an equally pressing need to evaluate interventions implemented to improve patrol effectiveness. Conservation, in general, has been hindered by a paucity of rigorous yet flexible methods for evaluating interventions. In this thesis, I set out to address these issues by drawing on methods from other policy arenas to evaluate whether and how a popular technological intervention aiming to improve patrol effectiveness, SMART, contributed to reduced illegal activity. In so doing I have made contributions in two areas. Firstly, I developed a theory of change for SMART and produced evidence for verifying successive aspects of this theory. Secondly, I advanced conservation evaluation by developing and demonstrating the application of pragmatic, formative methods and designs. In section 6.2 I discuss and synthesise evidence, reassess the strength of SMART's contribution claim, and reflect on the utility of the approaches applied.

6.2 Ranger-led law enforcement patrols in theory and practice

An important contribution made by this thesis has been to deepen understanding of the theory and practice of patrolling as a means to reduce illegal nature resource use in in protected areas. Patrol research and management has been hindered by historically inadequate monitoring in protected areas. Ranger-based monitoring – the collection of data by rangers on patrol – has filled this gap (Gray and Kalpers, 2005). The approach has been embraced by protected area managers, who need information on focal species, wildlife crime and patrol activity to inform decision-making. In recent years, the popularity of ranger-based monitoring has grown rapidly, stimulated by the development of Law Enforcement Monitoring (LEM) tools: technology-enabled systems which facilitate standardised collection, management, and analysis of ranger-based monitoring data (Stokes, 2010). I exploited widespread, standardised monitoring of illegal activity and patrol activity, via a popular LEM tool, to assemble data from a broad sample of sites globally. I used these data to advance knowledge in two key areas, described below.

6.2.1 Patrolling in practice

Broad, regular patrol presence throughout protected areas is thought to be key to reducing illegal activity (Singh *et al.*, 2015). Yet the extent of patrolling undertaken in protected areas is poorly understood, despite concerns that presence may be generally low and potentially inadequate. In chapter 4, I addressed this gap by conducting the first global analysis of trends in spatiotemporal patrol presence. I estimated spatial and temporal coverage provided by ranger patrols within and across sites and over time. I also evaluated results with respect to industry benchmarks and assessed factors influencing patrol presence. In so doing, I developed two key findings.

Firstly, I found that patrol presence was poor in the majority of protected areas around the world. Specifically, patrols typically provided very low spatial coverage at monthly scales, and low coverage at annual scales. Accordingly, almost all sites fell short of spatial coverage targets derived from industry benchmarks (Singh *et al.*, 2015). These results are in line with existing evidence (Plumptre *et al.*, 2014; Hötte *et al.*, 2016), but my contribution is to demonstrate that limited patrol presence is widespread throughout protected areas globally.

This finding sheds doubt on the efficacy of patrolling, as it is currently practiced in the majority of sites, as a means to reduce illegal activity. Patrolling is assumed to reduce illegal activity through detection and deterrence of crime (Nagin, 2013b). In protected area contexts, these mechanisms are poorly understood, yet the levels of patrolling observed are unlikely to be sufficient for either to be achieved. The protected areas I assessed were generally large (mean=2,789 km^2) and often contained dense habitats. These factors, in combination with infrequent, low coverage patrols, provide ample space, time and refugia for offenders to avoid detection. Even snares, which persist for weeks or months in the landscape, are difficult to detect (O'Kelly *et al.*, 2018). Consequently, high levels of survey effort will be required to reduce their threat. The failure of patrolling to reduce the threat of snaring in South East Asia suggests effort levels are insufficient (Gray *et al.*, 2017). The deterrence mechanism of ranger patrols is less well understood, but in contexts with far higher patrolling levels (e.g., city policing) the evidence that patrols deter crime is mixed (Nagin, 2013a).

This finding also sheds doubt on the efficacy of ranger-based monitoring as a robust method for monitoring wildlife or illegal activity. Ranger-based monitoring generally and SMART specifically are increasingly touted and implemented as a solution to data paucity in protected areas, and resultant data are used to predict spatiotemporal trends in illegal activity and inform management decisions (Critchlow *et al.*, 2015, 2017). Yet ranger-based monitoring is only as robust as the sampling design employed (Keane *et al.*, 2011). My contribution is to demonstrate that in most sites survey effort may be too narrow and inconsistent to generate reliable trends, at least using conventional methods which are feasible for most sites. These results suggest that ranger-based monitoring, via SMART, as commonly practiced, is not a robust method for monitoring wildlife or illegal activity. Indeed, SMART should only be used for this purpose to the extent that patrolling is able to mimic robust sampling designs (Keane *et al.*, 2011). However, doing so may detract from patrols' primary objective of law enforcement.

Secondly, I found that patrol presence, at least over monthly timescales, was constrained by budgets for law enforcement. Moreover, my results suggest that managers may be circumventing this limitation by varying patrol locations throughout the year to achieve broader spatial coverage over longer annual timescales, at the expense of regular ranger presence in any one area. That patrolling is limited by funding available for the activity is not surprising. Shortfalls in financial support limit protected area effectiveness, in general (Watson *et al.*, 2014). Yet my analysis represents the first investigation of factors influencing patrolling and provides clear evidence, which indicates how levels of patrolling might be improved – by increasing financial support to protected areas. Employing tactics to achieve coverage goals over long timeframes, such as varying deployment, cannot make up for the lack of consistent patrol broad presence.

Future research could build on these contributions in a number of constructive ways. Firstly, coverage is a relatively crude measure of patrol presence. I chose this measure because it is commonly used for monitoring and evaluation of patrols. Yet a more representative and relevant measure would better account for variation in levels and distribution of spatial and temporal activity through time at finer scales. Secondly, my analysis of the relationship between patrol presence and budgets for law enforcement was correlative, not causative. It remains to be seen whether and how changes in budgets within sites translate into changes in presence.

More broadly, I evaluated presence using industry targets for patrol coverage (Singh *et al.*, 2015), but these targets are not based on empirical evidence. Indeed, there is limited understanding of relationships between ranger patrolling and crime. Consequently, whilst presence was low, at present it is unclear what adequate levels of patrolling should be, and if achieving such levels is cost-effective relative to alternative approaches. Relatedly, while ranger-based monitoring should only be used to the extent that patrolling is able to mimic robust sampling designs, it is unclear whether doing so would detract from patrols' primary purpose of law enforcement. A major hindrance to better understanding of the relationship between patrols and illegal activity is limited appreciation of the mechanisms involved. It is essential to develop a deeper understanding of the theory underlying how patrols are assumed to work, in order to be able to provide clear guidance for how patrolling can be improved.

6.2.2 Patrolling in theory

In chapter 5, I addressed this gap by investigating whether and in what contexts patrols effectively deter crime, using ranger-based monitoring data. Deterrence is the principle mechanism through which patrols are assumed to operate, but is rarely demonstrated in practice, because the most common source of data on illegal activity in protected areas – ranger-based monitoring – is particularly vulnerable to bias (Keane *et al.*, 2011). Common metrics of deterrence that use ranger-based monitoring data fail to account for these biases. I applied a novel metric of deterrence ('differenced plots' (Dobson *et al.*, 2018)), which is designed to be more robust to confounding temporal effects and autocorrelation, and that had reliably identified deterrence using simulated data, to real patrol data for the first time. I used difference plots to assess whether I find evidence that patrolling deters illegal activity in four diverse protected areas, with relatively high levels of patrol effort. I used two different indicators of illegal activity: snares and direct observations of people. These activities occupy alternate ends of the persistence spectrum, with snares remaining detectable in the landscape for weeks or months (Coad, 2007), while encounters with people are much more ephemeral. I also examined whether differences in deterrence can be explained by habitat type and explored methods for applying differenced plots to real data, including the effect of temporal scales. In so doing, I made a number of key contributions to this understudied area.

Firstly, I found that deterrence may have been operating in one of four sites for snares, and two out of four sites for direct observations of people, but the relationships were weak and were not consistent across multiple temporal resolutions within the same site. Moreover, I found that deterrence may have been operating in one savannah-dominated site for snaring and in one forest- and one savannah-dominated site for people. These results are in line with existing research: deterrence may be operating in protected areas, but the evidence remains weak (e.g., Beale *et al.* (2018)). The absence of strong evidence of deterrence in sites with relatively high patrol effort implies that patrols in sites with lower patrol effort may be even less effective deterrents. As low patrol presence is the norm in protected areas (chapter 4), patrolling in general may rarely deter crime. If deterrence is truly rare, my analysis suggests limited resources may be more effectively directed towards other activities (e.g., intelligence-led policing, or non-enforcement approaches).

Secondly, whilst the absence of consistent evidence of deterrence across sites precluded a comprehensive analysis of factors associated with stronger or weaker deterrence, these results were also in line with my hypotheses that: (1) illegal activity types which persist in the landscape, such as snaring, will display weaker relationships than those which are ephemeral, such as people; and (2) I would find greater evidence of deterrence in sites dominated by open habitat, such as savannah, than in closed habitat, such as forests. The latter suggests patrol may be most effective for deterring people in open habitat sites. The former suggests that differenced plots will be more able to detect deterrence if the illegal activity types are either ephemeral, or if the age of signs can be reliably estimated.

Lack of evidence for deterrence could indicate that patrols do not reliably deter illegal activity, or that the approach I used was not sufficiently sensitive. At present, understanding of patrol deterrence is too poor to provide guidance on the temporal or spatial scales over which the mechanism operates, or which aspects of patrolling are important. To advance knowledge in this area I explored the effect of a range of plausible time intervals and lags and developed three contributions: (1) I found evidence for deterrence over only individual temporal resolutions within sites, suggesting that selection of an appropriate time interval is crucial for identification of deterrence; (2) no one resolution was favoured across sites, suggesting that exploration of multiple timesteps is warranted; (3) deterrence was only apparent over the default time lag (t-1), suggesting deterrence is unlikely to be identified over longer lags. These findings can be used to guide future research into deterrence, yet questions remain for appropriate spatial scales, which may be finer than I assessed. Finally, I explored methods for aggregating illegal activity data. My findings suggest that aggregating data using a unique grid cell occupancy approach may be more effective for identifying deterrence than a simple observation count method, at least for analyses conducted at broad spatial scales which use spatial coverage as a patrol effort measure. Direct counts may be more effective at finer scales and when using an effort measure which captures variation in effort within cells.

Lastly, in chapters 4 and 5 I assessed one aspect of patrol presence – coverage of protected areas – but multiple other aspects of patrol activity or law enforcement in protected areas may be important for deterring crime and should be explored in future analyses. For example, presence around park borders may deter ingress by offenders, or offenders may be mindful of the predictable presence of patrols in space and time, in which case predictability of patrolling may be as important as activity levels. Finally, patrols will only deter crime if potential offenders perceive that the risk of punishment is sufficiently certain, swift and severe (Nagin, 2013b). I am aware of no studies that have investigated relationships between punishment and offending in protected area contexts, but such analyses may provide insights as to why deterrence may be weak and might be strengthened.

My analysis has shown that differenced plots are a useful metric for diagnosing deterrence in ranger-based monitoring data but should be applied and interpreted with care. Whilst differenced plots may account for temporal biases in ranger-based monitoring data, multiple uncontrolled biases, such as non-random sampling effort and imperfect detection are still uncontrolled (Keane *et al.*, 2011). I selected sites with high patrol effort to minimise issues associated with patchy sampling effort, but other biases are less easily controlled. Sampling effort of ranger patrols in protected areas is generally lower and less consistent than the sites I assessed (chapter 4) suggesting differenced plots will not be an appropriate deterrence metric for ranger-based monitoring data gathered in the majority of sites. Ranger-based monitoring and differenced plots are a useful source of data for exploring deterrence in practice, and alternative methods do exist for controlling sampling biases in ranger-collected data which could usefully be applied to extend the approach (e.g., Critchlow *et al.* (2015)). Nevertheless, there is an urgent need for better and more detailed investigation of the spatiotemporal relationship between patrols and illegal activity, to inform management. This will require rigorous, in-depth analysis in controlled circumstances and illegal activity data collected independently of ranger-patrols. Future analyses should also examine the extent to which ranger-based monitoring can be achieved rigorously and concurrently to law enforcement.

In conclusion, more fundamental research drawing on robust data is needed to determine whether, how and in what contexts patrols can effectively deter crime, and if achieving deterrence can be achieved economically. If it is not, there is an urgent need to develop a better understanding of alternative, cost-effective methods for reducing illegal activity in protected areas.

6.3 Improving patrol effectiveness: Evaluation of SMART

To attempt to improve patrol effectiveness, tools which use ranger-based monitoring data to inform adaptive patrol management, such as SMART (SMART Partnership, 2018), are increasingly implemented in protected areas globally. Yet, despite substantial resources expended in development and implementation, there has been no rigorous evaluation. In this thesis, I evaluated SMART, drawing on a mixture of flexible, formative approaches from policy arenas outside conservation. In so doing, I developed evidence which can be used to judge whether SMART represents a wise allocation of resources, to improve the intervention, and to inform decisions about future deployment. I guided aspects of the evaluation using contribution analysis, a structured, theory-based design for assessing whether an intervention contributed to observed impacts by verifying a theory of change with empirical evidence (Mayne, 2012). Using contribution analysis, causality is inferred if: (1) assumptions underlying the theory of change are plausible and agreed upon by stakeholders, (2) the intervention was implemented as outlined in the theory of change, (3) the chain of expected results occurred, and (4) the relative contribution of external factors was recognised (Mayne, 2012; Delahais and Toulemonde, 2017). In this section, I discuss evidence developed in the course of the evaluation, and whether this evidence supports the case for SMART having contributed to a reduction in illegal activity.

6.3.1 Theory of change

Impact evaluation of SMART is hindered by limited formal understanding of what implementation involves and the mechanisms through which it effects change. This information is essential to guide and interpret causal attribution. In chapter 2, I illustrated how SMART aims to reduce illegal activity via a chain of expected results, using a theory of change approach (Vogel, 2012), in collaboration with the intervention's developers (Fig 2.1). I also interrogated the construct, by articulating causal assumptions underlying SMART's Theory of Change, and assessing the strength of extant evidence supporting these assumptions. Construction and interrogation of SMART's theory of change revealed weak or lacking evidence for certain aspects, shedding doubt on SMART's contribution claim. In subsequent chapters I investigated these aspects to generate new evidence and reassess the claim.

Firstly, I found that implementation of SMART is complex, highly adaptable and success may be influenced by multiple external factors. Consequently, implementation will vary between participating sites and influence the extent to which outcomes are achieved, suggesting concurrent evaluation of how the intervention is delivered in practice is essential. In chapter 3 I undertook such an evaluation and, in chapter 4, explored the effects of variability in implementation on intermediate outcomes. Secondly, I found that SMART acts via a series of changes to the behaviour of rangers, managers, and offenders. As a result, illegal activity may respond slowly to implementation and effects may be subtle (Howe and Milner-Gulland, 2012). Consequently, impact evaluation should examine whether SMART contributed to incremental change in successive intermediate outcomes, such as effects of implementation on patrol performance (e.g., patrol coverage), and subsequent effects of patrol performance on deterrence of illegal activity. I explored the former in chapter 4. Thirdly, a review of supporting evidence identified potential weak links in the theory of change, which warranted further research. In particular, there was mixed evidence that increases in patrol presence derived from SMART implementation can reliably deter illegal activity. I explored this aspect in chapter 5. Finally, the review provided low confidence that ranger-based monitoring tools, such as SMART, can provide reliable information on threats in protected areas, particularly if patrols provide inconsistent or patchy patrol effort.

6.3.2 Implementation evaluation

In chapter 3 I presented a novel framework for evaluating implementation of conservation and applied the framework to SMART. I examined how faithfully SMART was implemented in practice, how implementation was achieved, and whether the contexts in which it was implemented were conducive to success. I found that SMART implementation was generally strong (i.e., implemented as intended, with consistent inputs, and in supportive contexts), but heterogenous, varying between participating sites and aspects of implementation. Implementation is rarely evaluated in conservation, but that SMART varies between participating sites is consistent with implementation of complex interventions in other fields (e.g., public health (Craig *et al.*, 2008)). Specifically, while SMART was commonly implemented as a tool for monitoring, it was implemented less frequently and faithfully as a tool for adaptive management. This finding was not surprising as adaptive management is difficult to achieve in complex social-ecological systems (Game *et al.*, 2014), yet has implications for the extent to which SMART's outcomes related to management, such as increased patrol presence, will be achieved.

In-depth investigation of SMART's implementation also provided insights into how the intervention might be strengthened to improve outcomes. For example, training in aspects of adaptive management was often weak, suggesting that focusing on enhancing training provision is important for future deployments. Moreover, SMART was sometimes implemented in contexts where resources for patrolling were low (e.g., low ranger density). Yet SMART's theory of change identified that adequate resources for patrolling was an important factor in successful implementation (chapter 2). This finding suggests SMART may be more likely to achieve impacts if future deployments are directed to sites with higher resources for patrolling, or if concurrent support is provided to increase resources.

6.3.3 Intermediate outcomes

In chapter 4 I measured whether an important intermediate outcome in SMART's causal chain – increased patrol presence – had been achieved. I also employed dose-response analysis to assess whether SMART had contributed to this outcome. Results were mixed. Across SMART sites, I found evidence of positive trends in presence over time, for two of three outcome measures. Clear, positive trends in annual spatial patrol coverage were apparent for 44% of sites, and in temporal coverage for 33% of sites (vs. negative trends for 13% and 6%, respectively). Positive trends in these measures were particularly apparent for sites with the poorest coverage. However, trends in monthly spatial patrol coverage, which may be essential for achievement of final impacts (see section 6.2.1), were negligible for the majority of sites. Moreover, differences in outcomes between sites appeared unrelated to implementation of SMART, although the analysis was not conclusive.

Importantly, I only assessed the effect of SMART management mechanisms overtime following implementation (e.g., performance evaluation), but one objective of SMART is oversight of patrol activities. Prior to SMART implementation, in sites without GPS-enabled ranger-based monitoring, patrolling was largely unmonitored. Consequently, rangers in some contexts may have misreported their activities whilst on patrol (pers. comm.). Implementation may have improved this situation, and the change would not be detected in my analysis, precisely because patrols were not previously monitored. Yet that would not justify the expense of implementation of management mechanisms.

6.3.4 Can a contribution claim be made for SMART?

In summary, evidence developed in the course of this thesis did not support the case for SMART having contributed to a reduction in illegal activity (Table 6.1). Using contribution analysis, causality can be inferred if evidence supports four conditions (Table 6.1; Mayne, 2012). This thesis did not fully assess whether each condition had been met. However, the evidence I did develop would not have been sufficient to make a credible causal claim. Had the evidence supported the conditions, it would also have been necessary to verify that final impacts had been achieved (i.e., that illegal activity had decreased) and to fully recognise the relative contribution of external factors.

Table 6.1. Summary of evidence for causal effect of SMART on reduced illegal activity in protected areas via ranger-based patrolling. Contribution analysis approach to impact evaluation argues that causality can be inferred if aspects of a theory of change are verified by empirical evidence (Mayne, 2012).

Evidence required to infer causality	Evidence for SMART	Justification
Assumptions underlying the theory of change are plausible and agreed upon by stakeholders.	Mixed	Assumptions agreed by stakeholders, but review of existing evidence found some unsupported links in causal chain (chapter 2). Novel investigation of one integral link lacking evidence – whether patrols deter crime – found some weak evidence that the assumption held in practice (chapter 5).
Intervention was implemented as outlined in the theory of change.	Mixed	Implementation was generally strong (i.e., as outlined in theory) but varied between sites and aspects of implementation, and was weakest in management activities related to achievement of important intermediate outcomes (chapter 3).
Chain of expected results occurred.	Intermediate outcomes: Mixed. Final impacts: not assessed.	Key intermediate outcome – increased patrol presence – achieved for two of three measures assessed (chapter 4). However, outcomes varied broadly between sites, and it was unclear if changes were related to SMART. Outcome not achieved for third measure, which may be important for causal effects. Final impact (levels of illegal activity) not assessed.
Relative contribution of external factors was recognised.	Not fully assessed	Factors contributing to success of implementation identified (chapter 2) and assessed (chapter 3). Future research should recognise external factors contributing to each step in the causal chain.

My analysis does not suggest that SMART is incapable of achieving its long-term goals or is flawed in design, only that a causal claim cannot be made at this time. Importantly, the intervention was not consistently implemented as intended in aspects important for achieving intermediate outcomes, such as increased patrol coverage. If SMART was implemented as intended outcomes might be achieved. Yet that monthly spatial patrol coverage was constrained by budgets suggests any additional achievable increases may be constrained. Moreover, even if increases in patrol presence were large, evidence from chapter 5 suggests that it is unclear whether this would be consistently sufficient to deter or detect illegal activity. In this respect SMART is comparable to conservation interventions in general, which often aim to achieve long-term, ultimate goals (e.g., reducing illegal activity) but act on immediate, proximate problems (e.g., limited resources for patrolling and inadequate monitoring in protected areas) without clear evidence that addressing the latter will lead to the former (Kapos *et al.*, 2009).

Nevertheless, the evaluation also provided suggestions for how the intervention might be adapted to achieve impact. Firstly, evidence from chapter 5 suggests patrols are likely to be stronger deterrents in open habitats, such as savannah or marine sites, and future deployments such focus on these environments. Secondly, in theory, instead of increasing patrol presence, SMART could also act by facilitating targeting of patrols towards areas at high risk of illegal activity, drawing on data derived from ranger-based monitoring (chapter 2). Interrogation of SMART's theory of change revealed that evidence for this causal chain is weak, as ranger-based monitoring is not a reliable source of data on illegal activity and adapting patrol effort to mimic robust survey effort would detract from enforcement patrolling. However, if ranger-based monitoring were replaced with an alternative source of data on illegal activity, which was sufficiently reliable and cost-effective (e.g., passive acoustic monitoring (Hill *et al.*, 2018)), SMART could act as a platform to analyse such data and to plan patrols and motivate rangers to achieve targets accordingly. Whether targeted deployment of patrols to high-risk areas can enhance detection and deterrence sufficiently to reduce illegal activity is unknown but warrants further research.

6.3.5 Broadening the range of designs and methods for conservation evaluation

In this thesis I advanced conservation evaluation by developing and demonstrating the application of pragmatic, formative methods and designs from other policy arenas, including contribution analysis, dose-response analysis and implementation evaluation, all of which necessitate a clear theory of change for the intervention being assessed. In this section, I reflect on the usefulness of these methods, their limitations, and suggest avenues for further research and development.

The usefulness of a theory of change for informing evaluation, counterfactual or otherwise, is widely acknowledged (Vogel, 2012). All implementers make assumptions about why and how an intervention is necessary and will work, although these assumptions are often implicit. Theory of change approaches and similar conceptual methods force evaluators to make those assumptions explicit (Margoluis *et al.*, 2009b). Of course, there is a risk that assumptions are insufficiently examined or justified, but the risk of unexamined assumptions is probably greater (Archibald *et al.*, 2016). Within the context of this thesis, the theory of change approach facilitated mapping of the 'missing middle' between what SMART does and aims to achieve, permitting assumptions to be interrogated and providing a framework for subsequent aspects of the evaluation.

The utility of theory-based impact evaluation approaches, such as contribution analysis, or other non-counterfactual methods which draw on theories of change, such as dose-response analysis, is less clear. The methods' value in strict impact evaluation terms (i.e., their ability to detect the degree to which changes in outcomes can be attributed to an intervention rather than to other factors (Ferraro, 2009)) may be strongly limited. Firstly, theory-based approaches cannot determine the extent of an effect (Stern *et al.*, 2012). This is particularly problematic as it is essential that decision-makers are able to compare the relative effects of different approaches to enable effective allocation of resources. Secondly, while dose-response analysis can produce an estimate of the extent of an effect, all theory-based methods are strongly subject to bias (Stern *et al.*, 2012). For example, without a sufficiently robust control group for comparison, it is not possible to rule out the possibility that change observed is a result of selection bias. Consequently, such methods struggle to infer causality. Contribution analysis, for example, may strengthen confidence in causal claims, but any conclusion will be qualified and inconclusive (Delahais and Toulemonde, 2017).

Nevertheless, interventions will continue to be implemented for which evaluation methods that can robustly identify contributed effects will not be applicable or feasible. Consequently, whilst conservation scientists must continue to push for application of rigorous counterfactual methods, we must also develop and employ alternative evaluation methods which provide less certainty in conclusions but greater flexibility. In part, this will mean embracing uncertainty and drawing on approaches from fields such as decision theory to make decisions which are good enough, given uncertainty (Milner-Gulland and Shea, 2017). Moreover, while individual methods may lack precision, conclusions derived using one approach can be confirmed and corroborated using multiple, complementary methods (i.e., triangulation (Stern *et al.*, 2012)). For example, contribution analysis can be combined with process tracing to strengthen inference (Befani and Stedman-Bryce, 2017).

In less strict terms, however, theory-based evaluation can be extremely useful. While the result of an evaluation – a causal claim – may be inconclusive, the process can generate other information of equal importance to decision-makers. Specifically, by elucidating the mechanisms through which the intervention is thought to operate and investigating weaknesses, the methods permit identification of how it might be improved. This information is essential for programme development, especially for relatively new and untested interventions (Baylis *et al.*, 2016). With regard to SMART, following a contribution analysis framework enabled generation of such evidence.

Similarly, implementation evaluation, which focuses on the processes by which an intervention operates rather than the intervention's effects, are generally intended to be formative in nature (Rossi *et al.*, 2004). In chapter 3, I presented a novel framework for evaluating heterogeneity in conservation implementation, which considers three critical aspects: what was delivered in practice (Activities), how delivery was achieved (Inputs), and contextual contexts which influence implementation or impact (Moderators). Application of the framework generated evidence which suggested how implementation might be improved. Such an approach could be criticised for focusing on short-term aspects of conservation which, whilst commonly reported to donors to demonstrate achievement, may be poorly related to impacts (Kapos *et al.*, 2009; Pressey *et al.*, 2015). However, to be able to improve interventions, it is essential to understand why correlations between implementation and impact are often weak. Implementation evaluation can help in this regard (e.g., by identifying implementation failure (Montgomery *et al.*, 2013)).

6.4 Conclusion

Law enforcement approaches specifically and protected areas in general typify the "fences and fines" approach to nature conservation, which is almost as old as the field itself; the idea that nature needs to be protected from people (Mace, 2014). In certain respects, conservation thinking has evolved, at least in the scientific literature. Whilst debates proliferate over whether nature should be protected for its own sake or for the good of people (Sandbrook, 2015), most sides acknowledge that the poorest, particularly in the global south, often bear the greatest costs for conservation (Poudyal *et al.*, 2018). Moreover, as illegal natural resource use in protected areas in some contexts is driven by poverty and lack of alternative livelihoods (Duffy *et al.*, 2016), ethical and effective protection often requires incentivisation for local communities to support conservation activities, alongside disincentives to engaging in illegal activity (Cooney *et al.*, 2017). Yet, despite awareness in the literature that a plurality of approaches may be necessary, law enforcement remains the dominant mode in practice. Consequently, it is essential that enforcement is socially just, to ensure people's rights are not infringed (Duffy *et al.*, 2015). Evidence suggests that increasing militarization in some contexts may be undermining enforcement's ability to meet these ideals (Duffy *et al.*, 2019). Equally, it is vital that enforcement is effective, or risk rendering these infringements meaningless. More and better research into patrolling will be essential in this regard.

References

Abernethy, K. A., Coad, L., Taylor, G., Lee, M. E. and Maisels, F. (2013) 'Extent and ecological consequences of hunting in Central African rainforests in the twenty-first century', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. doi: 10.1098/rstb.2012.0303.

Archibald, T., Sharrock, G., Buckley, J. and Cook, N. (2016) 'Assumptions, conjectures, and other miracles: The application of evaluative thinking to theory of change models in community development', *Evaluation and Program Planning*. Pergamon, 59, pp. 119–127. doi: 10.1016/J.EVALPROGPLAN.2016.05.015.

Balmford, A., Gaston, K. J., Blyth, S., James, A. and Kapos, V. (2003) 'Global variation in terrestrial conservation costs, conservation benefits, and unmet conservation needs.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), pp. 1046–50. doi: 10.1073/pnas.0236945100.

Barichievy, C., Munro, L., Clinning, G., Whittington-Jones, B. and Masterson, G. (2017) 'Do armed field-rangers deter rhino poachers? An empirical analysis', *Biological Conservation*. Elsevier, 209, pp. 554–560. doi: 10.1016/J.BIOCON.2017.03.017.

Barnes, M. D., Glew, L., Wyborn, C. and Craigie, I. D. (2018) 'Prevent perverse outcomes from global protected area policy', *Nature Ecology & Evolution*. Nature Publishing Group, 2(5), pp. 759–762. doi: 10.1038/s41559-018-0501-y.

Basch, C. E., Sliepcevich, E. M., Gold, R. S., Duncan, D. F. and Kolbe, L. J. (1985) 'Avoiding type III errors in health education program evaluations: a case study.', *Health education quarterly*, 12(4), pp. 315–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4077544 (Accessed: 16 January 2019).

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., … Wunder, S. (2016) 'Mainstreaming Impact Evaluation in Nature Conservation', *Conservation Letters*. Wiley/Blackwell (10.1111), 9(1), pp. 58–64. doi: 10.1111/conl.12180.

Beale, C. M., Brewer, M. J. and Lennon, J. J. (2014) 'A new statistical framework for the quantification of covariate associations with species distributions', *Methods in Ecology and Evolution*, 5, pp. 421–432. doi: 10.1111/2041-210X.12174.

Beale, C. M., Hauenstein, S., Mduma, S., Frederick, H., Jones, T., Bracebridge, C., … Kohi, E. M. (2018) 'Spatial analysis of aerial survey data reveals correlates of elephant carcasses within a heavily poached ecosystem', *Biological Conservation*. Elsevier, 218, pp. 258–267. doi: 10.1016/J.BIOCON.2017.11.016.

Becker, M., McRobb, R., Watson, F., Droge, E., Kanyembo, B., Murdoch, J. and Kakumbi, C. (2013) 'Evaluating wire-snare poaching trends and the impacts of by-catch on elephants and large carnivores', *Biological Conservation*. Elsevier, 158, pp. 26–36. doi: 10.1016/J.BIOCON.2012.08.017.

Befani, B. and Stedman-Bryce, G. (2017) 'Process Tracing and Bayesian Updating for impact evaluation', *Evaluation*. SAGE PublicationsSage UK: London, England, 23(1), pp. 42–60. doi: 10.1177/1356389016654584.

Van Belle, S. B., Marchal, B., Dubourg, D., Kegels, G., Chowdhary, N., Lund, C., … Prince, M. (2010) 'Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions', *BMC Public Health*. BioMed Central, 10(1), p. 741. doi: 10.1186/1471-2458-10-741.

Benítez-López, A., Alkemade, R., Schipper, A. M., Ingram, D. J., Verweij, P. A., Eikelboom, J. A. J. and Huijbregts, M. A. J. (2017) 'The impact of hunting on tropical mammal and bird populations.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 356(6334), pp. 180–183. doi: 10.1126/science.aaj1891.

Bernard, H. R. (2011) *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. 5th edn. AltaMira Press.

Bertzky, M. and Stoll-Kleemann, S. (2009) 'Multi-level discrepancies with sharing data on protected areas: What we have and what we need for the global village', *Journal of Environmental Management*. Academic Press, 90(1), pp. 8–24. doi: 10.1016/J.JENVMAN.2007.11.001.

Biggs, D., Cooney, R., Roe, D., Dublin, H. T., Allan, J. R., Challender, D. W. S. and Skinner, D. (2017) 'Developing a theory of change for a community-based response to illegal wildlife trade', *Conservation Biology*, 31(1), pp. 5–12. doi: 10.1111/cobi.12796.

Blackburn, S., Hopcraft, J. G. C., Ogutu, J. O., Matthiopoulos, J. and Frank, L. (2016) 'Human-wildlife conflict, benefit sharing and the survival of lions in pastoralist community-based conservancies', *Journal of Applied Ecology*. Edited by N. Singh. John Wiley & Sons, Ltd (10.1111), 53(4), pp. 1195–

1205. doi: 10.1111/1365-2664.12632.

Bonell, C. P., Hargreaves, J., Cousens, S., Ross, D., Hayes, R., Petticrew, M. and Kirkwood, B. R. (2011) ‘Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions.’, *Journal of epidemiology and community health*. BMJ Publishing Group Ltd, 65(7), pp. 582–7. doi: 10.1136/jech.2008.082602.

Bottrill, M. C. and Pressey, R. L. (2012) ‘The effectiveness and evaluation of conservation planning’, *Conservation Letters*. Wiley/Blackwell (10.1111), pp. 407–420. doi: 10.1111/j.1755-263X.2012.00268.x.

Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L. and Resnick, B. (2010) ‘Implementation fidelity in community-based interventions.’, *Research in nursing & health*. NIH Public Access, 33(2), pp. 164–73. doi: 10.1002/nur.20373.

Bruner, A. G., Gullison, R. E., Rice, R. E. and da Fonseca, G. A. (2001) ‘Effectiveness of parks in protecting tropical biodiversity.’, *Science (New York, N.Y.)*, 291(5501), pp. 125–8. doi: 10.1126/science.291.5501.125.

Cascio, W. F. and Aguinis, H. (2011) *Applied psychology in human resource management*. Prentice Hall.

CBD (2010) *Strategic plan for biodiversity 2011–2020 - COP 10, decision X/2*. Montreal, Canada.

Chamlin, M. B., Grasmick, H. G., Bursik, R. J. and Cochran, J. K. (1992) ‘Time aggregation and time lag in macro-level deterrence research’, *Criminology*. John Wiley & Sons, Ltd (10.1111), 30(3), pp. 377–396. doi: 10.1111/j.1745-9125.1992.tb01109.x.

Chandler, R. E. and Scott, E. M. (2011) *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*, *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. doi: 10.1002/9781119991571.

Chhatre, A. and Agrawal, A. (2008) ‘Forest commons and local enforcement.’, *Proceedings of the National Academy of Sciences of the United States of America*, 105, pp. 13286–13291. doi: 10.1073/pnas.0803399105.

CITES (2010) ‘Monitoring of illegal hunting in elephant range states’, *Fifteenth meeting of the Conference of the Parties Doha (Qatar), 13-25 March 2010*, 2, p. 27. Available at: http://www.cites.org/eng/cop/15/doc/E15-44-02.pdf.

Cleveland, R. B., Cleveland, W. S., McRae, J. E. and Terpenning, I. (1990) 'STL: A seasonal-trend decomposition procedure based on loess', *Journal of Official Statistics*. doi: citeulike-article-id:1435502.

Coad, L. (2007) *Bushmeat hunting in Gabon: Socio-economics and hunter behaviour*. University of Cambridge and Imperial College London. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.711.1905&rep=rep1&type=pdf (Accessed: 2 January 2019).

Cooney, R., Roe, D., Dublin, H., Phelps, J., Wilkie, D., Keane, A., … Biggs, D. (2017) 'From Poachers to Protectors: Engaging Local Communities in Solutions to Illegal Wildlife Trade', *Conservation Letters*. Wiley/Blackwell (10.1111), 10(3), pp. 367–374. doi: 10.1111/conl.12294.

Craig, P., Dieppe, P., Macintyre, S., Nazareth, I. and Petticrew, M. (2008) 'Developing and evaluating complex interventions: the new Medical Research Council guidance', *The BMJ*, 337, p. a1655. doi: 10.1136/bmj.a1655.

Craigie, I. D., Baillie, J. E. M., Balmford, A., Carbone, C., Collen, B., Green, R. E. and Hutton, J. M. (2010) 'Large mammal population declines in Africa's protected areas', *Biological Conservation*. Elsevier Ltd, 143(9), pp. 2221–2228. doi: 10.1016/j.biocon.2010.06.007.

Critchlow, R., Plumptre, A. J., Alidria, B., Nsubuga, M., Driciru, M., Rwetsiba, A., … Beale, C. M. (2017) 'Improving Law-Enforcement Effectiveness and Efficiency in Protected Areas Using Ranger-collected Monitoring Data', *Conservation Letters*. Wiley/Blackwell (10.1111), 10(5), pp. 572–580. doi: 10.1111/conl.12288.

Critchlow, R., Plumptre, A. J., Driciru, M., Rwetsiba, A., Stokes, E. J., Tumwesigye, C., … Beale, C. M. (2015) 'Spatiotemporal trends of illegal activities from ranger-collected data in a Ugandan national park', *Conservation Biology*, 29(5), pp. 1458–1470. doi: 10.1111/cobi.12538.

CSIR (2019) *Cmore – A Platform for Shared Awareness*. Available at: https://cmore.csir.co.za/ (Accessed: 25 February 2019).

Curran, B., Sunderland, T., Maisels, F., Oates, J., Asaha, S., Balinga, M., … Roth, P. (2009) 'Are central Africa's protected areas displacing hundreds of thousands of rural poor?', *Conservation and Society*. Medknow Publications and Media Pvt. Ltd., 7(1), p. 30. doi: 10.4103/0972-4923.54795.

CyberTracker Conservation (2013) *CyberTracker GPS Field Data Collection System*. Available at: https://www.cybertracker.org/ (Accessed: 25 February 2019).

Delahais, T. and Toulemonde, J. (2017) 'Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management?', *Evaluation*. SAGE PublicationsSage UK: London, England, 23(4), pp. 370–388. doi: 10.1177/1356389017733211.

DENR (2016) *Joint FMB-BMB Technical Bulletin, No. 2016-01: Enhancing forest protection through application of the LAWIN forest and biodiversity protection system*. Available at: http://forestry.denr.gov.ph/images/article/LAWIN_JTB.pdf (Accessed: 25 February 2019).

Dobson, A. D. M., Milner-Gulland, E. J., Beale, C. M., Ibbett, H. and Keane, A. (2018) 'Detecting deterrence from patrol data', *Conservation Biology*. Wiley/Blackwell (10.1111). doi: 10.1111/cobi.13222.

Donovan, K. P. (2018) 'The rise of the randomistas: on the experimental turn in international aid', *Economy and Society*. Routledge, 47(1), pp. 27–58. doi: 10.1080/03085147.2018.1432153.

Dudley, N. (ed.) (2008) *Guidelines for Applying Protected Area Management Categories*. Gland, Switzerland. Available at: www.iucn.org/pa_guidelines (Accessed: 16 January 2019).

Duffy, R., Massé, F., Smidt, E., Marijnen, E., Büscher, B., Verweijen, J., … Lunstrum, E. (2019) 'Why we must question the militarisation of conservation', *Biological Conservation*. Elsevier, 232, pp. 66–73. doi: 10.1016/J.BIOCON.2019.01.013.

Duffy, R., St John, F. A. V., Büscher, B. and Brockington, D. (2015) 'The militarization of anti-poaching: undermining long term goals?', *Environmental Conservation*. Cambridge University Press, 42(04), pp. 345–348. doi: 10.1017/S0376892915000119.

Duffy, R., St John, F. A. V, Büscher, B. and Brockington, D. (2016) 'Toward a new understanding of the links between poverty and illegal wildlife hunting.', *Conservation biology : the journal of the Society for Conservation Biology*. Wiley-Blackwell, 30(1), pp. 14–22. doi: 10.1111/cobi.12622.

Durlak, J. A. and DuPre, E. P. (2008) 'Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation', *American Journal of Community Psychology*, 41(3–4), pp. 327–350. doi: 10.1007/s10464-008-9165-0.

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F. and Schabenberger, O. (2008) 'An R2 statistic for fixed effects in the linear mixed model.', *Statistics in medicine*. NIH Public Access, 27(29), pp. 6137–57. doi: 10.1002/sim.3429.

Escoffery, C., Bundy, L., Haardoerfer, R., Berg, C. J., Savas, L. S., Williams, R. S. and Kegler, M. C. (2016) 'A process evaluation of an intervention to promote home smoking bans among low income households', *Evaluation and Program Planning*. Pergamon, 55, pp. 120–125. doi: 10.1016/J.EVALPROGPLAN.2015.12.008.

Fang, F., Tambe, M., Dilkina, B. and Plumptre, A. J. (2019) *Artificial Intelligence and Conservation*. Cambridge University Press. Available at: https://www.cambridge.org/us/academic/subjects/computer-science/artificial-intelligence-and-natural-language-processing/artificial-intelligence-and-conservation?format=PB&isbn=9781108464734 (Accessed: 17 February 2019).

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

Ferraro, P. J. and Hanauer, M. M. (2014) 'Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1307712111.

Ferraro, P. J. and Pattanayak, S. K. (2006) 'Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments', *PLoS Biology*. Edited by G. Mace. Public Library of Science, 4(4), p. e105. doi: 10.1371/journal.pbio.0040105.

Ferraro, P. J. and Pressey, R. L. (2015) 'Measuring the difference made by conservation initiatives: Protected areas and their environmental and social impacts', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2014.0270.

Fisher, B., Balmford, A., Ferraro, P. J., Glew, L., Mascia, M., Naidoo, R. and Ricketts, T. H. (2014) 'Moving Rio Forward and Avoiding 10 More Years with Little Evidence for Effective Conservation Policy', *Conservation Biology*. John Wiley & Sons, Ltd (10.1111), 28(3), pp. 880–882. doi: 10.1111/cobi.12221.

Fleishman, E., Costa, D. P., Harwood, J., Kraus, S., Moretti, D., New, L. F., … Wells, R. S. (2016) 'Monitoring population-level responses of marine mammals to human activities', *Marine Mammal Science*. Wiley/Blackwell (10.1111), 32(3), pp. 1004–1021. doi: 10.1111/mms.12310.

Game, E. T., Bremer, L. L., Calvache, A., Moreno, P. H., Vargas, A., Rivera, B. and Rodriguez, L. M. (2018) 'Fuzzy Models to Inform Social and Environmental Indicator Selection for Conservation Impact Monitoring', *Conservation Letters*. Wiley/Blackwell (10.1111), 11(1), p. e12338. doi: 10.1111/conl.12338.

Game, E. T., Meijaard, E., Sheil, D. and McDonald-Madden, E. (2014) 'Conservation in a Wicked Complex World; Challenges and Solutions', *Conservation Letters*. Wiley/Blackwell (10.1111), 7(3), pp. 271–277. doi: 10.1111/conl.12050.

Gavin, M. C., Solomon, J. N. and Blank, S. G. (2010) 'Measuring and monitoring illegal use of natural resources', *Conservation Biology*. Blackwell Publishing Inc, 24(1), pp. 89–100. doi: 10.1111/j.1523-1739.2009.01387.x.

Geldmann, J., Barnes, M., Coad, L., Craigie, I. D., Hockings, M. and Burgess, N. D. (2013) 'Effectiveness of terrestrial protected areas in reducing habitat loss and population declines', *Biological Conservation*. Elsevier, 161, pp. 230–238. doi: 10.1016/J.BIOCON.2013.02.018.

Geldmann, J., Coad, L., Barnes, M. D., Craigie, I. D., Woodley, S., Balmford, A., … Chelsea, C. (2018) 'A global analysis of management capacity and ecological outcomes in terrestrial protected areas Kathryn Knights 15 Michael B. Mascia 16 Louise McRae 17', 11, p. 14. doi: 10.1111/conl.12434.

Gray, M. and Kalpers, J. (2005) 'Ranger based monitoring in the Virunga-Bwindi region of East-Central Africa: A simple data collection tool for park management', *Biodiversity and Conservation*, 14(11), pp. 2723–2741. doi: 10.1007/s10531-005-8406-x.

Gray, T., Lynam, A., Seng, T., Laurance, W., Long, B., Scotson, L. and Ripple, W. (2017) 'Wildlife-snaring crisis in Asian forests.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 355(6322), pp. 255–256. doi: 10.1126/science.aal4463.

Harrison, R. D. (2011) 'Emptying the Forest: Hunting and the Extirpation of Wildlife from Tropical Nature Reserves', *BioScience*. Oxford University PressAmerican Institute of Biological Sciences, 61(11), pp. 919–924. doi: 10.1525/bio.2011.61.11.11.

Henson, D. W., Malpas, R. C. and D'Udine, F. A. C. (2016) *Wildlife law enforcement in Sub-Saharan African protected areas : a review of best practices*. IUCN International Union for Conservation of Nature. doi: 10.2305/IUCN.CH.2015.SSC-OP.58.en.

Herzberg, F. (1959) *The motivation to work*. 2nd ed. New York: Wiley. Available at: https://www.worldcat.org/title/motivation-to-work/oclc/803681223 (Accessed: 6 April 2018).

Hilborn, R., Arcese, P., Borner, M., Hando, J., Hopcraft, G., Loibooki, M., … Sinclair, A. R. E. (2006) 'Effective Enforcement in a Conservation Area', *Science*, 314(5803), p. 1266. doi: 10.1126/science.1132780.

Hill, A. B. (1965) 'The Environment and Disease: Association or Causation?', *Proceedings of the Royal Society of Medicine*. Royal Society of Medicine Press, 58(5), pp. 295–300. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14283879 (Accessed: 18 February 2019).

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L. and Rogers, A. (2018) 'AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment', *Methods in Ecology and Evolution*. Edited by N. Isaac. John Wiley & Sons, Ltd (10.1111), 9(5), pp. 1199–1211. doi: 10.1111/2041-210X.12955.

Hoare, R. (2015) 'Lessons From 20 Years of Human–Elephant Conflict Mitigation in Africa', *Human Dimensions of Wildlife*. Routledge, 20(4), pp. 289–295. doi: 10.1080/10871209.2015.1005855.

Hötte, M. H. H., Kolodin, I. A., Bereznuk, S. L., Slaght, J. C., Kerley, L. L., Soutyrina, S. V., … Miquelle, D. G. (2016) 'Indicators of success for smart law enforcement in protected areas: A case study for Russian Amur tiger (Panthera tigris altaica) reserves', *Integrative Zoology*. Wiley/Blackwell (10.1111), 11(1), pp. 2–15. doi: 10.1111/1749-4877.12168.

Howe, C. and Milner-Gulland, E. J. (2012) 'The view from the office is not all bad: conservation evaluation as a "sexy" research goal', *Animal Conservation*. Wiley/Blackwell (10.1111), 15(3), pp. 231–232. doi: 10.1111/j.1469-1795.2012.00563.x.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) *Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf (Accessed: 26 April 2018).

Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P. and Roy, D. B. (2014) 'Statistics for citizen science: extracting signals of change from noisy ecological data', *Methods in Ecology and Evolution*. Edited by B. Anderson, 5(10), pp. 1052–1060. doi: 10.1111/2041-210X.12254.

IUCN and UNEP-WCMC (2018) 'The World Database on Protected Areas (WDPA)'. Cambridge, UK: UNEP-WCMC. Available at: www.protectedplanet.net (Accessed: 1 December 2018).

Jachmann, H. (2008) 'Monitoring law-enforcement performance in nine protected areas in Ghana', *Biological Conservation*, 141, pp. 89–99. doi: 10.1016/j.biocon.2007.09.012.

Jachmann, H. and Billiouw, M. (1997) 'Elephant Poaching and Law Enforcement in the Central Luangwa Valley, Zambia', *The Journal of Applied Ecology*, 34(1), p. 233. doi: 10.2307/2404861.

Jacobson, S. K. (1991) 'Evaluation model for developing, implementing, and assessing conservation education programs: Examples from Belize and Costa Rica', *Environmental Management*. Springer-Verlag, 15(2), pp. 143–150. doi: 10.1007/BF02393846.

Johnson, A., Goodrich, J., Hansel, T., Rasphone, A., Saypanya, S., Vongkhamheng, C., … Strindberg (2016) 'To protect or neglect? Design, monitoring, and evaluation of a law enforcement strategy to recover small populations of wild tigers and their prey', *Biological Conservation*. Elsevier, 202, pp. 99–109. doi: 10.1016/J.BIOCON.2016.08.018.

Jones, K. W. and Lewis, D. J. (2015) 'Estimating the Counterfactual Impact of Conservation Programs on Land Cover Outcomes: The Role of Matching and Panel Regression Techniques', *PLOS ONE*. Edited by M. S. Crowther. Public Library of Science, 10(10), p. e0141380. doi: 10.1371/journal.pone.0141380.

Joppa, L. N., Loarie, S. R. and Pimm, S. L. (2008) 'On the protection of "protected areas"', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(18), pp. 6673–8. doi: 10.1073/pnas.0802471105.

Joppa, L. N. and Pfaff, A. (2009) 'High and far: biases in the location of protected areas.', *PloS one*, 4(12), p. e8273. doi: 10.1371/journal.pone.0008273.

Kahler, J. S. and Gore, M. L. (2012) 'Beyond the cooking pot and pocket book: factors influencing noncompliance with wildlife poaching rules', *International Journal of Comparative and Applied Criminal Justice*. Taylor & Francis, 36(2), pp. 103–120.

Kapos, V., Balmford, A., Aveling, R., Bubb, P., Carey, P., Entwistle, A., … Manica, A. (2009) 'Outcomes, not implementation, predict conservation success', *Oryx*. Cambridge University Press, 43(03), p. 336. doi: 10.1017/S0030605309990275.

Kauano, É. E., Silva, J. M. C. and Michalski, F. (2017) ‘Illegal use of natural resources in federal protected areas of the Brazilian Amazon’, *PeerJ*, 5, p. e3902. doi: 10.7717/peerj.3902.

Keane, A., Jones, J. P. G., Edwards-Jones, G. and Milner-Gulland, E. J. (2008) ‘The sleeping policeman: understanding issues of enforcement and compliance in conservation’, *Animal Conservation*, 11(2), pp. 75–82. doi: 10.1111/j.1469-1795.2008.00170.x.

Keane, A., Jones, J. P. G. and Milner-Gulland, E. J. (2011) ‘Encounter data in resource management and ecology: pitfalls and possibilities’, *Journal of Applied Ecology*. Wiley/Blackwell (10.1111), 48(5), pp. 1164–1173. doi: 10.1111/j.1365-2664.2011.02034.x.

Keith, D. A., Martin, T. G., McDonald-Madden, E. and Walters, C. (2011) ‘Uncertainty and adaptive management for biodiversity conservation’, *Biological Conservation*, 144(4), pp. 1175–1178. doi: 10.1016/j.biocon.2010.11.022.

Kuempel, C. D., Adams, V. M., Possingham, H. P. and Bode, M. (2018) ‘Bigger or better: The relative benefits of protected area network expansion and enforcement for the conservation of an exploited species’, *Conservation Letters*. Wiley/Blackwell (10.1111), p. e12433. doi: 10.1111/conl.12433.

Laurance, W. F., Carolina Useche, D., Rendeiro, J., Kalka, M., Bradshaw, C. J. a., Sloan, S. P., … Zamzani, F. (2012) ‘Averting biodiversity collapse in tropical forest protected areas’, *Nature*. Nature Publishing Group, 489, pp. 290–294. doi: 10.1038/nature11318.

Leader-Williams, N., Albon, S. D. and Berry, P. S. M. (1990) ‘Illegal Exploitation of Black Rhinoceros and Elephant Populations: Patterns of Decline, Law Enforcement and Patrol Effort in Luangwa Valley, Zambia’, *The Journal of Applied Ecology*. British Ecological Society, 27(3), p. 1055. doi: 10.2307/2404395.

Leader-Williams, N. and Milner-Gulland, E. J. (1993) ‘Policies for the Enforcement of Wildlife Laws: The Balance between Detection and Penalties in Luangwa Valley, Zambia’, *Conservation Biology*. John Wiley & Sons, Ltd (10.1111), 7(3), pp. 611–617. doi: 10.1046/j.1523-1739.1993.07030611.x.

Lemieux, A., Bernasco, W. and Rwetsiba, A. (2014) ‘Tracking poachers in Uganda: spatial models of patrol intensity and patrol efficiency’, in Lemieux, A. (ed.) *Situational Prevention of Poaching*. Routledge, pp. 126–143. doi: 10.4324/9780203094525-15.

Lendrum, A. and Humphrey, N. (2012) 'The importance of studying the implementation of interventions in school settings', *Oxford Review of Education*. Routledge, 38(5), pp. 635–652. doi: 10.1080/03054985.2012.734800.

Leverington, F., Costa, K. L., Pavese, H., Lisle, A. and Hockings, M. (2010) 'A global analysis of protected area management effectiveness', *Environmental Management*, 46, pp. 685–698. doi: 10.1007/s00267-010-9564-5.

Linkie, M., Martyr, D. J., Harihar, A., Risdianto, D., Nugraha, R. T., Maryati, … Wong, W.-M. (2015) 'Safeguarding Sumatran tigers: evaluating effectiveness of law enforcement patrols and local informant networks', *Journal of Applied Ecology*. Edited by M. Hayward. John Wiley & Sons, Ltd (10.1111), 52(4), pp. 851–860. doi: 10.1111/1365-2664.12461.

Lyons, J. E., Runge, M. C., Laskowski, H. P. and Kendall, W. L. (2008) 'Monitoring in the Context of Structured Decision-Making and Adaptive Management', *Journal of Wildlife Management*, 72, pp. 1683–1692. doi: 10.2193/2008-141.

Mace, G. M. (2014) 'Whose conservation?', *Science*. American Association for the Advancement of Science, 345(6204), pp. 1558–1560. doi: 10.1126/science.1254704.

Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G. and Pawson, R. (2013) 'Realist RCTs of complex interventions - An oxymoron', *Social Science and Medicine*. doi: 10.1016/j.socscimed.2013.06.025.

Margoluis, R., Stem, C., Salafsky, N. and Brown, M. (2009a) 'Design alternatives for evaluating the impact of conservation projects', *New Directions for Evaluation*. Wiley Subscription Services, Inc., A Wiley Company, 2009(122), pp. 85–96. doi: 10.1002/ev.298.

Margoluis, R., Stem, C., Salafsky, N. and Brown, M. (2009b) 'Using conceptual models as a planning and evaluation tool in conservation', *Evaluation and Program Planning*. Pergamon, 32(2), pp. 138–147. doi: 10.1016/J.EVALPROGPLAN.2008.09.007.

Mayne, J. (2012) 'Contribution analysis: Coming of age?', *Evaluation*. SAGE PublicationsSage UK: London, England, 18(3), pp. 270–280. doi: 10.1177/1356389012451663.

Mayo, E. (1933) *The human problems of an industrial civilization,*. New York: Macmillan Co. Available at: http://www.worldcat.org/title/human-problems-of-an-industrial-civilization/oclc/304903

(Accessed: 6 April 2018).

de Merode, E., Smith, K. H., Homewood, K., Pettifor, R., Rowcliffe, M. and Cowlishaw, G. (2007) 'The impact of armed conflict on protected-area efficacy in Central Africa.', *Biology letters*, 3(3), pp. 299–301. doi: 10.1098/rsbl.2007.0010.

Milner-Gulland, E. J. and Shea, K. (2017) 'Embracing uncertainty in applied ecology', *Journal of Applied Ecology*. Edited by A. Punt. John Wiley & Sons, Ltd (10.1111), 54(6), pp. 2063–2068. doi: 10.1111/1365-2664.12887.

Di Minin, E. and Toivonen, T. (2015) 'Global Protected Area Expansion: Creating More than Paper Parks.', *Bioscience*. Oxford University Press, 65(7), pp. 637–638. doi: 10.1093/biosci/biv064.

MoEF (no date) *MSTrIPES: Monitoring system for Tigers – Intensive Protection and Ecological Status*. Available at: http://www.moef.nic.in/downloads/public-information/mstripes.pdf (Accessed: 25 February 2019).

Montgomery, P., Underhill, K., Gardner, F., Operario, D. and Mayo-Wilson, E. (2013) 'The Oxford Implementation Index: a new tool for incorporating implementation data into systematic reviews and meta-analyses', *Journal of Clinical Epidemiology*, 66(8), pp. 874–882. doi: 10.1016/j.jclinepi.2013.03.006.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., … Baird, J. (2015) 'Process evaluation of complex interventions: Medical Research Council guidance', *BMJ*. BMJ Publishing Group Ltd, 350. doi: 10.1136/bmj.h1258.

Moore, J. F., Mulindahabi, F., Masozera, M. K., Nichols, J. D., Hines, J. E., Turikunkiko, E. and Oli, M. K. (2018) 'Are ranger patrols effective in reducing poaching-related threats within protected areas?', *Journal of Applied Ecology*, 55, pp. 99–107. doi: 10.1111/1365-2664.12965.

Moreto, W. D. (2015) 'Introducing intelligence-led conservation: bridging crime and conservation science', *Crime Science*. SpringerOpen, 4(1), p. 15. doi: 10.1186/s40163-015-0030-9.

Moreto, W. D. (2016) 'Occupational stress among law enforcement rangers: insights from Uganda', *Oryx*. Cambridge University Press, 50(04), pp. 646–654. doi: 10.1017/S0030605315000356.

Moreto, W. D. and Gau, J. M. (2017) 'Deterrence, Legitimacy, and Wildlife Crime in Protected Areas', in *Conservation Criminology*. Chichester, UK: John Wiley & Sons, pp. 45–58. doi:

10.1002/9781119376866.ch3.

Moreto, W., Lemieux, A. and Rwetsiba, A. (2014) ‘Law enforcement monitoring in Uganda: the utility of official data and time/distance-based ranger efficiency measures’, in Lemieux, A. (ed.) *Situational Prevention of Poaching*. Routledge, pp. 106–125. doi: 10.4324/9780203094525-14.

Morrison, S. A. (2015) ‘A framework for conservation in a human-dominated world’, *Conservation Biology*. Wiley/Blackwell (10.1111), 29(3), pp. 960–964. doi: 10.1111/cobi.12432.

Nagin, D. S. (2013a) ‘Deterrence: A Review of the Evidence by a Criminologist for Economists’, *Annual Review of Economics*. Annual Reviews , 5(1), pp. 83–105. doi: 10.1146/annurev-economics-072412-131310.

Nagin, D. S. (2013b) ‘Deterrence in the Twenty-First Century’, *Crime and Justice*. University of Chicago PressChicago, IL, 42(1), pp. 199–263. doi: 10.1086/670398.

Newing, H., Eagle, C. M., Puri, R. K. and Watson, C. W. (2011) *Conducting research in conservation: A social science perspective*. Routledge.

Nichols, J. D. and Williams, B. K. (2006) ‘Monitoring for conservation.’, *Trends in ecology & evolution*, 21(12), pp. 668–73. doi: 10.1016/j.tree.2006.08.007.

Nolte, C. (2016) ‘Identifying challenges to enforcement in protected areas: empirical insights from 15 Colombian parks’, *Oryx*. Cambridge University Press, 50(02), pp. 317–322. doi: 10.1017/S0030605314000891.

O’Kelly, H. J., Rowcliffe, J. M., Durant, S. M. and Milner-Gulland, E. J. (2018) ‘Experimental estimation of snare detectability for robust threat monitoring’, *Ecology and Evolution*, 8(3), pp. 1778–1785. doi: 10.1002/ece3.3655.

Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J. and RIPPLE Study Team (2006) ‘Process evaluation in randomised controlled trials of complex interventions.’, *BMJ (Clinical research ed.)*. British Medical Journal Publishing Group, 332(7538), pp. 413–6. doi: 10.1136/bmj.332.7538.413.

Ogunjinmi, A., Umunna, M. and Ogunjinmi, K. (2009) ‘Factors Affecting Job Satisfaction Of Rangers In Yankari Game Reserve, Bauchi, Nigeria’, *Journal of Agriculture and Social Research (JASR)*. Agricultural Volunteers of Nigeria (AgVON), 8(2). doi: 10.4314/jasr.v8i2.43332.

Pantel, S. (2007) *MIST Specialist – Evaluation Report. Srepok Wilderness Area Project. Technical Paper*

Series, 2, 35. Available at: http://pubs.iied.org/pdfs/G02322.pdf (Accessed: 6 April 2018).

Papworth, S. K., Bunnefeld, N., Slocombe, K. and Milner-Gulland, E. J. (2012) 'Movement ecology of human resource users: using net squared displacement, biased random bridges and resource utilization functions to quantify hunter and gatherer behaviour', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd (10.1111), 3(3), pp. 584–594. doi: 10.1111/j.2041-210X.2012.00189.x.

Paternoster, R. (2010) 'How Much Do We Really Know about Criminal Deterrence', *Journal of Criminal Law and Criminology*, 100(3). Available at: https://scholarlycommons.law.northwestern.edu/jclc/vol100/iss3/6 (Accessed: 5 January 2019).

Pawson, R. (2006) *Evidence-based policy : a realist perspective*. SAGE.

Pfaff, A., Robalino, J., Sandoval, C. and Herrera, D. (2015) 'Protected area types, strategies and impacts in Brazil's Amazon: public protected area strategies do not yield a consistent ranking of protected area types by impact', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1681), p. 20140273. doi: 10.1098/rstb.2014.0273.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2018) '{nlme}: Linear and Nonlinear Mixed Effects Models'. Available at: https://cran.r-project.org/package=nlme.

Plumptre, A. J. (2019) 'Law Enforcement for Wildlife Conservation', in Fang, F., Tambe, M., Dilkina, B., and Plumptre, A. J. (eds) *Artificial Intelligence and Conservation*. Cambridge University Press. Available at: https://books.google.co.uk/books?hl=en&lr=lang_en&id=tXuIDwAAQBAJ&oi=fnd&pg=PA17&dq=law+enforcement+conservation&ots=yPuvq1OXlM&sig=uW36wnA7AC-e6mAx_bN5H6CkwLs&redir_esc=y#v=onepage&q=law enforcement conservation&f=false (Accessed: 25 February 2019).

Plumptre, A. J., Fuller, R. A., Rwetsiba, A., Wanyama, F., Kujirakwinja, D., Driciru, M., … Possingham, H. P. (2014) 'Efficiently targeting resources to deter illegal activities in protected areas', *Journal of Applied Ecology*. Edited by E. J. Milner-Gulland, 51(3), pp. 714–725. doi: 10.1111/1365-2664.12227.

Poudyal, M., Jones, J. P. G., Rakotonarivo, O. S., Hockley, N., Gibbons, J. M., Mandimbiniaina, R., … Ramamonjisoa, B. S. (2018) 'Who bears the cost of forest conservation?', *PeerJ*, 6, p. e5106. doi: 10.7717/peerj.5106.

Pratt, T. C. and Cullen, F. T. (2005) ‘Assessing Macro-Level Predictors and Theories of Crime: A Meta-Analysis’, *Crime and Justice*. The University of Chicago Press, pp. 373–450. doi: 10.2307/3488363.

Pressey, R. L., Visconti, P. and Ferraro, P. J. (2015) ‘Making parks make a difference: poor alignment of policy, planning and management with protected-area impact, and ways forward’, *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 370(1681), p. 20140280. doi: 10.1098/rstb.2014.0280.

Pringle, R. M. (2017) ‘Upgrading protected areas to conserve wild biodiversity’, *Nature*. Nature Publishing Group, 546(7656), pp. 91–99. doi: 10.1038/nature22902.

Pullin, A. S. and Stewart, G. B. (2006) ‘Guidelines for Systematic Review in Conservation and Environmental Management’, *Conservation Biology*. John Wiley & Sons, Ltd (10.1111), 20(6), pp. 1647–1656. doi: 10.1111/j.1523-1739.2006.00485.x.

Qiu, J., Game, E. T., Tallis, H., Olander, L. P., Glew, L., Kagan, J. S., … Weaver, S. K. (2018) ‘Evidence-Based Causal Chains for Linking Health, Development, and Conservation Actions’, *BioScience*. Oxford University Press, 68(3), pp. 182–193. doi: 10.1093/biosci/bix167.

R Core Team (2018) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria. Available at: https://www.r-project.org/.

Ratcliffe, J. H., Taniguchi, T., Groff, E. R. and Wood, J. D. (2011) ‘The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots’, *Criminology*. John Wiley & Sons, Ltd (10.1111), 49(3), pp. 795–831. doi: 10.1111/j.1745-9125.2011.00240.x.

Redpath, S. M., Linnell, J. D. C., Festa-Bianchet, M., Boitani, L., Bunnefeld, N., Dickman, A., … Milner-Gulland, E. J. (2017) ‘Don’t forget to look down - collaborative approaches to predator conservation’, *Biological Reviews*. Wiley/Blackwell (10.1111), 92(4), pp. 2157–2163. doi: 10.1111/brv.12326.

Reynolds, A. J. (1998) ‘Confirmatory Program Evaluation: A Method for Strengthening Causal Inference’, *American Journal of Evaluation*. Sage PublicationsSage CA: Thousand Oaks, CA, 19(2), pp. 203–221. doi: 10.1177/109821409801900204.

Rogers, P. (2014) ‘Overview: Strategies for Causal Attribution’, in *Methodological Briefs: Impact Evaluation 6*. Florence: UNICEF Office of Research.

Rossi, P. H. (Peter H., Lipsey, M. W. and Freeman, H. E. (2004) *Evaluation : a systematic approach*. Sage.

Rowcliffe, J. M., de Merode, E. and Cowlishaw, G. (2004) 'Do wildlife laws work? Species protection and the application of a prey choice model to poaching decisions', *Proceedings of the Royal Society B: Biological Sciences*, 271(December), pp. 2631–2636. doi: 10.1098/rspb.2004.2915.

Sandbrook, C. (2015) 'What is conservation?', *Oryx*. Cambridge University Press, 49(04), pp. 565–566. doi: 10.1017/S0030605315000952.

SANParks (2018) *South African National Parks Annual Performance Plan 2018/2019*. Pretoria, South Africa. Available at: https://www.sanparks.org/assets/docs/about/annual_performance_plan_2018-2019.pdf (Accessed: 25 February 2019).

Schulze, K., Knights, K., Coad, L., Geldmann, J., Leverington, F., Eassom, A., … Katharina Schulze, C. (2018) 'An assessment of threats to terrestrial protected areas', *Conservation Letters*. doi: 10.1111/conl.12435.

Schwartz, M. W., Deiner, K., Forrester, T., Grof-Tisza, P., Muir, M. J., Santos, M. J., … Zylberberg, M. (2012) 'Perspectives on the Open Standards for the Practice of Conservation', *Biological Conservation*, 155, pp. 169–177. doi: 10.1016/j.biocon.2012.06.014.

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C. and Patel, V. (2014) 'Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions', *Trials*, 15(1), p. 267. doi: 10.1186/1745-6215-15-267.

Simkin, K. (2016) *Trialling the Theory of Change approach to impact evaluation in conservation using the Saiga Conservation Alliance's work in Uzbekistan as a case study*. University of Oxford. Available at: https://www.iccs.org.uk/sites/www.iccs.org.uk/files/inline-files/K Simkin UG project_0.pdf (Accessed: 17 April 2018).

Singh, R., Barlow, C. and Lawson, D. (2015) *PACS: Protection Audit for Conservation Sites*. Available at: http://zeropoaching.org/pdfs/Protection_Audit_Booklet.pdf (Accessed: 29 August 2018).

Sintov, N., Seyranian, V. and Lyet, A. (2018) 'Fostering adoption of conservation technologies: a case study with wildlife law enforcement rangers', *Oryx*. Cambridge University Press, pp. 1–5. doi: 10.1017/S0030605317001533.

SMART Partnership (2017) *SMART: A Guide To Getting Started*. Available at: http://smartconservationtools.org/wp-content/uploads/2017/11/SMART_GettingStarted2017_English_sm.pdf (Accessed: 11 January 2019).

SMART Partnership (2018) *SMART Partnership Annual Report 2017*. Available at: http://smartconservationtools.org/wp-content/uploads/2018/05/SMART_AnnualReport_2017_FINAL_sm.pdf.

Stebbings, E., Copsey, J., Tatayah, V., Black, S. A., Zuël, N. and Ferriere, C. (2016) 'Applying Systems Thinking and Logic Models to Evaluate Effectiveness in Wildlife Conservation', *Open Journal of Leadership*. Scientific Research Publishing, 05(03), pp. 70–83. doi: 10.4236/ojl.2016.53007.

Steckler, A. B. and Linnan, L. (2002) 'Process Evaluation for Public Health Interventions and Research: an Overview', in *Process Evaluation for Public Health Interventions and Research*. doi: 100324917.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012) *Broadening the range of designs and methods for impact evaluations*. Available at: https://www.oecd.org/derec/50399683.pdf (Accessed: 19 February 2019).

Stokes, E. J. (2010) 'Improving effectiveness of protection efforts in tiger source sites: Developing a framework for law enforcement monitoring using MIST', *Integrative Zoology*, 5(4), pp. 363–377. doi: 10.1111/j.1749-4877.2010.00223.x.

Tesfaw, A. T., Pfaff, A., Golden Kroner, R. E., Qin, S., Medeiros, R. and Mascia, M. B. (2018) 'Land-use and land-cover change shape the sustainability and impacts of protected areas.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(9), pp. 2084–2089. doi: 10.1073/pnas.1716462115.

Tittensor, D. P., Walpole, M., Hill, S., Boyce, D., Britten, G. L., Burgess, N., … Reagan, E. (2014) 'A mid-term analysis of progress towards international biodiversity targets', *Science*, Accepted, pp. 1–182. doi: 10.1126/science.1257484.

Tranquilli, S., Abedi-Lartey, M., Abernethy, K., Amsini, F., Asamoah, A., Balangtaa, C., … Sommer, V. (2014) 'Protected Areas in Tropical Africa: Assessing Threats and Conservation Activities', *PLoS ONE*. Edited by F. Moreira. Public Library of Science, 9(12), p. e114154. doi: 10.1371/journal.pone.0114154.

Tranquilli, S., Abedi-Lartey, M., Amsini, F., Arranz, L., Asamoah, A., Babafemi, O., … Kuehl, H. (2012) 'Lack of conservation effort rapidly increases African great ape extinction risk', *Conservation Letters*, 5(1), pp. 48–55. doi: 10.1111/j.1755-263X.2011.00211.x.

UNEP-WCMC, IUCN and NGS (2018) *Protected Planet Report 2018*. Cambridge, UK; Gland, Switzerland; and Washington, D.C., USA. Available at: www.unep-wcmc.org (Accessed: 22 February 2019).

USAID (2018) *Partners in the Philippines' Development*. Manila, Philippines. Available at: https://www.usaid.gov/sites/default/files/documents/1861/USAID_Philippines_Development_Handbook_as_of_July_2018.pdf (Accessed: 25 February 2019).

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

Veríssimo, D. and Wan, A. K. Y. (2018) 'Characterizing efforts to reduce consumer demand for wildlife products', *Conservation Biology*. Wiley/Blackwell (10.1111). doi: 10.1111/cobi.13227.

Vogel, I. (2012) *Review of the use of Theory of Change in International Development*. London: UK Department of International Development.

Vulcan Inc. (2018) *EarthRanger*. Available at: https://earthranger.com/ (Accessed: 25 February 2019).

Walmsley, S. F. and White, A. T. (2003) 'Influence of social, management and enforcement factors on the long-term ecological effects of marine sanctuaries', *Environmental Conservation*. Cambridge University Press, 30(04), pp. 388–407. Available at: http://journals.cambridge.org/abstract_S0376892903000407 (Accessed: 6 March 2015).

Walters, C. (2003) 'Folly and fantasy in the analysis of spatial catch rate data', *Canadian Journal of Fisheries and Aquatic Sciences*. NRC Research Press Ottawa, Canada , 60(12), pp. 1433–1436. doi: 10.1139/f03-152.

Watson, J. E. M., Dudley, N., Segan, D. B. and Hockings, M. (2014) 'The performance and potential of protected areas', *Nature*. Nature Publishing Group, 515(7525), pp. 67–73. doi: 10.1038/nature13947.

Wearn, O. R., Freeman, R. and Jacoby, D. M. P. (2019) 'Responsible AI for conservation', *Nature Machine Intelligence*. Nature Publishing Group, 1(2), pp. 72–73. doi: 10.1038/s42256-019-0022-7.

White, H. (2009) *Theory-Based Impact Evaluation: Principles and Practice*. Available at:

http://www.3ieimpact.org/media/filer_public/2012/05/07/Working_Paper_3.pdf (Accessed: 24 May 2018).

Wiafe, E. D. and Amoah, M. (2012) 'The Use of Field Patrol in Monitoring of Forest Primates and Illegal Hunting Activities in Kakum Conservation Area, Ghana', *African Primates*, 7(2), pp. 238–246. Available at: http://static1.1.sqspcdn.com/static/f/1200343/22603725/1367508101707/Afr_Prim_72_Wiaf_+_Amoah.pdf?token=7kXOiDi8uJ36kDbhHyzRMrTODCw%3D (Accessed: 18 April 2018).

Wilfred, P., Milner-Gulland, E. J. and Travers, H. (2017) 'Attitudes to illegal behaviour and conservation in western Tanzania', *Oryx*. Cambridge University Press, pp. 1–10. doi: 10.1017/S0030605317000862.

Wilson, E. (2016) *Half-earth: our planet's fight for life*. Available at: https://books.google.com/books?hl=en&lr=&id=gft1CQAAQBAJ&oi=fnd&pg=PT7&ots=VdLCetyn2O&sig=BqDMv5FjYdH_jBLDvPqVi0oYHhQ (Accessed: 18 October 2018).

Wittemyer, G., Northrup, J. M., Blanc, J., Douglas-Hamilton, I., Omondi, P. and Burnham, K. P. (2014) 'Illegal killing for ivory drives global decline in African elephants.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(36), pp. 13117–21. doi: 10.1073/pnas.1403984111.

Zeileis, A. and Grothendieck, G. (2005) 'zoo: S3 Infrastructure for Regular and Irregular Time Series', *Journal of Statistical Software*, 14(6), pp. 1–27. doi: 10.18637/jss.v014.i06.

Appendices

1. Closed ended key informant interview questions and protected area manager questionnaire (Chapter 3).
2. Sensitivity analyses (Chapter 5).

Appendix 1

A1 Interview questions and questionnaires

A1.1 Interview questions

Table A1.1. Closed-ended interview questions administered to key informants for SMART implementation evaluation and possible responses.

Question	**Responses**
Contexts	
How supportive were the government or relevant management authority of SMART implementation?	Yes, No, Neutral
How would you describe leadership at the site? For example, do you believe the site has weak or strong leadership?	Yes, No, Neutral
How committed would you say management were to implementation?	Yes, No, Neutral
How sufficient were resources in place to implement SMART? For example, financial, staff, equipment.	Yes, No, Neutral
Data collection and entry	
What proportion of patrols are collecting SMART data? Less than half, half to 80%, 80 to 90%, 90-95%, or 95 to 100%?	0-50%,50-80%,80-90%,90-95%,95-100%
Are rangers formally trained in data collection?	Yes, No
How quickly are patrol data entered into SMART?	<1 week, 1 week-1 month, 1-3 months, 3-6 months, 6 months+
In terms of responsibility for entering patrol data into SMART, are one (or more) individuals specifically employed to spend some or all of their time entering data?	Dedicated, shared
Are data entry staff based on-site or off-site?	On, off
Are these staff formally trained in data entry?	Yes, No
Data analysis and reporting	
Are reports produced using SMART?	Yes, No
Are these reports standardised?	Yes, No
Is there a reporting schedule?	Yes, No
On average, how often are reports produced?	Weekly, bi-weekly, monthly, quarterly, bi-annually (or less)
In terms of responsibility for analysing SMART data and producing reports, is one (or more) individual specifically employed to spend some or all of their time on the activity?	Dedicated, shared, none
Are data analysis and reporting staff based on-site or off-site?	On, off, both
Do staff receive formal training in analysis and reporting?	Yes, No
Who receives reports produced using SMART?	
Rangers	Yes, No

Patrol coordinators/Section heads/Patrol leaders	Yes, No
Protected area manager	Yes, No
Which of the following information types is included in reports	
Patrol effort measures, such as number of patrols, distance patrolled, or number of days.	Yes, No
Measures of individual ranger performance. For example, measures of distance patrolled per ranger.	Yes, No
Measures of team or section performance. For example, measures of distance patrolled per team.	Yes, No
Maps of patrol routes	Yes, No
Maps of illegal activity	Yes, No
Maps of wildlife observations	Yes, No
Table of patrol activities (such as seizures, arrests and other actions)?	Yes, No
Trends in patrolling performance	Yes, No
Trends in illegal activities	Yes, No
Trends in wildlife observations	Yes, No
A patrol plan with targets/objectives for the reporting period	Yes, No
Intelligence received and acted upon	Yes, No
Recommendations for follow-up action	Yes, No
Constraints or problems encountered	Yes, No
Performance evaluation and incentives	
Is individual ranger performance measured or evaluated using SMART data?	Yes, No
Is team or section performance measured or evaluated using SMART?	Yes, No
Are results from SMART used to calculate any type of staff incentives or reward scheme?	Yes, No
If yes, which incentives?	
Ranger salaries	Yes, No
Bonuses	Yes, No
Promotions	Yes, No
Do staff or management receive formal training in using SMART to evaluate performance and develop incentives?	Yes, No
Were rangers consulted about potential incentive systems before SMART was implemented?	Yes, No
Meetings	
Are meetings held to discuss results or reports from SMART?	Yes, No
Are meetings held regularly, to a schedule?	Yes, No
How often are meetings held?	Weekly, bi-weekly, monthly, quarterly, bi-annually (or less)
Are meetings held on-site or off-site?	On, off
Who attends these meetings?	
Rangers	Yes, No

Patrol coordinators/Section heads/Patrol leaders	Yes, No
Are team leaders invited to comment on results? For example, patrol coordinators, section heads, or patrol leaders.	Yes, No
Are rangers invited to comment on results?	Yes, No
Patrol planning	
Are results from SMART used to inform patrol planning?	Yes, No
How frequently would you say this takes place? How often are patrol routes or activities reviewed in light of information from SMART?	Weekly, bi-weekly, monthly, quarterly, bi-annually (or less)
Do staff or management receive formal training in using SMART to inform patrol planning?	Yes, No
What information from SMART is used to inform patrol planning?	
Patrol effort. Such as number of patrols, distance patrolled, or number of days.	Yes, No
Patrol routes.	Yes, No
Spatial patterns in illegal activity discussed	Yes, No
Spatial patterns in wildlife observations	Yes, No
Is information from SMART used to set patrol targets for the next reporting period?	Yes, No

A1.2 Questionnaire questions

The following questions were distributed to site managers via a questionnaire for completion under the supervision of SMART technical partners, as part of SMART's implementation evaluation.

Site information

1. IUCN protected area management category (only if known/applicable – please check one)

Ia - Strict Nature Reserve	
Ib - Wilderness Area	
II - National Park	
III - Natural Monument or Feature	
IV - Habitat/Species Management Area	
V - Protected Landscape/Seascape	
VI - Protected Area with sustainable use of natural resources	
Not applicable	
Unknown	

2. Governance type (please check one)

State managed (e.g., by federal or national ministry/agency)	
Privately managed by *non-profit* organisation (NGO)	
Privately managed by *for-profit* organisation or individual	
Community managed	

Shared governance (please describe)	
Other (please specify)	

3. Most important threats to site (please check **one primary threat** and **one secondary threat**)

	Primary	Secondary
Commercial poaching		
Subsistence poaching		
Illegal logging		
Illegal harvesting of non-timber forest products (NTFPs)		
Encroachment (e.g., village expansion, clearance for agriculture, pastoralism, over-grazing)		
Illegal fishing		
Fire		
Illegal mining/extraction		
Human-wildlife conflict		
Other (please specify)		

4. Approximate size of area under management (in km^2)

5. Was the site created, or is it primarily managed, to protect any flagship, high-profile or commercially-valuable species? If yes, please specify species name/s.

SMART implementation

NB: By SMART implementation we mean both initial set-up of SMART and ongoing SMART-based activities.

SMART Software

6. a) Was there a software-based monitoring system in use prior to SMART (e.g., MIST or CyberTracker)? If yes, please specify name of system.

   b) If there was a previous software-based system, how many years was it in use for?

   c) And has monitoring information from the previous system been integrated into your SMART database? If so, please specify when data from the previous system ends and SMART data begins.

SMART Equipment

7. Which of the following items of equipment are available *on-site* and used for SMART activities? Please check all that apply.

Paper forms	
GPS unit/s (+ batteries)	
Hand-held electronic recording device/s (e.g., SMART phones or PDAs + batteries)	
Computer/s (+ power source)	
Internet connection	

SMART Personnel

8. Are staff employed, either full-time or part-time, specifically to undertake entry, management, analysis or reporting of SMART data for this site? If so, please indicate the number of individuals employed. (NB: Staff may have secondary, non-SMART duties, but we refer to individuals whose primary responsibility is SMART-related.)

No individuals are employed with specific responsibility for entry, management, analysis or reporting of SMART data	
Number of individuals employed on a part-time basis	
Number of individuals employed on a full-time basis	

9. Who employs any staff indicated in Q13? (i.e. which organisation is responsible for paying their salaries or wages?)

State or national ministry/agency	
Non-profit organisation (NGO)	
Private individual or for-profit organisation	
Community	
Other (please specify)	

SMART Technical support

10. How regularly does the site receive *on-site* technical support for SMART implementation from an NGO partner?

Permanent on-site NGO technical support for SMART implementation	
NGO visits site to provide technical support for SMART once or more per month	
NGO visits site to provide technical support for SMART once per quarter	
NGO visits site to provide technical support for SMART once every 6 months	
NGO visits site to provide technical support for SMART less than once every 6 months	
Other (please specify)	

Costs of SMART implementation

By costs of SMART implementation, we mean new and additional costs incurred to set-up and run SMART at this site, over and above usual management budgets.

11. **Capital/set-up costs**: Approximately, how much did *initial* set-up of SMART cost? Set-up costs are one-off costs to get the system started (e.g., for new equipment, new infrastructure or initial training). This is distinct from annual operating costs (see Q17).

Less than $5,000	
$5,000 to $10,000	
$10,000 to $25,000	
$25,000 to $50,000	
$50,000 to $100,000	
More than $100,000	

12. **Annual operating costs**: Approximately, what is the site's *annual* budget for SMART activities? Operating costs are recurring (e.g., salaries for SMART support staff, equipment maintenance and replacement, consumables, SMART-informed rewards or incentives, or ongoing training provision).

Less than $5,000	
$5,000 to $10,000	
$10,000 to $25,000	
$25,000 to $50,000	
$50,000 to $100,000	
More than $100,000	

Monitoring protocols

13. a) Do rangers on patrol record automatic tracklogs (e.g., using a GPS unit's automatic recording mode)?

    b) If patrols do record tracklogs (if not, skip to Q19), are they uploaded to your SMART database?

    c) If tracklogs are not uploaded to SMART, are tracklogs shareable by some other means?

14. If patrols *do not* record tracklogs, do rangers on patrol record regular waypoints even when no observations are made (i.e. manually record georeferenced locations according to a set schedule)? If so, how regularly? (please indicate interval)

Patrols only record waypoints when they observe indications of wildlife or human activity	
Patrols record waypoints per a time schedule (e.g., every 5/15/30 mins) (please specify time interval)	
Patrols record waypoints per a distance schedule (e.g., every 500 m/1km) (please specify distance interval)	
Patrols record waypoints without observations but not to a defined schedule	
Other (please specify)	

15. How do rangers record SMART observations (e.g., of wildlife or human activity)?

Paper forms (+ GPS unit for waypoint)	

Hand-held mobile data recording device (smartphone or other)	
Other	

Law enforcement

These questions relate to all law enforcement and protection activities at the site, not just those involving SMART.

16. How frequently are enforcement patrols conducted?

Once or more per week	
Once per fortnight	
Once per month	
Less than once per month	
Other (please specify)	

17. How many field rangers are employed at the site (i.e. rangers primarily responsible for enforcement patrolling)?

18. How many rangers or staff with powers of arrest are employed at the site (e.g., powers of detention necessary until the suspect has been handed over to the police)?

19. How frequently do rangers attend continuation training courses for enforcement and protection? (please check one)

At least once per quarter	
At least once every six months	
At least once per year	
Less frequently than once per year	
Other (please specify)	

20. Approximately, what percentage of patrol staff:

Are issued patrol boots?	
Are issued with radios during patrols?	
Are issued with a weapon during patrols?	

21. Annual operating costs: Approximately, what is the site's *annual* budget for law enforcement activities? Operating costs can include, for example, rangers' salaries, travel costs (e.g., fuel), equipment purchases, annual enforcement training and administrative costs.

Less than $50,000	
$50,000 to $100,000	
$100,000 to $250,000	
$250,000 to $500,000	
More than $500,000	

Appendix 2

A2 Sensitivity analyses

A2.1 Time lags

		(t-1) lag						MA2 lag						MA3 lag					
Site	Step	Slope	SE	*F*	DF	*P*	r^2	Slope	SE	*F*	DF	*P*	r^2	Slope	SE	*F*	DF	*P*	r^2
Forest-dominated sites																			
	14	-0.000038	0.0028	0.00018	(1,36)		0.00	-0.0051	0.0079	0.42	(1,35)		0.01	0.0014	0.0079	0.032	(1,34)		0.00
1	28	0.0044	0.0055	0.64	(1,17)		0.04	0.0043	0.012	0.13	(1,16)		0.01	-0.019	0.016	1.4	(1,15)		0.09
	42	-0.0055	0.0061	0.81	(1,10)		0.08	-0.034	0.0081	18	(1,9)	**	0.66	-0.021	0.024	0.79	(1,8)		0.09
	14	0.0096	0.0054	3.1	(1,47)	.	0.06	-0.011	0.015	0.56	(1,46)		0.01	0.038	0.018	4.8	(1,45)	*	0.10
2	28	-0.014	0.01	1.9	(1,21)		0.08	-0.023	0.016	2	(1,20)		0.09	0.0043	0.029	0.022	(1,19)		0.00
	42	-0.038	0.018	4.6	(1,13)	.	0.26	-0.017	0.035	0.25	(1,12)		0.02	-0.012	0.051	0.054	(1,11)		0.00
Savannah-dominated sites																			
	14	-0.00028	0.0041	0.0046	(1,101)		0.00	-0.0069	0.0068	1	(1,100)		0.01	-0.0079	0.0095	0.68	(1,99)		0.01
3	28	-0.0097	0.0034	7.9	(1,49)	**	0.14	-0.0032	0.0061	0.27	(1,48)		0.01	-0.0042	0.0078	0.29	(1,47)		0.01
	42	0.0064	0.0039	2.7	(1,31)		0.08	0.0018	0.0059	0.089	(1,30)		0.00	0.0089	0.0075	1.4	(1,29)		0.05
	14	-0.0064	0.023	0.075	(1,39)		0.00	-0.042	0.039	1.2	(1,38)		0.03	-0.00063	0.053	0.00014	(1,37)		0.00
4	28	-0.0067	0.015	0.18	(1,17)		0.01	0.014	0.025	0.31	(1,16)		0.02	-0.00096	0.057	0.00029	(1,15)		0.00
	42	0.018	0.018	1	(1,10)		0.09	0.062	0.059	1.1	(1,9)		0.11	-0.026	0.1	0.063	(1,8)		0.01

Figure A2.1. Regression output for differenced plots of CPUE snare occupancy over patrol effort, with timesteps of 14, 28 and 42 days, the default time lag (t-1), and lags with effort calculated over a moving average of two and three timesteps (MA2 and MA3, respectively), for 2 forest-dominated and 2 savannah-dominated sites.

		(t-1) lag						MA2 lag						MA3 lag					
Site	Step	Slope	SE	*F*	DF	*P*	r^2	Slope	SE	*F*	DF	*P*	r^2	Slope	SE	*F*	DF	*P*	r^2
Forest-dominated sites																			
	14	0.0013	0.0022	0.37	(1,36)		0.01	-0.0011	0.0064	0.031	(1,35)		0.00	0.0039	0.0066	0.35	(1,34)		0.01
1	28	0.0006	0.0024	0.061	(1,17)		0.00	0.0007	0.0049	0.021	(1,16)		0.00	-0.011	0.0065	3	(1,15)		0.17
	42	0.00056	0.0026	0.046	(1,10)		0.00	-0.0082	0.0052	2.5	(1,9)		0.21	-0.014	0.012	1.4	(1,8)		0.15
	14	-0.0051	0.0076	0.46	(1,47)		0.01	0.0034	0.02	0.029	(1,46)		0.00	-0.0093	0.025	0.14	(1,45)		0.00
2	28	-0.0075	0.02	0.14	(1,21)		0.01	-0.026	0.03	0.75	(1,20)		0.04	-0.0074	0.052	0.02	(1,19)		0.00
	42	-0.064	0.027	5.7	(1,13)	*	0.30	-0.092	0.049	3.6	(1,12)	.	0.23	-0.084	0.078	1.2	(1,11)		0.09
Savannah-dominated sites																			
	14	0.0015	0.0038	0.15	(1,101)		0.00	0.01	0.0061	2.7	(1,100)		0.03	0.0053	0.0082	0.42	(1,99)		0.00
3	28	0.0016	0.0023	0.5	(1,49)		0.01	0.0037	0.0037	1	(1,48)		0.02	-0.00035	0.0046	0.0058	(1,47)		0.00
	42	0.0026	0.0031	0.72	(1,31)		0.02	0.0026	0.0043	0.36	(1,30)		0.01	0.0041	0.0044	0.86	(1,29)		0.03
	14	-0.039	0.015	6.9	(1,39)	*	0.15	-0.018	0.027	0.43	(1,38)		0.01	-0.058	0.036	2.7	(1,37)		0.07
4	28	-0.0028	0.0094	0.086	(1,17)		0.01	-0.0063	0.015	0.17	(1,16)		0.01	-0.018	0.033	0.28	(1,15)		0.02
	42	-0.0012	0.011	0.012	(1,10)		0.00	0.017	0.034	0.24	(1,9)		0.03	-0.0068	0.058	0.014	(1,8)		0.00

Figure A2.2. Regression output for differenced plots of CPUE people occupancy over patrol effort, with timesteps of 14, 28 and 42 days, the default time lag (t-1), and lags with effort calculated over a moving average of two and three timesteps (MA2 and MA3, respectively), for 2 forest-dominated and 2 savannah-dominated sites.