Integrated Natural Resources Management (INRM)

Impact Evaluation Feasibility Assessment of the USAID/Zambia Eastern Kafue Nature Alliance Activity

Last updated: May 27, 2022

Activity Title:	Integrated Natural Resource Management
Sponsoring USAID Office:	Center for Environment, Energy, and Infrastructure
Contract Number:	7200AA20F00010
Task Order Number:	STARR II – 7200AA18D00020
Period of Performance:	July 30, 2020 – July 19, 2025
Contracting Officer:	Stella Alexander-Sergeeff
COR:	Ioana Bouvier
Ceiling Price:	$34,976,131
Contractor:	DAI
Date of Publication:	Draft submitted for review and approval 27 May 2022
Authors:	Heather Huntington, Thomas Morrison, and Christina Seybolt
Front Cover photo:	Kafue National Park https://piperandheath.com/zambia-kafue-national-park/

Submitted by:
Robin Martino, Chief of Party
Integrated Natural Resource Management
DAI
7600 Wisconsin Ave, Suite 200
Bethesda, MD, 20814, USA
Telephone: 301.771.7600
Fax: 301.771.7777
www.dai.com

This document was produced for review by the United States Agency for International Development. It was prepared with support from the Integrated Natural Resource Management Task Order, under the Strengthening Tenure and Resource Rights II (STARR II) IDIQ.

Contents

Table of Figures

Table of Tables

Acronyms

ADS	Automated Directive Systems
CFMG	Community Forest Management Group
CHW	Community Health Worker
CRB	Community Resource Board
CWP	Community Wildlife Protection
DHS	Demographic and Health Surveys
DiD	Difference-in-Differences
DNA	Deoxyribonucleic Acid
DNPW	Department of National Parks and Wildlife
eDNA	Environmental Deoxyribonucleic Acid
FA	Feasibility Assessment
GESI	Gender Equity and Social Inclusion
GhG	Greenhouse Gas
GIS	Geographic Information System
GKE	Greater Kafue Ecosystem
GMA	Game Management Area
GPS	Global Positioning System
GRI	Game Rangers International
HEARTH	Health, Ecosystems, and Agriculture for Resilient Thriving Societies
IE	Impact Evaluation
INRM	Integrated Natural Resource Management
IUCN	International Union for Conservation Nature
IP	Implementing Partner
MCH	Maternal and Child Health
MDES	Minimum Detectable Effect Sizes
MERL	Monitoring, Evaluation, Research, and Learning
M&E	Monitoring and evaluation
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetative Index
PE	Performance Evaluation
RRT	Randomized Response Technique
SA	Strategic Approach
SE	Standard Errors
TOC	Theory of Change
TNC	The Nature Conservancy
USAID	United States Agency for International Development
WCP	Wildlife Crime Prevention

Executive Summary

Objective

The objective of this feasibility assessment (FA) is to assess the possible evaluation options for the **Health, Ecosystems, and Agriculture for Resilient Thriving Societies (HEARTH) Eastern Kafue Nature Alliance activity ("Kafue Activity") in Zambia**. The assessment considers design options, including impact evaluation (IE) and performance evaluation (PE), that meet Agency-wide HEARTH and Mission learning interests, with the goal of determining the most rigorous options that can be applied given implementation, resource, and other constraints for this activity.

Kafue Activity Overview

The Kafue Activity will be implemented in several Game Management Areas and conservancies that border the eastern side of Zambia's Kafue National Park, by a consortium of public and private sector partners led by The Nature Conservancy. High poverty rates in the area lead to a dependence on natural resources and income from forests for many households, contributing to deforestation and forest degradation from wood extraction, agricultural expansion, and fires. To address these issues and impact both conservation and human well-being outcomes, the Kafue Activity is comprised of **four strategic approaches** (SAs):

- SA1: Strengthen natural resource compliance and management systems
- SA2: Develop inclusive ecosystem-based markets for local prosperity
- SA3: Strengthen community maternal and child health and improve access to clean water
- SA4: Develop effective land and resource use planning, tenure, and governance systems

Evaluation Approaches Considered

IEs measure the causal impact of a program, or the difference in outcomes caused by the program and not by other external factors. While assessing the feasibility of an IE was one of the primary objectives of this assessment, at USAID's direction, the FA team considered PE approaches. In particular, a variety of evaluation methods were considered including (1) **experimental approaches**, which measure the causal impact of programs through randomized assignment (e.g., randomized control trials); (2) **quasi-experimental**, which also attempt to measure causal impacts but without randomization (e.g., difference-in-difference [DID] and statistical matching); and (3) **non-experimental approaches**, which can answer descriptive questions about differences but cannot measure causality with the same degree of rigor or confidence. Non-experimental approaches include PEs, which generally include before-after comparisons without a rigorously defined counterfactual, and case studies, which include in-depth learning from an instance through extensive description and analysis. A mixed-method evaluation integrates two or more evaluation methods, usually drawing on both quantitative and qualitative data. Generally, mixed methods evaluations can provide a deeper understanding of why change is or is not occurring and capture a wider range of perspectives.

Summary of Findings

The FA team finds that the Kafue Activity presents an important opportunity to improve USAID's understanding of conservation and biodiversity programming through a **mixed methods evaluation, including both IE and PE components.** Specifically, several components of the program being implemented in Year 1 (SA1 resource protection and SA2 agricultural markets and out-grower support) might be amenable to evaluation through a quasi-experimental difference-in-difference (DiD) approach, with matching to improve rigor. DiD is a quasi-experimental evaluation design that estimates programmatic impact by comparing (1) changes in outcomes among program participants with (2) changes in outcomes among non-participants and is one of the most commonly used designs for IEs. For some individual SAs, experimental methods like a randomized lottery around eligibility cut-off for health and water (SA3) activities can be further explored after the needs assessments and situational analyses have occurred. It should be emphasized that final decisions about the evaluation design and methodology for activities under SA3, SA4, and some components of SA1 and SA2 can only be made when the interventions and sites are determined at the end of Year 1.

Given the dearth of counterfactual-based studies on the Kafue Activity strategic approaches, even knowledge generated through a well-designed PE for some components would advance USAID's and the HEARTH portfolio's learning agenda. Furthermore, an evaluation would add value by strengthening the program's theory of change and promoting a deeper understanding of where to focus on intervention integration and quality. Baseline data will provide a key source of monitoring and evaluation (M&E) data and provide important contextual information that can be used to promote more effective, adaptive programming.

In addition, the Kafue Activity presents a **unique opportunity to measure the effect of conservation programming on biodiversity outcomes**. This is due to the large amount of biodiversity monitoring that will take place as part of the program implementation; this large-scale wildlife monitoring makes it feasible to pursue a cost-effective, rigorous, and long-term study of biodiversity outcomes. Extensive observation data collection through a combination of SMART monitoring, spoor surveys, and camera traps may provide the necessary data to measure biodiversity outcomes in the context of an IE approach.

Recommendations

In addition to the key findings above, the FA team recommends the followings to USAID:

- **Baseline data collection at the end of Year 1:** Given the phased implementation design, the FA team recommends waiting until the end of Year 1/ beginning of Year 2 to conduct one comprehensive baseline household data collection effort after all activities and locations have been determined.
- **Need for Pause and Reflect:** The site locations and content of most interventions will not be finalized until the end of Year 1. The FA team recommends a series of regular coordination and information exchange meetings as implementation information becomes available. In addition, at the end of Year 1 the MERL plan will need to be updated and there should be a Pause and Reflect of all stakeholders.
- **Biodiversity Measures:** Overall, we recommend a combination of approaches for monitoring biodiversity outcomes that leverages existing datasets and data streams from partners with new data sources. Remote sensing data, particularly forest cover, provides rich and readily available proxies for biodiversity, and as well as important habitat outcomes. Additionally, the most likely direct biodiversity indicators will involve changes in wildlife behavior or spatial distributions near treatment sites. Camera traps will provide an efficient balance between cost and field effort, yielding high-quality data for a broad diversity of large mammals, and the long record of aerial censuses and recent SMART monitoring activities make valuable baselines for understanding biodiversity outcomes for common large-bodied species.
- **Long term evaluation:** The FA team recommends a long-term evaluation, including follow-up data collection about five years after the end of the activity. The primary biophysical outcomes of interest will take a longer time to materialize than the standard USAID program cycle.
- **Strong coordination and collaboration are required throughout design and implementation:** A rigorous evaluation will require detailed M&E tracking of inputs, outputs, and specific site locations, along with significant coordination among the IPs and between the IPs and the evaluation team, to ensure the design is appropriate as implementation plans evolve.
- **USAID and IP focus on integration and quality of programming**: Integration is a key underlying assumption for the whole of project theory of change, as well as for the theories of change for several SAs. Thus, site selection for activities should prioritize overlapping implementation to the extent possible, to answer key learning questions related to integrated programming.

I. Introduction and Background

The following section introduces the assessment, including the primary objectives, purpose, audience and intended users, and information sources, as well as some background and context for the Kafue Activity.

Introduction

The objective of this feasibility assessment is to assess the possible options for a rigorous evaluation of the effectiveness of the Health, Ecosystems, and Agriculture for Resilient Thriving Societies (HEARTH) Global Development Alliance Eastern Kafue Nature Alliance activity ("Kafue Activity") in Zambia (five years, beginning in late 2021). This assessment, conducted under the Integrated Natural Resource Management Task Order, includes identifying illustrative impact evaluation (IE) and/or performance evaluation (PE) design options that meet Agency-wide HEARTH and Mission learning interests and are considered feasible for a credible assessment of impacts, should the United States Agency for International Development (USAID) decide to conduct an IE of the activity. This report will provide an assessment of the Kafue Activity's current theory of change (TOC), evaluation design options and potential methods, challenges, and limitations to conducting an IE, potential outcomes and data sources, illustrative costs, and next steps.

Purpose, Audience, and Intended Uses

USAID has commissioned this assessment to conduct a desk-based feasibility assessment of IE design options that could be used to rigorously evaluate the impacts of the Kafue Activity. The feasibility assessment will help to inform broader development of design options for a rigorous IE, if USAID decides it would like to conduct such an activity. The primary audiences for the IE feasibility assessment are USAID/Zambia, USAID/Bureau for Africa/Office of Sustainable Development, and USAID/Environment, Energy, and Infrastructure/Natural Environment. Secondary audiences include the implementing and private sector partners for the Kafue Activity. USAID will use the results of this IE feasibility assessment to gain an understanding of available design options and methods that could be used for an IE of the Kafue Activity, the types of outcomes that could be measured under such designs, the additional information that would be required to proceed with an IE design, and an illustrative indication of costs.

Background and Context

Zambia's Greater Kafue Ecosystem (GKE) is about 80 miles west from Lusaka, Zambia's capital, as the crow flies. The GKE consists of Kafue National Park, which is Zambia's oldest and largest national park, and nine Game Management Areas (GMAs). GMAs are co-managed by the Department of National

Parks and Wildlife (DNPW) and communities through Community Resource Boards (CRBs), each with varying capacity to govern the surrounding landscapes. The GKE covers about 67,000 sq km and suffers from high poaching rates, unsustainable forest clearing, and rampant fires.[1] The ecosystem is also home to 200,000 people (about half of which reside in the Kafue Activity project area), mostly smallholder farmers, but livelihoods are also based on fishing and cutting trees to produce charcoal, tourism, mining, and timber (see Figure 1). For more detailed information including the overlap of GMAs and administrative districts, areas of high conservation value, and human settlements, see Annexes 1, 2, and 3.

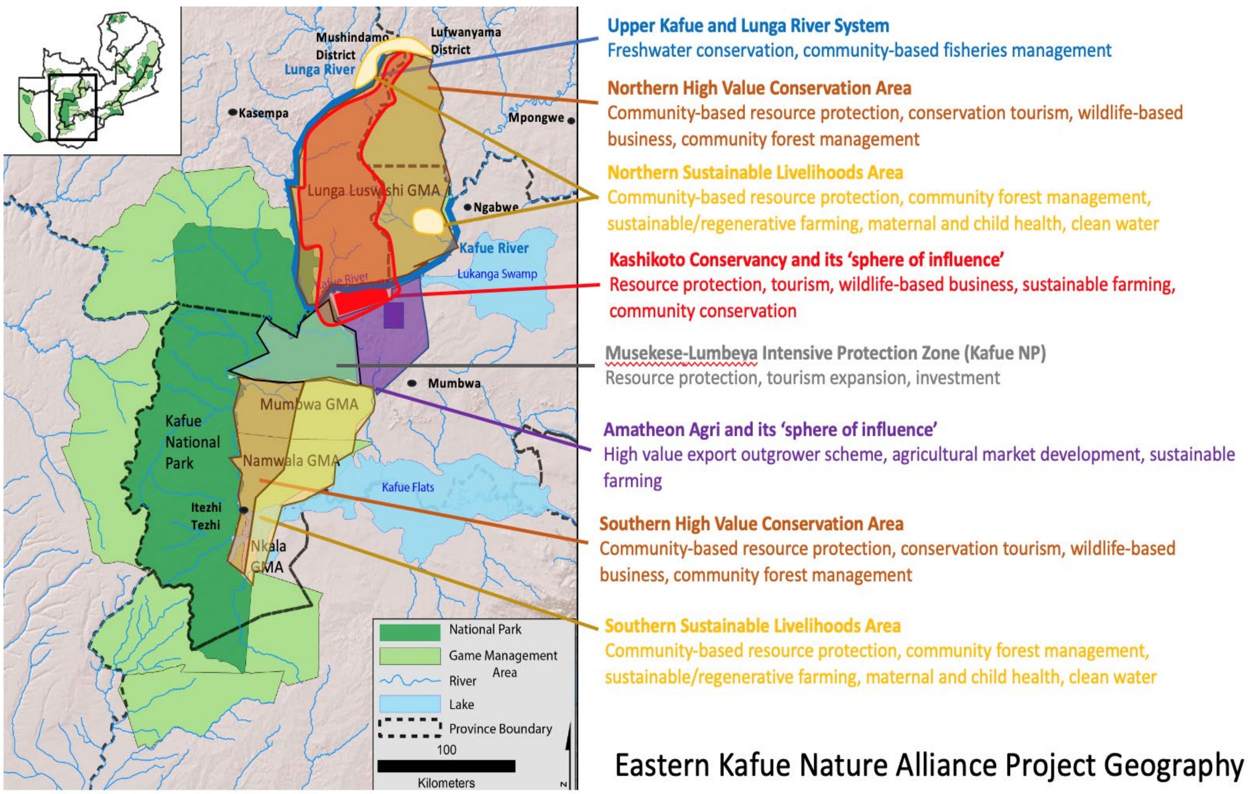


Figure 1. *Eastern Kafue Nature Alliance Project Geography2.*

In rural Zambia, poverty leads to a high dependence on natural resources and income from forests for day-to-day survival.[3] Many households live without electricity or public water and sanitation services and are largely reliant on charcoal production and subsistence agricultural activities for livelihoods. According to the most recent World Bank data, 13.9 percent of the rural population has access to electricity,[4] and only 48 percent of the rural population has basic drinking water services.[5] The primary

[1] USAID, Impact Evaluation Feasibility Assessment of the Eastern Kafue Nature Alliance Activity: Statement of Work (2021).
[2] Shared by TNC on January 21, 2021
[3] Bwalya, Samuel M., "Household Dependence on Forest Income in Rural Zambia," *Zambia Social Science Journal 2*, no. 1, (2011): 6, https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=1021&context=zssj.
[4] "Access to Electricity (% of Population)," The World Bank, accessed February 17, 2022, https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=ZM.
[5] "People using at least basic drinking water services, rural (% of rural population) – Zambia," The World Bank, accessed February 17, 2022, https://data.worldbank.org/indicator/SH.H2O.BASW.RU.ZS?locations=ZM.

drivers of deforestation and forest degradation across Zambia and the GKE include wood extraction, agricultural expansion, and fires.[6] Wood extraction encompasses logging, collection of fuel wood, and charcoal production. Charcoal production is a significant driver of deforestation, providing livelihoods for producers in rural areas and driven by high demand for charcoal from urban areas due to the lack of affordable and reliable alternative energy sources.[7] For rural households, charcoal production is an important risk-mitigation strategy, as subsistence farmers use additional income from charcoal during the lean season after harvest and in the face of increasing droughts.[8]

Agricultural expansion is another primary driver of deforestation in Zambia. Subsistence agriculture is the main source of food and income for many Zambians, including those in the GKE.[9] The use of unsustainable cultivation practices, such as slash-and-burn and overgrazing, increase the time needed for deforested land to regenerate. Fires are frequently used in Zambia to hunt wild game, clear fields for cultivation, control brush, and manage pastures. These fires are often not well managed, and wildfires, particularly late in the dry season, can be devastating to forest cover, as they slow the regeneration and survival of young plants.[10] For a detailed map of natural resource-based activities, see Annex 4.

The Kafue Activity will be implemented by public and private sector partners, Kashikoto Conservancy Limited (Kashikoto), Amatheon Agri Zambia (Amatheon), The Nature Conservancy (TNC), Game Rangers International (GRI), Panthera, Musekese Conservation, Wildlife Crime Prevention (WCP), and i4Life. For a detailed map of partner project areas, see Annex 5. The Kafue Activity consists of four Strategic Approaches (SAs), which are discussed in more detail in Section II below.

Table 1. *Kafue Activity Strategic Approaches with Expected End of Project Outcomes*[11]

Strategic Approach (SA)*	Expected Outcome
SA1: Strengthen natural resource compliance and management systems	900,000 hectares have improved biophysical condition
SA2: Develop inclusive ecosystem-based markets for local prosperity	10,000 people (including 30 percent youth) have increased income from sustainable enterprises
SA3: Strengthen community Maternal and Child Health (MCH) and improve access to clean water	5,000 women have improved access to MCH services and 20,000 people have improved access to safe drinking water

[6] Vinya, R., Syampungani, S., Kasumu, E.C., Monde, C. & Kasubika, R., "Preliminary Study on the Drivers of Deforestation and Potential for REDD+ in Zambia," *Forestry Department and FAO, National UN-REDD+ Programme Ministry of Lands & Natural Resources*, (2011), https://www.unredd.net/documents/un-redd-partner-countries-181/africa-335/zambia-182/studies-reports-and-publications-526/8023-preliminary-study-on-the-drivers-of-deforestation-and-potential-for-redd-in-zambia-8023.html.
[7] Vinya, Kasumu, Syampungani, Monde, and Kasubika, 2011.
[8] Gonzalez, Ahtziri, "Engaging Zambian Charcoal Producers in Sustainability Efforts," *Forest News*, August 31, 2021, https://forestsnews.cifor.org/74333/engaging-zambian-charcoal-producers-in-sustainability-efforts?fnl=en.
[9] Ministry of Tourism, Environment and Natural Resources, 2002.
[10] Livestock grazing itself is not a threat to conservation (and may even be symbiotic if managed properly), but associated behaviors such as over-grazing, starting fires, setting snares, etc. as well as human-wildlife conflict have a negative impact. There is overlap between conservation areas and cattle grazing, which is not permitted under the Game Management Plans. Cattle are primarily kept by communities in the Mumbwa, Namwala, and Nkala GMAs, with livestock exerting a significant influence in the Nkala GMAs.
[11] This information is from the December 2021 Draft Monitoring, Evaluation, Research, and Learning (MERL) Plan.

Strategic Approach (SA)*	Expected Outcome
SA4: Develop effective land and resource use planning, tenure, and governance systems	1 million hectares of community lands are under improved management

Information Sources

The Feasibility Assessment team ("FA team") consulted a variety of documents and other information sources to inform this assessment, including:

- Discussions with USAID/Zambia, TNC, and other Implementing Partners (IPs);
- Detailed logic model, TOC, and results chains specific to the Kafue Activity;
- Draft Year 1 Workplan which outlines the nature of ongoing and planned activities for the Kafue Activity, and general anticipated timing of key activities;
- Draft MERL plan for the Kafue Activity;
- Participation in the Activity Start-Up Workshop (facilitated by Measuring Impact II);
- Documentation or description of any key criteria that may have been used to select beneficiaries or areas identified to receive the program;
- Information on which implementing partners are working where and when via the draft Workplan and additional maps provided by the implementing partners (to the extent available);
- Details on the planned interventions provided by the implementing partners (to the extent available);
- Geospatial datasets of activity locations and area of intervention boundaries;
- Other secondary data and implementation information from USAID and implementing partners, including shapefiles and boundaries or point locations of communities. This includes analysis of remote sensing data available on forest cover, vegetation, and fire trends for the program area, as well as Demographic and Health Survey (DHS) data for relevant provinces; and
- Literature review of academic and grey literature on measuring biodiversity and forest conservation, as well as evaluations of wildlife monitoring, sustainable livelihoods, and conservation enterprises.

II. Intervention and Theory of Change

The following section provides further details on the interventions planned under each of the Kafue Activities' SAs, including the implementing partners and timing of activities. This is followed by a discussion of the whole of project TOC and specific results chains for each SA, with a focus on the key underlying assumptions and potential weaknesses of each. The FA team recommends that the whole of project TOC be updated, along with individual SA results chains, at the end of project Year 1 to reflect the final implementation plans (which are still to be determined for many activities). Finally, there is a high-level literature review of conservation and biodiversity evaluations, which finds that there are few well-designed counterfactual-based studies, especially which assess the effects of programming on both conservation and poverty reduction. Therefore, an evaluation of the Kafue Activity presents an important opportunity to advance USAID's learning interests about programming and policies designed to improve conservation and biodiversity outcomes.

Intervention Details

The below section is based on information provided in the Year 1 Workplan and the map of the settlement areas where each implementing partner will be conducting their activities provided by TNC. According to the workplan, the primary interventions in Year 1 are beginning activities for SA1 (resource protection/enforcement and wildlife crime monitoring) and SA2 (conservation agriculture/out-grower schemes and tourism investments). The districts and settlements for these interventions have largely been identified, and many of these activities have already begun in the first quarter of Year 1 or are continuations of on-going activities from the private sector (Kashikoto and Amatheon) and non-governmental (WCP) partners.

The remaining interventions that comprise the Kafue Activity will not begin implementation on the ground until Year 2, with Year 1 being focused on needs assessments, situational analyses, etc. that will inform the specific intervention details and locations. Specifically, SA3 (community health workers and boreholes), SA4 (natural resource governance) as well as some activities under SA1 (investigation/prosecution guidelines and awareness raising) and SA2 (conservation enterprises) will not begin until Year 2 of the program. A simplified version of the Year 1 Workplan is provided in Annex 6, and a summary for each strategic approach is below including the main activities, implementing partners (IPs), and timeline:

SA1: Strengthen natural resource compliance and management systems. SA1 includes three main components: resource protection/law enforcement (e.g., increasing scouts and anti-poaching patrols, resource protection infrastructure/equipment, and fire management teams), investigation/prosecution of wildlife crime (e.g., courtroom and prison monitoring, development of investigation/prosecution guidelines, media/awareness raising), and developing systems for monitoring biodiversity/wildlife (e.g., aerial surveys, biodiversity assessments, focal species monitoring and protection plans).[12] All activities are either on-going from before the Kafue Activity or will launch in Year 1. Specifically, for the investigation/prosecution of wildlife crime activities, monitoring cases will be on-going but not technically funded by the Kafue Activity until Year 2.[13] SA1 resource protection/law enforcement activities will be implemented primarily by Kashikoto, Muskese Conservation, and GRI in their respective areas; investigation/prosecution of wildlife crime activities will be implemented by WCP; and monitoring biodiversity/wildlife will be conducted by Kashikoto and Panthera. SA1 activities will take place across the entire project area, with resource protection/law enforcement targeting specific districts based on the IP, and investigation/prosecution and monitoring occurring across the whole of project area at the relevant subordinate courts.

SA2: Develop inclusive ecosystem-based markets for local prosperity. For SA2, there are three main components: conservation enterprises (e.g., community outreach managers supporting alternative livelihoods, support for wildlife-based business models/relationships), conservation agriculture/out-grower schemes (e.g., farmer training and contracting), and support for tourism (e.g., investment and opportunity assessments). Agricultural activities are implemented by Amatheon and focused in the Mumbwa district and have already begun in Year 1 with the late 2021/early 2022 growing season. This includes supporting crop diversification into high-value, climate-smart crops (e.g., quinoa, chili) and the provision of market services, including input finance, extension services, and demonstration of sustainable farming practices. Amatheon guarantees the offtake of harvest at a fixed price stipulated in a contract prior to the growing season and is then able to connect smallholder farmers to high-level value chains. Tourism and conservation enterprise support from Kashikoto will be focused in Mumbwa and Kasempa and will begin in Year 1, with other support for these components from TNC still to be determined in Year 2.

SA3: Strengthen community MCH and improve access to clean water. SA3 is comprised of two main components: training community health workers (CHWs) for improved MCH and the construction or repair of boreholes. The CHW activities will be implemented by i4Life, and the borehole construction contracts are still to be determined (although, the site assessment will be done by TNC). Year 1 will consist of a baseline health assessment and a borehole site assessment to inform the specific locations/settlements for these activities in Year 2, as well as to inform the health focus of the CHW. The CHW activities will be centered in/around the northern project districts and the Lunga-Luswishi GMA.

SA4: Develop effective land and resource use planning, tenure, and governance systems. SA4 relates to the improved natural resource management and governance and includes activities such as conducting a situation analysis and engaging in community conservation planning, establishing new conservation governance structures (at least one Community Forest Management Group [CFMG] and one CRB), conducting a needs assessment to inform improvements in capacity for at least 10 existing CRBs, and identifying needs for developing/enforcing land use and resource management plans. The

[12] SA1 also includes the reintroduction of key antelope species by Kashikoto.

[13] Activities that WCP is currently undertaking primarily includes the monitoring of wildlife crime court cases in Itezhi-tezhi, Kaoma, Kasempa, Mumbwa, and Namwala Subordinate Courts around Kafue National Park.

situation analysis and needs assessments will begin in Year 1 but most activities are expected to begin in Year 2. TNC is the primary implementer.

Theory of Change

Initial results chains for each SA and the whole of project were developed during the co-design phase. These results chains were subsequently updated during the Kafue Activity start-up workshop, and further refined by the FA team, to add high-level summaries of the activities/outputs under each SA, as well as add more intermediary results, and finally reviewed by TNC to ensure that they reflect their understanding of the program logic. The results chains for each SA are provided below (see Figure 2 through Figure 5), and an overall TOC is provided in Annex 7.

The project's TOC is informed by the content (or substance) of its activities combined with the geography of their implementation. The original TOC for the Kafue Activity assumes the overlap of multiple different interventions. However, to-date, there is (1) limited geographic overlap for all the SAs, (2) uncertainty about the specific locations for many interventions, and (3) uncertainty about the content of the specific intervention activities for some SAs that have not been finalized. According to the project timeline, the details of the interventions across the SAs will be finalized by the end of project Year 1, along with the implementation plan, including locations for the interventions.

The FA team recommends that the whole of project TOC be updated, along with individual SA TOCs, at the end of project Year 1 to reflect the final implementation plans. Revisions to the TOC will likely require updates to the evaluation questions, methodology, and indicators as appropriate. For example, the health needs assessment will determine what aspects of health the CHWs will be trained in/provide services in, and the related indicators that should be updated to reflect the health outcomes that they would reasonably affect.

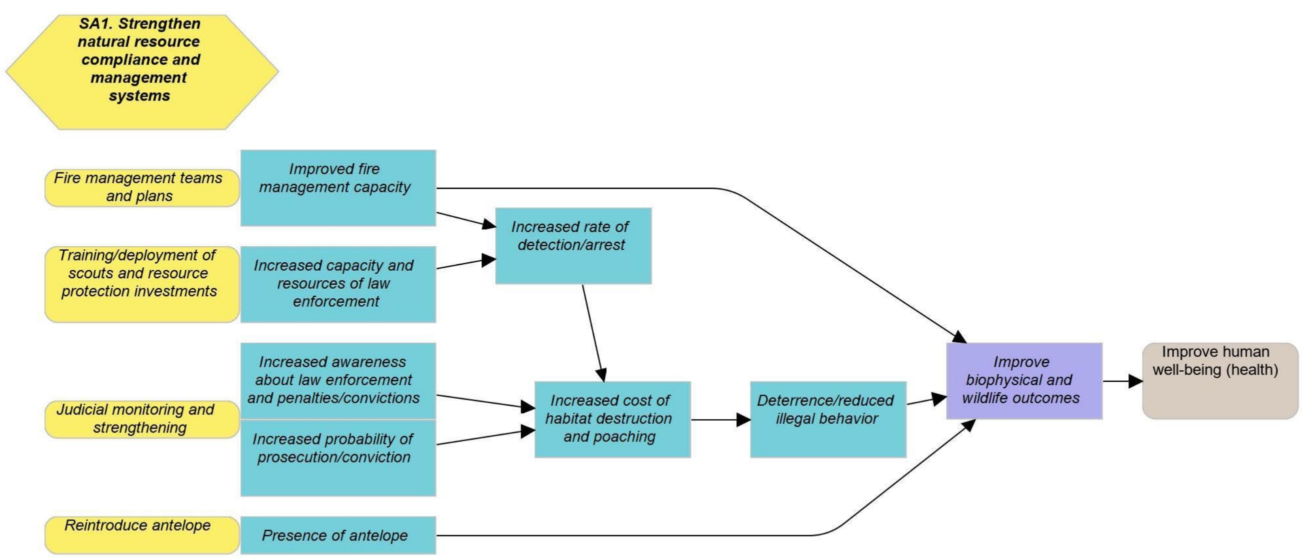


Figure 2. *Results Chain for SA1: Natural Resource Compliance and Management.*

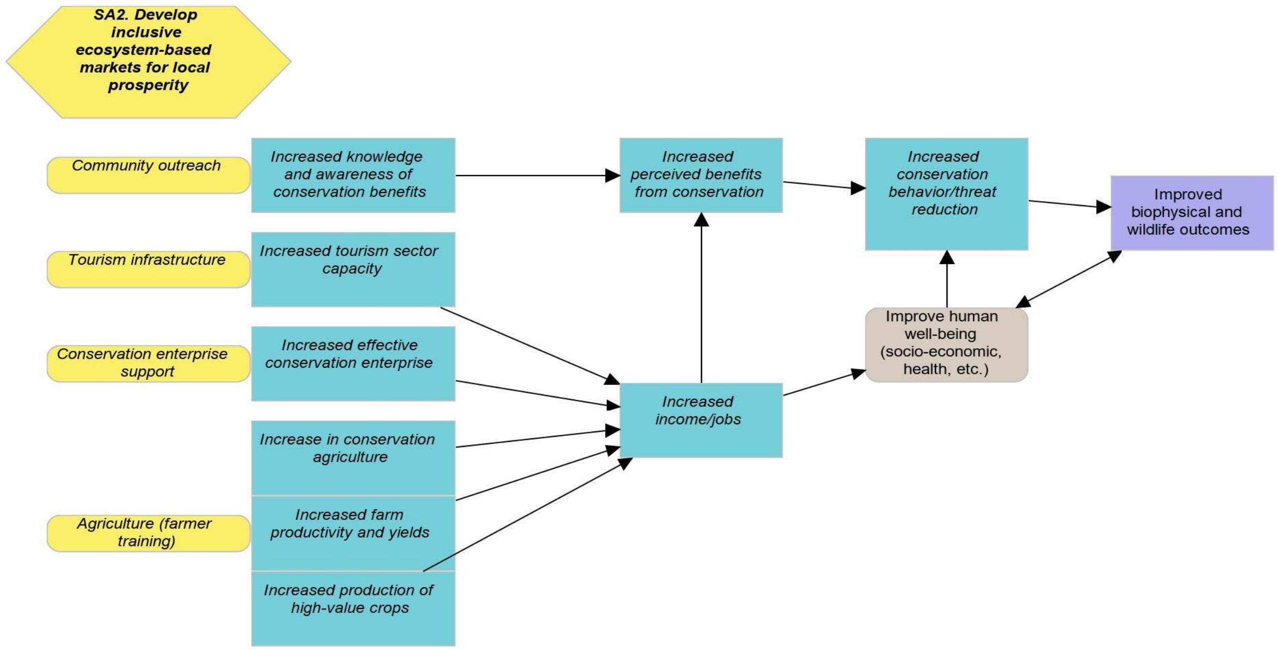


Figure 3. *Results Chain for SA2: Inclusive Ecosystem-based Markets.*

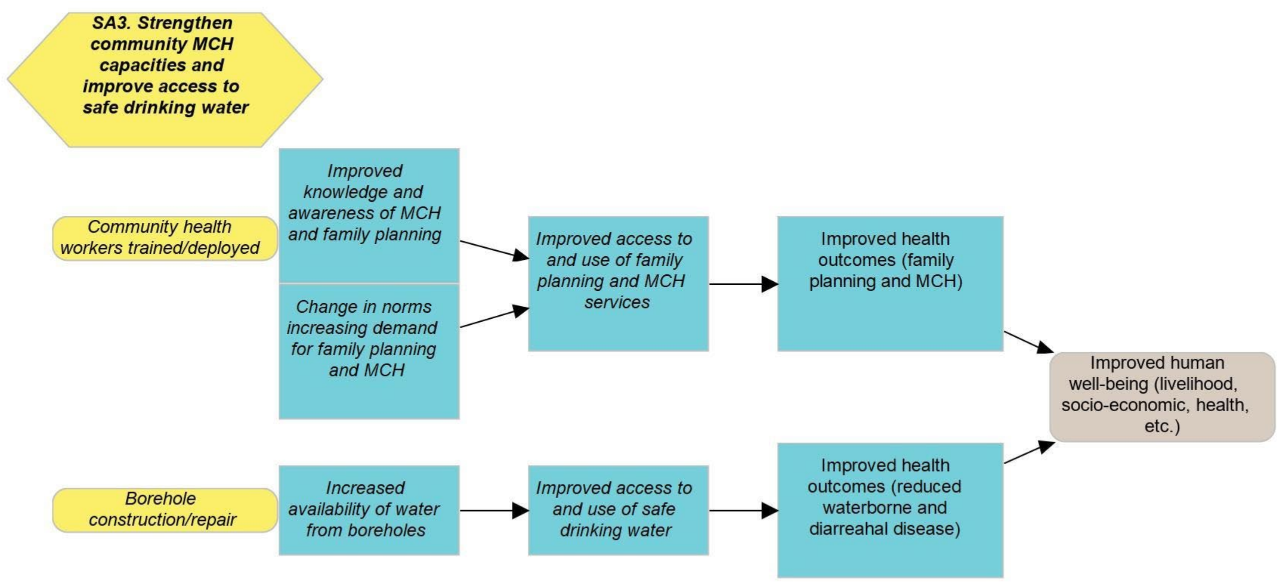


Figure 3. *Results Chain for SA3: Community MCH and Access to Safe Drinking Water.*

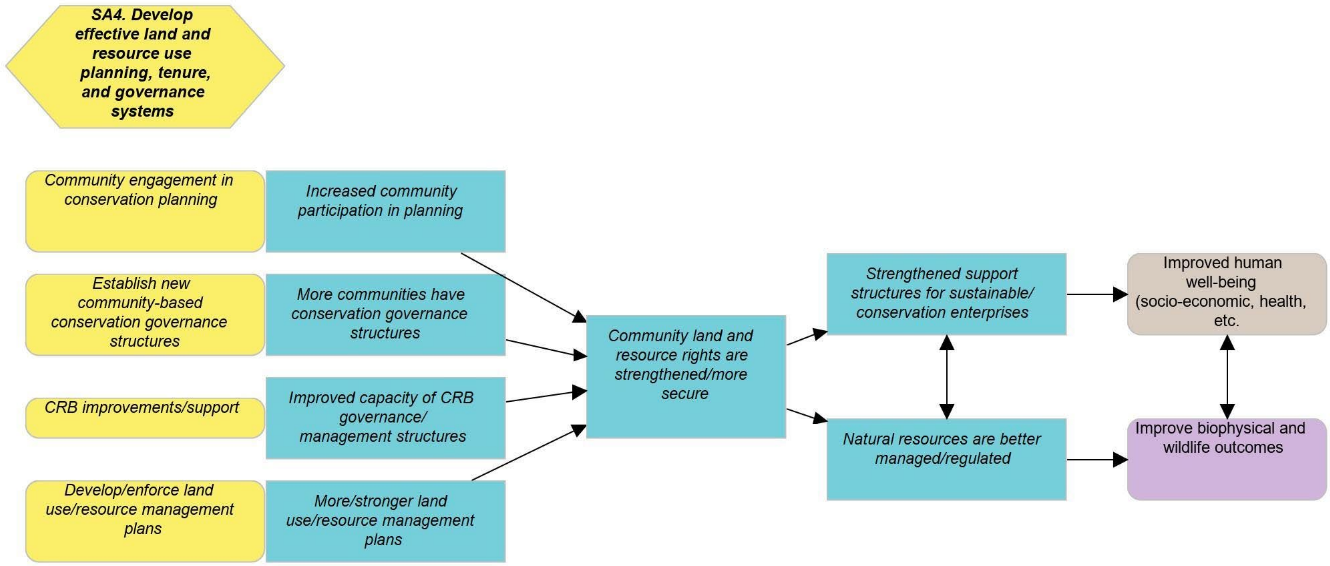


Figure 4. *Results Chain for SA4: Land and Resource Use Planning, Tenure, and Governance.*

Key Assumptions

TNC prepared a list of key assumptions underlying the TOC in the draft MERL plan. These assumptions were further discussed during the activity start-up workshop, and the FA team has revised them based on those discussions, focusing on one core assumption for each SA:

- SA1: Changing the likelihood of getting caught and facing penalties for illegal behavior will deter people from illegal activities that harm habitats and wildlife such as poaching and starting fires.

This assumption is based on the canonical Becker crime model, which is that criminal behavior is engaged in if the *expected benefits* from committing a crime exceed the *expected costs*.[14] Costs are a function of the rate of detection, the potential penalties, and the likelihood of facing those penalties. While there is much evidence to support this assumption, it might be invalid if the expected benefits remain greater than the expected cost, or if people do not have available alternatives for livelihoods.

- SA2: By engaging in more sustainable agriculture, conservation enterprise, and nature-based tourism, people will increase their perceived benefits of conservation.

This is a critical assumption for the Kafue Activity, yet evidence is lacking to support it. People directly benefiting from the goods and services that nature provides might engage in more sustainable and conservation-based practices, or people may continue to complement livelihood benefits from tourism/agriculture/conservation enterprises with continued extractive forest and

[14] Becker, Gary S. and William M. Landes, eds., "Crime and Punishment: An Economic Approach," *Essays in the Economics of Crime and Punishment* (1974): 13-68, https://www.nber.org/system/files/chapters/c3625/c3625.pdf.

wildlife activities. Indeed, the conservation enterprise and tourism activities are not expected to provide many people with direct benefits.

The link to perceived benefits of conservation is even more tenuous for sustainable agriculture. There is no rigorous empirical evidence to indicate whether people will see a link between conservation and increased agricultural productivity due to sustainable agricultural practices. If people do not identify a link between conservation and agricultural practices, then they will not be incentivized to halt land clearing. Alternatively, if households experience significant income benefits from the introduction of cash-crops, they may be motivated to clear additional land for agricultural activities. Lastly, there is a risk of unintended negative consequences, with people using increased incomes to engage in more—not less—extractive behavior.

- o As noted above, the project should consider rephrasing this to "by *engaging in more sustainable agriculture, conservation enterprise, OR nature-based tourism,"* as there is little indication that all three of these activities will overlap in the same locations. Additionally, agriculture, tourism, and conservation enterprises should be understood to affect income and jobs (intermediate results in the results chain) independently.

- SA3: Linking health services with conservation activities will improve both human health and reduce threats to habitat and wildlife.

Based on the current implementation plans, the FA team does not believe there is evidence to support this assumption. This is in part because the health and conservation interventions for the Kafue Activity are not integrated or linked in the way that they would be for other program models where these dual benefits have been shown to materialize, such as the Health in Harmony approach.[15] Also, as community health services will be provided by different IPs from those doing conservation programming, there is no reason to believe that communities would associate the two activities. Additionally, the workshop discussions did not provide support for the hypothesis that healthier people will not need to engage in unsustainable resource extraction or land use to pay for healthcare costs, as it is unclear if this is indeed a driver of this behavior to begin with in the program communities.

- o It should also be noted that the CHWs are expected to have a direct influence on family planning and MCH, and boreholes are expected to reduce the incidence of waterborne illness, which may have some indirect effects on MCH outcomes. However, until the extent of geographic overlap is determined, these two components are not expected to interact in terms of health impacts.

[15] For example, an evaluation of a program in rural Borneo found that a conservation–health care exchange reduced illegal logging and simultaneously improved human health and well-being, the intervention expanded health care access and use for communities living near a national park, with clinic discounts offsetting costs historically met through illegal logging, coupled with conservation, education, and alternative livelihood programs. Source: Jones, Isabel J., "Improving rural health care reduces illegal logging and conserves carbon in a tropical forest," *the National Academy of Sciences* 117, no. 45 (November 2020): 28515-28524, https://www.pnas.org/content/117/45/28515.

- SA4: Increased community participation, strengthened conservation governance structures, and land use and resource management plans will reduce threats to habitat and wildlife, improving biophysical conditions.

Overall, the results chain and assumptions for this SA should be revisited once there is more clarity on the interventions at the end of Year 1. While there is evidence that governance structures and management plans are important for threat reduction and more sustainable land and resource management, without strong enabling conditions these reforms will likely be insufficient for achieving the desired outcomes. For example, too often community participation is not sustained, plans are developed but not strongly enough enforced, or governance structures are captured by elites who use them to their own advantage. For improved management and governance to be effective, the project logic model also assumes that these must be layered with livelihood incentives for local communities to conserve forest and wildlife resources, such as increasing wildlife numbers that attract more private-sector investment.

At its core, the Kafue Activity—and the HEARTH portfolio more broadly—is based on the assumption that human well-being and conservation/biophysical outcomes are linked and have the potential to impact each other positively. For example, as human well-being improves, there might be a reduced need for extractive/unsustainable behavior, such as lower demand for forest resources during times of stress or shocks. Alternatively, as biophysical outcomes improve, this might improve agricultural productivity or human health. However, in many instances, the opposite may be just as likely to occur. For example, as human well-being improves, there might be greater demand on natural resources for things like energy, building materials, etc. Alternatively, as biophysical outcomes improve, this might occur at the expense of human well-being, as communities are barred from accessing resources which they previously relied on for their livelihoods with few available alternatives. Generating rigorous evidence about these relationships and how human well-being and conservation/biophysical outcomes impact each other—and what conditions might contribute to the realization of positive outcomes—is of primary interest not just for USAID, but the broader development sector.

For all assumptions, it will be important for both TNC to closely monitor outputs and intermediate outcomes as feasible, and for the FA team to include outcome and impact indicators that can be used to validate these assumptions for the evaluation. See more in Section IV on the proposed outcomes and data sources and linkages to the key assumptions.

Literature Review

Conservation and Biodiversity Evaluations

Conservation employs a variety of interventions, usually implemented as a suite that spans three levels in line with the Conservation Measures Partnership taxonomy of conservation actions: (1) interventions to improve the enabling environment for conservation, (2) interventions to change behavior/mitigate the threat, and (3) actions to relieve direct stress on species and ecosystems through land/water and species management.[16] Common USAID interventions include protected area management, conservation enterprises, law enforcement, demand reduction/behavior change campaigns, and strengthening enabling environments (legal/policy reform, conservation planning, education/training, institution strengthening), as well as more innovative market-based and direct economic payment schemes. Meta-analyses of impact evaluations, which measure the causal impact of programs, have unsurprisingly not identified a silver-bullet strategy for ensuring conservation outcomes.[17] Conservation programs typically include a bundle of interventions not easily disentangled, such as resource protection, habitat maintenance, and alternative animal sourced foods.

There is significant variation in the rigor of studies about the effectiveness of conservation programming. Many studies on the effectiveness of conservation strategies involve simple monitoring of indicators or case studies.[18] To date, impact evaluations are rare in conservation science; there are limited counterfactual-based studies that evaluate intervention effectiveness, and many are subject to a poor research design.[19] This is especially true for efforts to assess the effects of programming on both conservation and poverty reduction, with limited and methodologically weak efforts to assess poverty outcomes relative to measuring forest conditions.[20] Strong evidence has a patchy geographic distribution, and many studies lack long-term outcome measurements and/or focus on only a single outcome—forest cover change. Conservation programs have been biased towards locations facing relatively low threat levels and, by design, with high biodiversity value.[21] This is problematic for understanding impacts in partially degraded landscapes with dynamic land-use change. It also indicates the opportunity to find larger conservation impacts in areas facing more degradation and deforestation pressures.

[16] Faust, Christina, Tim Holland, and Heather Huntington, "Forest Evidence and Opportunities for Forest Conservation and Restoration to Achieve Multiple Sustainable Development Goals," Working Paper.

[17] Börner, Jan, Dario Schulz, Sven Wunder, and Alexander Pfaff, "The Effectiveness of Forest Conservation Policies and Programs," *Annual Review of Resource Economics*, Vol. 12, Issue 1 (2020): 45-64, https://doi.org/10.1146/annurev-resource-110119-025703. eprint: https://doi.org/10.1146/annurev-resource-110119-025703.

[18] Ferraro, Paul J., and Subhrendu K. Pattanayak, "Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments," *PLoS biology* 4, no. 4 (2006): e105–e105, https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040105; Stem, Caroline, Richard Margoluis, Nick Salafsky, and Marcia Brown, "Monitoring and Evaluation in Conservation: A Review of Trends and Approaches," *Conservation Biology* 19, no. 2 (2005): 295–309, DOI:10.1111/j.1523-1739.2005.00594.x.

[19] Ribas, Luiz Guilherme dos Santos, Robert L. Pressey, Rafael Loyola, and Luis M. Bini, "A Global Comparative Analysis of Impact Evaluation Methods in Estimating the Effectiveness of Protected Areas," *Biological Conservation* 246:108595 (2020): https://doi.org/10.1016/j.biocon.2020.108595; Burivalova, Zuzana, Thomas F Allnutt, Dan Rademacher, Annika Schlemm, David S Wilcove, and Rhett A Butler, "What Works in Tropical Forest Conservation, and What Does Not: Effectiveness of Four Strategies in Terms of Environmental, Social, and Economic Outcomes," *Conservation Science and Practice (Woodrow Wilson School of Public)* 1, no. 6 (2019), https://conbio.onlinelibrary.wiley.com/doi/full/10.1111/csp2.28.

[20] Samii, Cyrus, Matthew Lisiecki, Parashar Kulkarni, Laura Paler, and Larry Chavis, "Decentralised forest management for reducing deforestation and poverty in low- and middle-income countries," (2016), https://www.semanticscholar.org/paper/Decentralised-forest-management-for-reducing-and-in-Samii-Lisiecki/cbcf6e987be3eae660befdc861877449a03065d8.

[21] Joppa, Lucas N., and Alexander Pfaff, "High and Far: Biases in the Location of Protected Areas," *PLoS ONE 4*, no. 12: e8273 (2009), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008273.

Although biodiversity outcomes remain significantly understudied, over the past ten years, the rigor of conservation evaluations to measure forest cover change has significantly improved.[22] This includes the increasing use of statistical matching techniques as a correction for selection bias, which occurs when there are pre-existing, systematic differences between participants and non-participants that introduce bias into study designs. However, more rigorous study designs such as matching combined with difference-in-differences (which estimates impact by comparing changes in outcomes among program participants with changes in outcomes among non-participants), synthetic controls (which construct a weighted average of potential comparison units that best resembles the treated units), and randomized control trials (which use random assignment to ensure that those assigned to participate in the program are, on average, the same as those who are not) remain limited.[23]

Counterfactual/causal studies have not been prioritized in the conservation space relative to other development sectors. Several challenges have been raised about conducting rigorous research for this sector. These center around concerns that measuring impacts on biodiversity and conservation are methodologically challenging and expensive.[24] Specifically, challenges include: a historical legacy of prior interventions; purposeful selection of treatment areas; hard to identify comparison areas; large variability in ecological outcomes; long time lags between intervention and ecological response; programs with multiple interventions; complex spillover effects (e.g., forest use, species movement); large spatial scales of environmental processes; and data constraints, including an overreliance on self-reported behavioral indicators.

Of particular concern is the number of challenges to collecting biodiversity data and indicators in the context of counterfactual designs. This generally relates to difficulties in finding valid control sites and the high costs for collecting a sufficiently large sample of biodiversity outcomes. Many studies have noted that biodiversity is difficult to measure in the context of a statistically robust approach, especially an approach that would be viable to use biodiversity as an impact measure in the context of an impact evaluation.[25] Population trends pre- and post-intervention for selected species across a sample of forests that receive the program and a similar enough sample of comparison forests without the program are required. Fundamentally, measuring biodiversity is a costly data problem, as it is time intensive and expensive to measure biodiversity through standard methods such as transect sampling and netting. Current data sources that provide measures of forest extent, deforestation, and land cover change do not necessarily provide good proxies for biodiversity measures, as forest cover does not indicate the

---

[22] Baylis, Kathy, Jordi HoneyâRosés, Jan Börner, Esteve Corbera, Driss EzzineâdeâBlas, Paul J Ferraro, Renaud Lapeyre, U. Martin Persson, Alex Pfaff, and Sven Wunder, "Mainstreaming Impact Evaluation in Nature Conservation," *Conservation Letters (Agriculture) 9,* no. 1 (2016): 58–64, https://www.cifor.org/knowledge/publication/5829/.
[23] Börner, Schulz, Wunder, and Pfaff, 2020.
[24] Ferraro, Paul J., "Counterfactual Thinking and Impact Evaluation in Environmental Policy," *New Directions for Evaluation 2009*, vol. 122 (2009): 75–84, https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.297; Rissman, Adena R., and Robert Smail, "Accounting for Results: How Conservation Organizations Report Performance Information," *Environmental Management* 55, no. 4 (2015): 916–929, https://doi.org/10.1007/s00267-014-0435-3.
[25] Persha, Lauren and Phoebe Bui, "Conducting Randomized Controlled Trials (RCTs) to Evaluate the Impact of Land and Resource Governance Sector Interventions: Strengths, Practical, Challenges, and Best Practice Guidance," *Technical Report, USAID Communications, Evidence and Learning (CEL) Project* (2021); Erik Meijaard, Truly Santika, Kerrie A. Wilson, Sugeng Budiharta, Ahmad Kusworo, Elizabeth A. Law, Rachel Friedman, et al, "Toward Improved Impact Evaluation of Community Forest Management in Indonesia." *Conservation Science and Practice* 3, no. 1 (January 2021), https://conbio.onlinelibrary.wiley.com/doi/full/10.1111/csp2.189.

presence/absence or diversity of species, poaching, etc.[26] Global and publicly available remotely sensed spatial data often cannot be mapped at the site level to directly measure local biodiversity for most species, even if available at a high frequency and at a small enough scale.[27]

The lack of robust evidence makes it difficult to draw insights to inform future conservation efforts, and a number of studies over the past decade have emphasized the need for more rigorous experimental and quasi-experimental studies related to conservation outcomes.[28] Many of the challenges outlined above are not unique to forest conservation and biodiversity impact evaluations, but rather apply generally to impact evaluations in the sphere of international development. As the FA team discusses in more detail in Section IV below, several data and technological advancements have improved the potential for rigorous conservation evaluations.

Conservation organizations have made significant investments in alternative livelihoods, resource protection, conservation agriculture and outgrower schemes to incentivize behavioral changes to improve conservation and reduce threats to biodiversity. However, rigorous evidence for positive impact on outcomes related to conservation, well-being and biodiversity is lacking. Compared to the standard natural climate solutions interventions, such as payment for ecosystem services, protected areas, and land titling, many of the conservation activities that comprise the Kafue Activity are understudied. As a result, an evaluation of the Kafue Activity presents an important opportunity to advance USAID's learning interests about programming and policies designed to improve conservation and biodiversity outcomes.

---

[26] Burivalova, Allnutt, Rademacher, Schlemm, Wilcove, and Butler, 2019.

[27] Hill, Samantha L., Andy Arnell, Calum Maney, Stuart H. M. Butchart, Craig Hilton-Taylor, Carolyn Ciciarelli, Crystal Davis, Eric Dinerstein, Andy Purvis, and Neil D. Burgess, "Measuring Forest Biodiversity Status and Changes Globally." *Frontiers in Forests and Global Change* 2 (2019): 70, issn: 2624-893X, https://www.frontiersin.org/article/10.3389/ffgc.2019.00070; Luque, Sandra, N. Pettorelli, P. Vihervaara, and M. Wegmann, "Improving Biodiversity Monitoring Using Satellite Remote Sensing to Provide Solutions towards the 2020 Conservation Targets," *Methods in Ecology and Evolution* 9, no. 8 (2020): 1784–1786, https://hal.inrae.fr/hal-02607910.

[28] Curzon, Hannah Fay, and Andreas Kontoleon, "From Ignorance to Evidence? The Use of Programme Evaluation in Conservation: Evidence from a Delphi Survey of Conservation Experts," J Environmental Management 180 (2016): 466–475, issn: 1095-8630 (Electronic), https://doi.org/10.1016/j.jenvman.2016.05.062; Luiz Guilherme dos Santos Ribas, Robert L. Pressey, Rafael Loyola, and Luis M. Bini, "A Global Comparative Analysis of Impact Evaluation Methods in Estimating the Effectiveness of Protected Areas," *Biological Conservation,* vol. 246 (2020), https://doi.org/10.1016/j.biocon.2020.108595.

III. Illustrative Evaluation Questions

The FA team focused on key learning questions of interest to USAID, which at a high-level concern the impacts of each SA on human well-being (socio-economic status, food security, health, etc.) and reducing threats to habitats and wildlife, and thus improving biodiversity and conservation. Below is a set of simplified core questions that the FA team used to frame the evaluation design options:

To what extent does each SA (or combinations of SAs):

- Decrease stress on/reduce threats to biodiversity and improve biophysical conditions?
- Change behaviors and norms around conservation?
- Affect livelihoods,[29] well-being, and rural poverty? [30]
- Affect (human) health?
- Have differential effects, including negative externalities, for certain subgroups (such as women, youth, and those in extreme poverty)?
- Achieve sustainable outputs/outcomes/impacts?

In addition to the core evaluation questions, the FA team will address more specific research questions depending on the final evaluation approach for each SA in the evaluation design report. It should be emphasized that the feasibility of answering certain evaluation questions using impact evaluation or performance evaluation methods will in large part be determined by a variety of factors such as the implementation plans, availability of a counterfactual/comparison group, and available budget/resources. The constraints and proposed evaluation design options are described in more detail in Section V.

In terms of examining evaluation questions related to **integration or interaction between SAs**, at present, the FA team cannot confirm the extent of overlap for the various intervention streams. There is limited confirmation of implementation overlap between SA1 (resource protection) and SA2 (agricultural out-grower schemes), limited to the Kaindu area in the Mumbwa district. Plans are in place to ensure some overlap of additional SA programming for SA2, SA3, and SA4 across settlements. However, the extent and locations of overlap will not be determined and/or confirmed until the end of Year 1. The FA team will continue to explore opportunities to examine questions related to the integration of the different SAs as the implementation plans are finalized.

---

[29] Livelihoods include the means or methods that households engage in to earn a living or otherwise meet their basic needs. Livelihoods may be affected by the Kafue activity in a variety of ways, including by making existing livelihoods less extractive, shifting to new/different livelihood activities (e.g., engaging in tourism rather than charcoal production), and/or increasing income from existing or new livelihood activities.

[30] Well-being and poverty are multi-dimensional and include socio-economic status as well as other outcomes like resilience, food security, health, education, and other aspects of quality of life.

Another subset of questions of interest concern **differential impacts** (e.g., what are the impacts for the most vulnerable groups including those in extreme poverty, women, and youth) and what characteristics, conditions, and/or factors lead to better outcomes in some communities compared to others. While these types of questions can be explored with IE approaches, it is unlikely that the evaluation will be sufficiently powered to detect differences between these subgroups—meaning, for example, if the differences between outcomes for women and men are small, statistical analysis may be unable to distinguish these differences from zero. Rather, it is likely that these types of questions will be best answered through a rigorous performance evaluation approach, using a combination of descriptive analysis of quantitative and qualitative data sources.

Understanding Health and Conservation Linkages

Understanding the impact of health programming on conservation outcomes, and vice versa (the impact of conservation programming on human health), is of primary interest for USAID. However, as mentioned above, there are no plans to directly link the health and borehole interventions to conservation outcomes (for example, as would be done in a Health in Harmony type program approach). Therefore, the FA team does not anticipate these health interventions to lead to significant improved conservation or biodiversity outcomes based on the current program design. To the team's knowledge, there are also no plans to implement health interventions in areas without conservation programming, so it would not be possible measure their impact on conservation alone. However, the FA team will explore whether we can compare conservation outcomes in conservation-only program communities with conservation outcomes in health and conservation program communities—to estimate whether there is any added impact from the health programming. However, it should be emphasized that to do so, the difference in impacts between conservation-only communities and health and conservation communities will likely have to be very large—larger than what is probably reasonable to expect based on the proposed interventions. This will be discussed in further detail in the section below on statistical power.

IV. Illustrative Outcomes, Potential Data Sources, and Statistical Power

This section first provides an overview of illustrative intermediate outcomes with potential data sources, followed by illustrative human well-being and biophysical outcomes. We find a number of potential data sources that could be used to measure performance and impact indicators related to conservation and biodiversity. This is due, in-part, to the extensive biodiversity monitoring that will occur as part of Kafue's programmed activities. Finally, this section explores statistical power and implications for the evaluation design. Our power analysis indicates that some strategic approaches would be amenable to impact evaluation approaches that measure outcomes at the household level.

Illustrative Intermediate Outcomes

Below in Table 2 is an illustrative list of intermediate outcomes for each SA sub-approach important to include in monitoring and evaluation based on results chains. Intermediate outcomes are between the activity outputs and the higher-level human well-being and biophysical outcomes and impacts in the results chains, providing important indicators for the pathways or mechanisms through which the activities might affect change. The FA team does not include outputs here that should be covered by the IPs monitoring activities (e.g., number of scouts trained, farmers contracted, boreholes constructed).

Table 2. *SA Specific Intermediate Outcomes*

Strategic Approach	Sub Approach	Illustrative Intermediate Outcomes	Illustrative Data Sources
SA1: Strengthen natural resource compliance and management systems	Resource protection/law enforcement	Patrol intensity	Panthera SMART monitoring
		Encounters and apprehensions	Panthera SMART monitoring
		Weapons equipment seized, bushmeat/fish confiscated, etc.	Panthera SMART monitoring
	Investigation/ prosecution	Prosecution and conviction rates	WCP courtroom and prison monitoring
		Awareness of convictions/penalties	Recall questions in household surveys
SA2: Develop inclusive	Sustainable/	Average crop production, by targeted high-value crop	Amatheon data; Household surveys

Strategic Approach	Sub Approach	Illustrative Intermediate Outcomes	Illustrative Data Sources
ecosystem-based markets for local prosperity	conservation agriculture	Use of promoted agricultural technologies/practices	Amatheon data; Household surveys
	Tourism	Incomes from tourism activities	Kashikoto/TNC data; Household surveys
		Number of full-time equivalent jobs created	Kashikoto/TNC data
	Conservation enterprise	Incomes from conservation enterprises	Kashikoto/TNC data; Household surveys
		Number of full-time equivalent jobs created	Kashikoto/TNC data
SA3: Strengthen community MCH and improve access to clean water	Community health workers	Knowledge and attitudes regarding aspects of MCH and family planning	I4Life monitoring data; Household or community surveys
		Change in norms and demand for MCH and family planning services	I4Life monitoring data; Household or community surveys
	Boreholes	Use and condition of boreholes	TNC monitoring data; Household or community surveys
		Frequency and funding for borehole maintenance	TNC monitoring data; Community surveys
		Availability and quality of water[31]	TNC monitoring data; Community surveys
SA4: Develop effective land and resource use planning, tenure, and governance systems	Natural resource governance	Strength and security of community land and resource rights; collective action levels; trust in local leaders and accountability indicators; and assessments of satisfaction with land and resource governance	Household or community surveys; CRB monitoring survey Site-Level Assessment of Governance and Equity conducted by TNC

[31] Water quality may include a variety of indicators depending on what is most relevant for the GKE context. For example, presence/absence of *E. coli* in drinking water sources may be of interest given the link to health. Other indicators might include temperature and pH.

Illustrative Human Well-being Outcomes

Table 3 below includes illustrative human well-being outcomes for the Kafue Activity. For more details on the potential indicators, please reference the detailed guidance and Performance Indicator References Sheets in the HEARTH Monitoring and Evaluation Toolkit Guidance document.[32] These indicators are not linked to any one strategic approach, but rather are anticipated to be relevant across the Kafue Activity project areas and interventions. It is anticipated that all indicators below would be measured through household surveys.

Table 3. *Illustrative Human Well-being Outcomes from the HEARTH Global Toolkit*

Outcome Type	Illustrative Outcomes	Potential Indicators
Food Security and Nutrition	Increased dietary diversity	Percent of women of reproductive age consuming a diet of minimum diversity (MDD-W)
	Improved individual or household food security	Percent of households experiencing moderate and severe food insecurity, based on the Food Insecurity Experience Scale
	Reduction of potential exposure to zoonotic diseases	Percent of households consuming high-risk wild meat in the past year
Health	Improved household drinking water source	Percent of households with improved drinking water source; Percent of children under five with diarrhea in the past two weeks
	Increased use of family planning services	Percent of women/men of reproductive age who are using a contraceptive method; Percent of women given information on contraception methods (Method Information Index)
	Increased access to maternal health services	Percent of pregnant women who have attended at least two comprehensive antenatal clinics; Percent of pregnant women who deliver assisted by a skilled attendant at birth
Conservation Knowledge,	Improved knowledge and attitudes towards conservation and	Average score measuring the perceived importance of protecting nature and the environment

[32] Final draft to be published early 2022.

Outcome Type	Illustrative Outcomes	Potential Indicators
Attitudes, and Practices	natural resource management	
	Reduced unsustainable use of resources	Percent of households who engaged in unsustainable use of ecosystem resources in the past year; Percent of households that cleared land for cultivation in the past year
Agriculture and Land	Increased agricultural productivity	Average crop yield, by targeted high-value crop
	Increased use of sustainable/ regenerative practices	Number of hectares under improved management practices or technologies
Resilience	Increased household resilience	Average score on the ability to recover from shocks and stresses index
	Use of natural resources to reduce effects of shocks and stresses	Average score measuring the extent that households rely on natural resources during times of stress
Socio-economic Well-being	Increased socio-economic status	Percent of households below the comparative threshold for the poorest quintile of the Asset-Based Comparative Wealth Index
		Change in per capita household consumption/expenditures in key areas such as health, education, etc.
	Increased benefits from nature-based economic activities	Average household income from nature-based products and/or services

It should be flagged that there are no gender equality and social inclusion **(GESI)-specific indicators** proposed above. At minimum, the FA team will report data disaggregated by key marginalized groups identified by the activity (e.g., lower/higher socio-economic status, men/women, and youth). Additionally, the FA team recommends including a supplementary survey of primary female decision makers to address the gender data gap created by limiting household surveys to primary male decision makers (see more details in the Challenges and Limitations section below). While there are no GESI-specific outcomes in the TOC, several important gender issues were raised in the workshop including girls' education, domestic violence, and time use. The FA team will add gender-specific indicators in collaboration with USAID, TNC, and the IPs if identified as a priority by USAID. Some of these indicators might be relevant based on indirect effects along a TOC. For example, if the CHW training focuses on an alcohol reduction campaign, then we might expect a reduction in gender-based violence. We might also expect to see improvements in girls' education outcomes as a result of improved incomes and livelihoods. Additionally, if women are primarily responsible for extractive income generating activities like charcoal production, shifts away from these activities might result in changes in time use. Other gender outcomes from the global HEARTH toolkit that we can consider include women's economic empowerment and intra-household decision-making.

In addition, there are currently no proposed indicators for **watershed health** and related negative externalities from agricultural activities, as it is not clear from the TOC whether this is an important pathway through which the Kafue Activity is expected to affect biophysical outcomes. However, the FA team understands that generating evidence on the link between better watershed health and improved human health outcomes is of interest to USAID. If this is of primary interest, indicators can be considered as part of the evaluation design to measure watershed health. Illustrative indicators include turbidity of natural aquatic environments (surface, freshwater sources) near agricultural activity sites; pH of natural aquatic environments; presence/absence of *E. coli* bacteria in drinking water sources; and change in concentration of nitrites and nitrates in natural aquatic environments.

Project documents and the draft MERL plan include improved **maternal and infant mortality rates** as potential impacts. While the FA team can measure maternal and infant mortality rates either via household surveys or administrative data (if available at the right geographic level), it is not likely that the design will be statistically powered to detect change. This is because births are a rare event, which reduces the sample size for observation. Thus, the evaluation would only be powered to detect very large changes that might be unrealistic given (1) the type of intervention(s) planned (i.e., CHW training and outreach) and (2) the four-year timescale from baseline to end of project measures (although the evaluation design will propose a longer-term follow-up post program completion). We propose including the indicators above for increased access to maternal health services to serve as proxies for maternal health.

Illustrative Biophysical Outcomes

There are important factors to consider when choosing which indicators and methods are best suited to evaluating the impact of the Kafue Activity on biophysical outcomes. In particular, the scale at which the activity will be able to observe and detect biophysical change will depend on two factors:

1) The spatial and temporal scale at which the intervention is likely to have impact and at which change is hypothesized to occur; and
2) The spatial and temporal sampling strategy used to monitor change.

These considerations underscore the recommendations made below. The following section discusses biodiversity outcomes—in particular, measuring species presence/absence and population abundance.

Additionally, the following section discusses the proposed methods and data sources for measurement, utilizing existing monitoring data from the IPs and supplemented with other sources. This is followed by a discussion on using remote-sensing data to measure change in tree cover and other outcomes, which are important indicators for habitat condition in addition to being proxies for wildlife biodiversity, and measurement of threat reduction via household surveys.

The FA team will coordinate closely with Panthera and Kashikoto in development of the wildlife and habitat monitoring plan and explore options for ensuring that comparable data is collected for comparison areas as well. In cases where we might need to fill data collection gaps in control areas, we can consider engaging Panthera for expanded data collection, partnering with local universities, research institutions and existing long-term research projects (e.g., Zambian Carnivore Programme), and/or engaging local communities in a para-ecology models. The FA team recognizes the importance of interdisciplinary research teams, and that many new tech advancements for measuring biodiversity and forest outcomes require advanced skills and significant time requirements for data processing and analysis. The proposed options weight feasibility, value, and cost of different data collection methods.

Measuring Wildlife Biodiversity Outcomes

Direct measurement of biodiversity is a key interest for USAID. The FA team understands the following as potential outcomes of interest given the proposed interventions for the Kafue Activity: (1) reduce habitat destruction and fragmentation, in particular for riparian and miombo woodlands, and (2) reduce negative impacts on wildlife, in particular with respect to elephants, lions, overhunted bushmeat species, high-value trafficked species, and freshwater fish. Choosing the type of method to monitor species depends on the time and resources available, appropriate spatial scale for sampling, and types of species being monitored.

The global HEARTH MERL toolkit recommends monitoring changes in **species presence/absence** across study areas, as well as tracking changes in the **population abundance** in target species. Presence-absence data help to address simple questions, such as ‘is this species present in the target area?’, while abundance data address questions about trends, such as ‘is this population increasing or declining, and at what rate?’ Aggregating species presence-absence data over multiple species can give an overall picture of species diversity (as a measure of species richness). However, presence-absence and abundance data can only provide limited insights into biodiversity distribution and trends. Wide-ranging or rare species are difficult to detect, so observed absences of species may not reflect true absences in a particular area. Further, demographic, or spatial responses of many species occur too slowly to see significant changes over the lifespan of a project—except when a population is rapidly declining. This remains a significant issue with an IE of proposed target species in the Kafue Activity, such as lions and elephants: these species are unlikely to exhibit observable changes in abundance, and only somewhat likely to show observable changes in spatial distribution. Species with relatively small home ranges and high reproductive capacity, such as some harvested ungulates like sable, are more likely to show changes in distribution over the project lifespan. Finally, the large spatial extent of the GKE poses practical and logistical challenges to gathering *in situ* quantitative data on biodiversity indicators or species abundance and distribution across multiple sites and multiple environmental contexts. For more information on choosing target species, see Annex 8.

Despite the challenges to direct measurement of biodiversity outcomes described above, there are several innovative sources and approaches to help reduce costs and/or improve accuracy in the context of an evaluation of the Kafue Activity. In particular, the extensive biophysical monitoring that the IP will undertake as part of the program, combined with important data and technological advances provides a unique opportunity for a cost-effective impact analysis of biodiversity outcomes.

Planned and Recommended Methods for Direct Biodiversity Measurement

As part of the Kafue Activity, Panthera will be developing a full wildlife and habitat monitoring plan, including (1) a biodiversity baseline assessment and (2) a focal species monitoring and protection plan, each followed by annual monitoring/summary reports. The biodiversity assessment will utilize spoor surveys and camera trapping of large vertebrates, and the focal species monitoring will include Global Positioning System (GPS) collars and the collection of ecological data on group size, age/sex composition, etc. for lions (at minimum).[33] This is in addition to Panthera's community or ranger-based data monitoring of large-bodied vertebrates via SMART platforms.[34] Kashikoto will also conduct an annual aerial survey capturing large-bodied species counts. This extensive data monitoring presents a unique opportunity for biodiversity measures that can be used for an IE. The FA team therefore recommends utilizing the monitoring data from Panthera and Kashikoto to measure direct biodiversity outcomes and exploring options for both (1) expanding these approaches across the entire project area, and (2) comparing areas. The FA team understands that currently, the Panthera biodiversity assessment will report on data specific to the project area, as well as include analysis of data drawn from across the GKE for the same reporting period. More details are provided in Table 4 below on each approach.

[33] The biodiversity assessment and focal species monitoring will take place in the Lunga-Luswishi and Mumbwa GMAs.

[34] SMART wildlife monitoring from patrols will be summarized as part of the patrolling biological monitoring in monthly reports and annual operational reports. This data is used by Panthera for patrol planning, rather than biological impact monitoring. A similar SMART platform is used to capture biological monitoring data from spoor transects and focal species monitoring and for camera trap management (placement, status etc.), which feeds more directly into impact monitoring for population maintenance and recovery in the project area.

Table 4. Methods for Direct Biodiversity Outcome

Method	Tools Required	Considerations for Monitoring
Opportunistic observations Existing Source: Panthera SMART monitoring	Species identification guides, binoculars, surveying equipment (depending on species and ecosystem), data collection devices (notepads, tablets, phones, etc.)	Simple way to track species presence/absence and/or relative abundance. Can be collected by experts or non-experts (e.g., citizen scientists, community, or government patrols). Collection by non-experts will likely require the use of spot-checks or technology-based tools to ensure accurate species identification. Useful in that it is an opportunistic method, and bias will be limited as SMART monitoring will also record search effort (i.e., where observers looked). SMART monitoring will be useful to observe a variety of large mammals that may be difficult to census from the air, in addition to providing valuable information on illegal activities. An important potential limitation may be that past SMART data appear to be largely limited to Kafue National Park, so baseline data may be unavailable for project area GMAs unless monitoring is expanded.
Ground-based transects/point samplings/plot of visual sightings and spoor Existing Source: Panthera spoor surveys	Species identification guides, binoculars, surveying equipment (depending on species and ecosystem), data collection devices (notepads, tablets, phones, etc.)	This is a more rigorous but time-intensive (e.g., time spent making observations, surveying transects), method to survey plant and animal species. It is especially useful in that it provides a controlled spatial and temporal scale of observation with both recorded presence and absences of species. Requires skilled personnel who can accurately identify multiple species. The FA team understands that Panthera is planning spoor baseline surveys in 2022 in Lunga-Luswishi GMA and Mumbwa GMA. These surveys involve walking transects to sight signs of animals (such as footprints, tracks, etc.), rather than the animals themselves, from various species and allow a coarse measure of relative abundance/density.[35]
Aerial transects Existing Source: ZAWA &	Fixed-wing small aircraft, or drones	Aerial survey data can provide baselines for target species such as elephant and species targeted for bushmeat. A series of ecosystem-wide aerial surveys have been performed over the past 20 years (2002,

[35] "Research Findings on the Accuracy of Spoor Surveys as a Method of Calculating Carnivore Populations," *Wild Conservation Research Unit*, 2020, https://www.wildcru.org/news/research-findings-on-the-accuracy-of-spoor-surveys-as-a-method-of-calculating-carnivore-populations/.

Method	Tools Required	Considerations for Monitoring
Kashikoto aerial surveys		2006, 2011, 2015, 2019, 2021) to quantify the distribution and abundances of large-bodied wildlife, with a particular focus on elephant, buffalo, sable, eland, red lechwe,[36] and Kashikoto has also carried out aerial surveys over the past five years in their areas. While resource-intensive to implement, aerial surveys provide area-specific or ecosystem-wide information on vertebrate abundances. In recent years, these data have indicated that a number of species are increasing in number and spatial range across Kafue.[37] In particular, data from aerial surveys could be used to measure impacts on resident species with relatively small home ranges (e.g., buffalo, red lechwe, puku) that might respond locally to Kafue Activities.[38] However, nearly all sightings of these species are in Kafue National Park, rather than in GMAs. As interventions are focused on GMAs, aerial surveys could (in theory) detect a change in spatial distribution due to new arrivals. These surveys not only serve as assessments of trends for ecologically and economically important species, but they are also indicative of pressure from illegal human activities. Carcass ratios (i.e., the proportion of dead-to-live elephants) observed in aerial surveys are widely used as an indicator of poaching pressure.[39] Wildlife species such as buffalo and sable that are relatively numerous on the landscape, occur in open habitat, have relatively small range sizes, and have high sensitivity to illegal activities may be particularly suitable indicators for effective resource protection. Nonetheless, past aerial surveys in GKE suffer from low precision because animals are highly aggregated, which reduce precision in sample-based counts. Frederick (2012) recommended using total counts instead of sample-based counts for some species (e.g., buffalo and elephant) to avoid low

[36] Mkanda, F. X., S. Munthali, J. Milanzi, C. Chifunte, C. Kaumba, N. Muswema, A. Milimo, and A. Mwakifwamba. "The Giant Sleeps Again? Resource Protection and Tourism of Kafue National Park," *Zambia Parks* 24, 1 (2018), https://doi.org/10.2305/iucn.ch.2018.parks-24-1fxm.en.
[37] Mkanda, et al. 2018.
[38] Transect widths have been 2.5-5.0km.
[39] Douglas-Hamilton I, Hillman A, "Using Elephant Carcasses and Skeletons as Indicators of Population Trends in Low-level Aerial Survey Techniques," *CGSpace* (1981), https://cgspace.cgiar.org/handle/10568/4205.

Method	Tools Required	Considerations for Monitoring
		confidence intervals.[40] Evaluations of biodiversity outcomes should account for past demographic trends when assessing whether new interventions are responsible for any observed increases in population numbers. Age and sex data can be useful indicator of a population's current trajectory.[41]
Tools to remotely track animal movements Existing Source: Panthera GPS collars	Different types of technical options for tracking movements of wildlife across protected area boundaries. Physical tagging (e.g., GPS collars).	Provides rich information about individual-level patterns of movement (e.g., degree of connectivity between areas), changes in behavior (e.g., avoidance of human infrastructure or protected area boundaries), patterns of habitat use (e.g., preference for forest versus grassland sites), conflict behaviors (e.g., pattern of crop-raiding by elephants or livestock interaction from carnivores). Can be resource-intensive and are often difficult to implement over large spatial scales. Detection of changes in movement over the course of the program would only realistically be feasible with large sample sizes and large numbers of dispersing individuals. The FA team understands that Panthera plans on including lions, cheetahs, and wild dogs for focal monitoring. Lions have utility as an umbrella species and strong indicator of protection impact. Cheetah and wild dogs are highly cryptic with patchy presence and therefore we cannot be certain that in the timeframe, we will be able to focally monitor these species. Approaches will include monthly focal monitoring during dry season of group composition (age and sex) and size, range use (intensity of spatial use), range size and range heterogeneity, cub survival and group recruitment. Predators require more specialist and intensive surveys than other large vertebrates because of their nocturnal, cryptic behavior. Very High Frequency/Global Positioning System (GPS) telemetry and individual identification will be used (or possibly has been already)

[40] Frederick, Howard, "Aerial Survey: Kafue Ecosystem 2011," *Zambia Wildlife Authority*, ResearchGate (2012), DOI:10.13140/2.1.4466.8801.
[41] Morrison, T. A., A. B. Estes, S. A. R. Mduma, H. T. Maliti, H. Frederick, H. Kija, M. Mwita, A. R. E. Sinclair, and E. M. Kohi, "Informing Aerial Total Counts with Demographic Models: Population Growth of Serengeti Elephants Not Explained Purely by Demography," *Conservation Letters* 11 (2018), DOI:10.1111/conl.12413.

Method	Tools Required	Considerations for Monitoring
		to track individual animals. Behavior responses to GMA boundaries may be useful for IE purposes.
Camera trapping Existing Source: Panthera camera traps	Camera traps, security locks and cases, lithium batteries, data storage cards, computing software for analyzing data	The FA team understands that Panthera plans to deploy camera traps, which are a standard method for monitoring a variety of species over large areas. Camera traps can be useful for tracking presence/absence, and relative density of a wide range of species (large mammals, ground birds, cryptic or nocturnal species, species that can be attracted to baits (e.g., carnivores, small mammals), with little disturbance to habitat. Recently developed analytical tools to automated species detection/counts in images using machine learning classification can greatly reduce image processing time. Camera trapping can be resource intensive depending on number and arrangement of cameras. However, cameras can be easily damaged or stolen when deployed in remote areas, so may be less suitable in areas with low resource protection or high fire frequency. Suffers from relatively frequent damage by wildlife (e.g., elephants and insects). Current Panthera plans are for a systematic sampling design using a grid of 15 x 15 km. They will place a camera station (two cameras per station) within the core of each grid cell, in a selected location along roads and tracks where carnivores are likely to pass. The station placement will remain static from year to year. Traps are placed at thigh height on a stake driven into the ground. The total number of camera trap stations is still to be determined, as there are some grid cells that have limited access that will need to be assessed. However there appears from grid cell count to be approximately 84 stations maximum that would be required.
Collection and synthesis of existing data	Species occurrence databases including historical data, expert curated species distribution maps, existing research,	Relies on existing data and reduces cost for additional data collection, however, insights are usually limited to a coarser spatial resolution and may be prone to missing data and gaps depending on how well the species in question has been documented and how reliable the

Method	Tools Required	Considerations for Monitoring
	spatial distribution modeling	existing data are (e.g., publicly contributed data may have species identification issues).

It is expected that Panthera's wildlife and habitat monitoring plan, will determine which species to measure, the scale and frequency of sampling, and methods to use should be determined by reviews of relevant and reliable published literature and developed in collaboration with implementing partners, stakeholders, and experts knowledgeable about the target ecosystems.[42]

In addition to the methods discussed above, Annex 9 includes a summary of methods which were considered, but ultimately not recommended, along with brief details on resource considerations. Given feasibility concerns, limitations in value, cost considerations, as well as the various benefits from building on the existing tools and methods listed above, the FA team has designated this set of methods as secondary—and for consideration if the sources above are determined to be insufficient for biodiversity measures. For example, one potential approach is to use **satellite imagery to census large-bodied species** across large areas, using image-recognition algorithms to automate the count process. This approach might be cost-effective if the team can access free, high-resolution satellites and low-cost technical expertise. However, ultimately, this method is not recommended because (1) it is limited to elephants, whereas the other counting methods can capture multiple species, and (2) this method works best in areas with very high densities of elephants,[43] which is not the case in GKE. Additionally, **freshwater fish surveys**, either using direct catch methods in aquatic habitat or through surveys of fishermen, are not recommended at this time, as it is not clear whether freshwater fish are a key outcome for interest for the Kafue Activity.

Based on the team's preliminary assessment, the most likely direct biodiversity indicators will involve changes in wildlife behavior or spatial distributions near treatment sites. This shift could be detected with a variety of biodiversity methods. **Camera traps** provide an efficient balance between cost and field effort, yielding high-quality data for a broad taxonomic diversity of large mammals, including key target species (i.e., elephants, lions). The FA team therefore recommends expanding camera traps to collect data for both treatment and comparison areas. Data management and analysis of camera trap images is non-trivial, and camera are prone to theft and damage, so deployment should be carefully coordinated, and include a detailed data management/processing plan.

In addition, the long record of **aerial censuses** in GKE by ZAWA and Kashikoto, and recent **SMART monitoring** activities make valuable baselines for understanding biodiversity outcomes for common large-bodied species (ungulates and elephants). Ideally, aerial censuses would be run multiple times during Kafue Activity, during dry and wet seasons, in coordination with DNPW and Kashikoto. **GPS telemetry** will provide high quality information about behavioral patterns of lions, cheetahs, and wild dogs near settlements. However, the FA team does not recommend expanding this approach to

[42] For more details on recommended steps to define a species monitoring plan, please see the global HEARTH MERL toolkit for guidance.
[43] Gill, Victora, "Elephants counted from space for conservation," *BBC News*, January 21, 2021, https://www.bbc.com/news/science-environment-55737086.

comparison areas for an IE, since it is not expected to cover enough individual animals to generate statistical confidence in results. Rather, this information will provide useful contextual detail about range and behavior.

Planned and Recommended Methods for Forest Outcomes and Biodiversity Proxy Measures

The FA team proposes using remote-sensing data to measure change in tree cover and other outcomes, which are important indicators for habitat condition in addition to being proxies for biodiversity, and measurement of threat reduction via household surveys. For biodiversity, direct measures are often moderately expensive, time-intensive and (in some species or methodologies) require specialist identification, technical or analytical skills. Nonetheless, they provide valuable information on species abundances, distributions, and behavior. Indirect measures using remotely sensed satellite data only provide correlates of biodiversity but can be measured over large spatial extents in a time varying manner. Deforestation rates are the most common standard for evaluating forest conservation interventions.[44] Forests provide important structure, shade, microhabitats, and food resources to a suite of miombo-associated species, and thus habitat loss and particularly forest loss is an important driver of biodiversity loss. Maintaining (or increasing) the total extent of natural ecosystems across the GKE is a primary biophysical outcome of interest for the Kafue Activity. Other remotely sensed metrics such as fire intensity and frequency, vegetative greening and rainfall are correlated with biodiversity, and should at the very least be used to provide environmental context for patterns in direct observations. Additionally, literature suggests that estimating impacts on behavior change/threat reduction is correlated with ecosystem service stocks and flows.[45]

Table 5. *Planned and Recommended Methods and Subsequent Considerations for Monitoring*

Method	Considerations for Monitoring
Remote sensed data	Measuring **forest loss** and **forest gain** are important indicators to measure habitat degradation or ***habitat*** destruction, as well as proxies for deforestation and biodiversity loss/gain. Forest loss and forest gain can be directly monitored across Kafue from satellite imagery using various low-cost, publicly available global spatial datasets (e.g., Hansen dataset, with forest cover data available from 2000-2020; Hansen et al. 2013). **Normalized Difference Vegetative Index** (NDVI) is a widely used metric of vegetative growth which correlates with the intensity of illegal livestock grazing pressure, and inversely the loss of wildlife, within protected

[44] Luintel H, Bluffstone RA, Scheller RM, "The effects of the Nepal community forestry program on biodiversity conservation and carbon storage," *PLoS ONE* 13, vol 6 (2018): e0199526, https://doi.org/10.1371/journal.pone.0199526.
[45] Ferraro, Paul J., Merlin M. Hanauer, Daniela A. Miteva, Joanna L. Nelson, Subhrendu K. Pattanayak, Christoph Nolte, and Katharine RE Sims, "Estimating the impacts of conservation on ecosystem services and poverty by integrating modeling and evaluation," *Proceedings of the National Academy of Sciences* 112, no. 24 (2015): 7420-7425, https://www.pnas.org/content/pnas/112/24/7420.full.pdf.

Method	Considerations for Monitoring
	areas elsewhere in Sub-Saharan Africa.[46] Areas of high NDVI are often associated with higher soil moisture and riparian habitat, and thus attract wildlife (particularly herbivores). Fire plays a large ecological role in miombo woodlands, particularly Kafue, where much of the ecosystem burns each year. **Fire presence, timing and intensity** can similarly be measured from satellite-based remote sensing continuously. The diversity of these fire attributes (i.e., the 'pyrodiversity') is a correlate of mammal and bird species richness in Sub-Saharan Africa.[47] Other datasets provide information on extent and intensity of land use and land cover change, which are broadly associated with biodiversity loss.[48]
Behavior change/threat reduction via household surveys	Data can be collected via household surveys to measure behavior that threatens habitats and wildlife, such as poaching, setting fires, unsustainable charcoal production, etc. Literature suggests that estimating impacts on behavior change/threat reduction is correlated with ecosystem service stocks and flows.[49] However, for illicit activities (e.g., illegal poaching or land clearing) respondents are likely to underreport behavior. Overall, whether accurate/reliable data on illegal behaviors can be collected will depend on how taboo the behavior is. Based on discussions with the Center for International Forestry Research, the FA team understands that in Zambia, hunting is highly criminalized, and people are very reluctant to disclose their real hunting or wild meat-eating behaviors. However, there is a growing literature on a variety of approaches to encourage more truthful self-reporting of sensitive/illegal topics in conservation through anonymization techniques.[50] Randomized response techniques (RRTs) work to overcome bias in self-reporting by presenting respondents with a randomly selected statement about a sensitive topic (e.g., "I eat wild meat") who then report if the statement is true or false. The

[46] Veldhuis, M. P., M. E. Ritchie, J. O. Ogutu, T. A. Morrison, C. M. Beale, A. B. Estes, W. Mwakilema, G. O. Ojwang, C. L. Parr, J. Probert, P. W. Wargute, J. G. C. Hopcraft, and H Olff, "Cross-boundary human impacts compromise the Serengeti-Mara ecosystem," *Science* 363, (2019): 1424-1428, https://www.science.org/doi/10.1126/science.aav0564.
[47] Beale, C. M., C. J. Courtney-Mustaphi, T. A. Morrison, S. Archibald, T. M. Anderson, A. P. Dobson, J. E. Donaldson, G. P. Hempson, J. Probert, and C. L. Parr, "Pyrodiversity interacts with rainfall to increase bird and mammal richness in African savannas," *Ecology Letters* 21 (2018): 557–567, https://onlinelibrary.wiley.com/doi/full/10.1111/ele.12921.
[48] Lechner, Alex M., Giles M. Foody, and Doreen S. Boyd, "Applications in Remote Sensing to Forest Ecology and Management," *One Earth*, vol. 2, Issue 5 (2020): 405-412, https://doi.org/10.1016/j.oneear.2020.05.001.
[49] Ferraro et al., 2015
[50] Ibbett, H., Jones, J. P. G., and St John, F. A. V, "Asking sensitive questions in conservation using randomised response techniques," *Biological Conservation*, vol. 260 (2021), https://doi.org/10.1016/j.biocon.2021.109191.

Method	Considerations for Monitoring
	prevalence of the sensitive behavior is then estimated using the probability of selecting the sensitive statement, the total number of "yes" responses, and the total sample size. In this way, the response of the individual is masked/hidden, but the prevalence in the population can still be estimated. According to Ibbett et al. (2021), "By protecting respondents (who never reveal which statement they answered), and enumerators (who cannot tell which statement was answered), RRTs can reduce bias and yield higher estimates than asking people sensitive questions directly."[51]

Additional Methods for Climate Outcomes

In addition to the biophysical outcomes above, an evaluation design process can explore the cost and feasibility of measuring several indicators related to climate change including **soil carbon sequestration and greenhouse gas (GhG) emission mitigation**. Increasing soil organic carbon is important for agricultural resilience and productivity as well as reducing greenhouse gas emissions, and there are low-cost methods to quantify it based on integrating empirical models, expanded measurement and monitoring networks, remote sensing and crowdsourced management data.[52] GhG emissions, estimated in metric tons of CO2 equivalent, reduced, sequestered, or avoided is one of USAID's standard indicators, and is important for slowing the rate of climate change and reduce climate change impacts. Reducing GhG emissions and improved fire management can also have strong ancillary benefits for air and water pollution, energy security, and health. Moreover, by evaluating the Kafue Activity through a gendered lens, an evaluation can help fill the gap in the evidence base about linkages between reduced GhG emissions and improved gender outcomes. There are several options for measurement, including USAID's Agriculture, Forestry, and other Land Use Carbon Calculator, and remote sensing-based methods.[53]

Statistical Power

This section includes an analysis of statistical power related to outcomes such as well-being, livelihoods, behavioral change, health, and governance, followed by a discussion of similar considerations for remote sensing-based forest related outcomes and biodiversity/wildlife outcomes.

---

[51] Ibid.

[52] 52Paustian, Keith, Sarah M. Collier, Jeff Baldock, and Rach Burgess, "Quantifying carbon for agricultural soil management: from the current status toward a global soil information system," *Carbon Management* 10, no. 6: 1-21 (2019), DOI:10.1080/17583004.2019.1633231.

[53] Climate Links, "GCC standard indicator reporting templates," 2019, https://www.climatelinks.org/resources/gcc-standard-indicator-reporting-templates.

Individual, Household and Group-Level Outcomes

The Kafue Activity is a cluster-based intervention, whereby a group (cluster) of households comprising a community will be exposed to one, or a combination, of the strategic approaches. The package of activities is expected to impact individuals, households,[54] and the community. For clustered interventions such as the Kafue Activity, the total number of clusters in the IE sample is the most important factor for determining the statistical power of the IE design. Statistical power helps control the likelihood of a false negative—in other words, concluding that the program did not have an impact when in reality it did. Increases in power (or more confidence in measuring a statistically significant difference between treatment and comparison areas when, in fact, a difference exists) require a larger sample size and result in a smaller minimum detectable effect size, all else equal.

Generally, IEs of cluster-based interventions can be under-powered when there are a limited number of available treatment clusters (as larger sample sizes will result in larger statistical power) and when there is more heterogeneity (i.e., variation) across clusters, which is expected for the Kafue Activity given the very large geographical area covered by the program. That said, at this time, the FA team does not anticipate that the total number of settlements in the program area and other GMAs surrounding the Kafue National Park will be a limiting factor. Preliminary assessment of the similarity of non-program GMAs/districts and program GMAs/districts is provided in Section V, and the FA team will continue with further analysis depending on the final evaluation design.

The FA team has conducted preliminary power calculations to determine the minimum detectable effect sizes (MDES)—the smallest program impact that the evaluation can confidently detect through statistical analysis—for different sample sizes and evaluation design options.[55] It is important to consider the MDES and whether it is in line with policy and program expectations. For example, if the evaluation is powered only to detect impacts larger than realistically expected given the planned activities, it is more likely that the results will be statistically insignificant. Therefore, if the MDES is larger than expected program impacts, other designs or evaluation approaches should be considered.

The FA team conducted power calculations for measuring outcomes at both the individual/household level and cluster/group level. Figure 6 below illustrates the relationship between the MDES and the number of clusters for a variety of different sampling scenarios for **individual/household level outcomes**, varying the number of households surveyed per community (n) from 10 to 25, and varying the intra-cluster correlation (rho)[56] from 0.1 to 0.3. MDES are reported in standard deviations. It should be noted that increasing the number of households in the IE sample (i.e., by increasing the number of households surveyed per community) has only a minimal effect on MDES, particularly increasing above

---

[54] It should be noted that different implementation activities might affect some or all households within a program community, and that not all households in a program community might directly benefit from activities.

[55] The FA team conducted power calculations using Stata's clustersampsi command. Parameters: power = 0.80; alpha = 0.05. The FA team also accounted for 15 percent attrition, and 25 percent correlation with baseline values or other predicative covariates and the outcome.

[56] The intra-cluster correlation coefficient measures the relatedness/similarity of responses within a cluster. The higher the coefficient, the more similar households are within a community on key characteristics or outcomes and the higher the required sample size.

15. The benefit of increasing the sample size of households within a community above this level would be in ability to measure impacts for different types of households or individuals (subgroups) rather than increasing overall power.

With smaller sample sizes of 70 or 100 total communities, the IE would be powered to detect moderate effect sizes between 0.27 and 0.35 standard deviations depending on the parameters.[57] With larger sample sizes of 150 or 200 total communities, the IE would be powered to detect smaller effect sizes between 0.19 and 0.24 standard deviations.

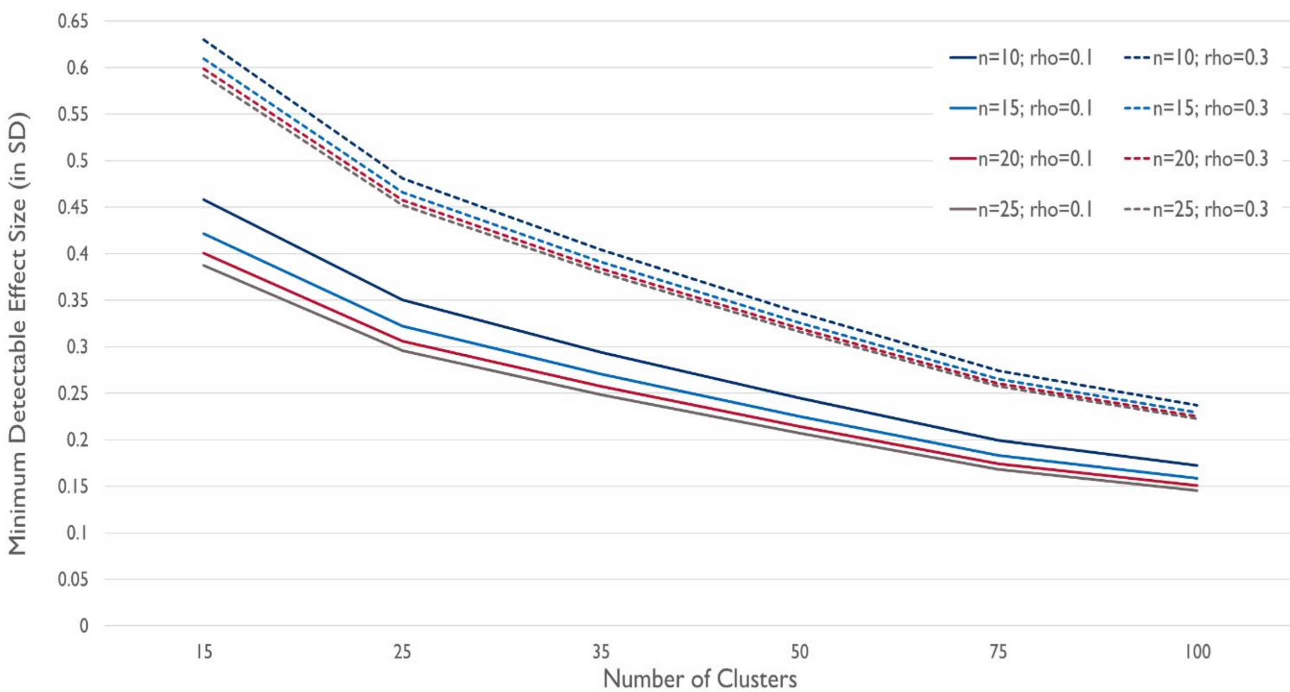


Figure 5. *Relationship between MDES and Number of Clusters – Household Level Outcomes.*

Figure 7 below illustrates the relationship between MDES and the number of clusters for **community level outcomes**. Overall, the IE will be powered to detect larger program impacts for the group level outcomes than it will for the household level outcomes. This is because for group level outcomes, there is only one observation or treated unit for each given period. Even with a total sample size of 200 communities, the MDES for group level outcomes is 0.42 standard deviations, which is a moderate effect size. Smaller sample sizes of 70 or 100 total communities have MDES estimated between 0.71 or 0.59, which are large. In other words, the Kafue Activity would need to have large impacts on community level outcomes like governance or natural resource management-related outcomes for the IE to distinguish real impacts from zero. If the Kafue Activity results in smaller changes in these group level outcomes, the IE will likely be unable to detect these impacts.

---

[57] Generally, MDES less than or equal to 0.20 standard deviations are considered small, between 0.20 and 0.50 moderate, and greater are considered large.

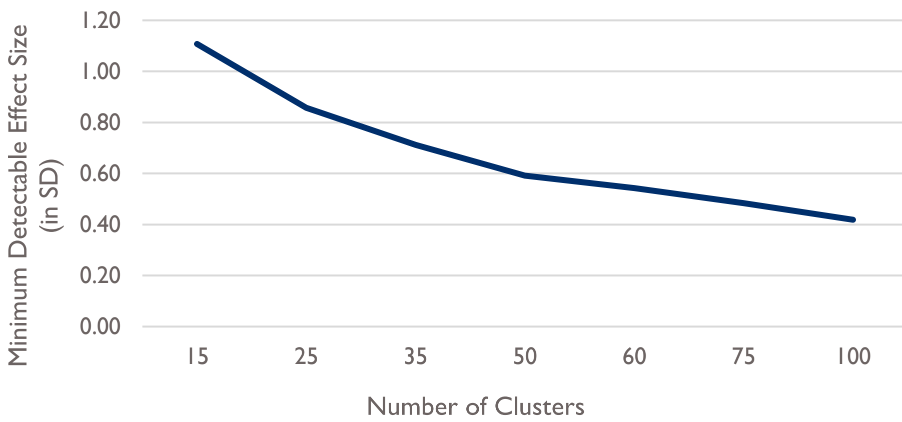


Figure 6. *Relationship between MDES and Number of Clusters – Group Level Outcomes.*

Table 6. *MDES for Different Sample Sizes, Matched Comparison Group DiD Design*

Minimum Detectable Effect SZEs for Different Sample Sizes								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of communities								
Kafue Activity group	35	35	50	50	75	75	100	100
Comparison group	35	35	50	50	75	75	100	100
Total communities	70	70	100	100	150	150	200	200
Number of households per community†	10	25	10	25	10	25	10	25
Total Household sample size†	700	1,750	1,000	2,500	1,500	3,750	2,000	5,000
MDES for individual/household level outcomes [a]	0.35	0.32	0.29	0.27	0.24	0.22	0.21	0.19
MDES for community level outcomes	0.71	0.71	0.59	0.59	0.48	0.48	0.42	0.42

Notes:
[a] MDES = Minimum detectable effect size (calculated in units of standard deviation from the mean)
[b] Calculations assumed a confidence level of 95 percent, two-tailed tests, 80 percent power, 15 percent non-response rate, 25 percent correlation with baseline values or other predicative covariates and the outcome, and 0.2 intra-cluster correlation coefficient.
† The number of households surveyed for the evaluation sample, not the total number of anticipated beneficiaries.

It is important to place the MDES in context to assess what effect sizes are realistic given the expected Kafue Activity impacts. To achieve this, the FA team analyzed data from the 2018 DHS in Zambia to contextualize the MDES for different key outcomes of interest. Data were restricted to provinces which overlap with the program areas and possible comparison areas: Central, Copperbelt, Northwestern, Southern, and Western. Table 7 below includes the means for an illustrative set of potential outcomes of interest at the individual and household level, including family planning, maternal and child health, water access, and socio-economic status.[58]

Generally, evaluation design options with 150 or 200 total communities are more likely to detect realistic impacts. These design options will be able to detect changes of between 6 and 12 percentage points (pp) depending on the outcome and other design parameters. Design options with only 70 or 100 total communities are less likely to detect realistic impacts, with changes between 8 and 17pp. It should be emphasized that these effect sizes are in *percentage points* and not *percent change*. For example, assuming a baseline mean of 33 percent of adult women using any contraception method, design options with 70 total communities would only be able to detect an increase in 16 or 17pp (almost a 50 percent increase from the baseline mean, which is very large) whereas design options with 200 total communities would be able to detect an increase in 9 or 10pp (about a 33 percent increase from the baseline mean, which is still large, but more realistic). It should be emphasized that this analysis is purely illustrative, as baseline values in the program and comparison areas are expected to deviate from the province-level DHS data.

Table 7. *MDES Contextualized for Key Individual and Household Level Outcomes*

Minimum Detectable Effect Sizes (MDES) Contextualized for Key Individual and Household Level Outcomes[59]									
	Mean[c]	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MDES (in Significant Difference) for household level outcomes [a]		0.35	0.32	0.29	0.27	0.24	0.22	0.21	0.19
Use any contraception method	0.33	0.17	0.16	0.14	0.13	0.12	0.11	0.10	0.09
Method Information Index	0.65	0.15	0.14	0.13	0.12	0.11	0.10	0.09	0.09
2 or more ante-natal care visits	0.87	0.10	0.09	0.08	0.08	0.07	0.07	0.06	0.06
Skilled assistance during delivery	0.66	0.15	0.14	0.13	0.12	0.11	0.10	0.09	0.09
Access to improved water source	0.59	0.16	0.15	0.14	0.13	0.11	0.10	0.10	0.09
Child diarrhea in past 2 weeks	0.20	0.16	0.14	0.13	0.12	0.10	0.09	0.09	0.08

[58] Each mean was calculated from the DHS data and then adjusted. Most indicators were reduced by 10 percent, childhood diarrhea was increased by 5 percent and the wealth quintile was kept as-is. This was done because the indicators were high (e.g., 75 percent of women with skilled assistance during delivery and 69 percent of households with access to improved drinking water), however the area immediately around the Kafue National Park has worse human well-being outcomes than the rest of the province areas due to limited livelihood options and lack of access to basic and social services.

[59] All individual and household level outcomes in this table are binary, meaning that they take a value of zero if the condition is false and a value of one if the condition is true. Outcomes for family planning (use of contraception and Method Information Index) and maternal and child health (antenatal-care visits, skilled assistance during delivery, and child diarrhea in the past 2 weeks) are at the individual level, while outcomes for water access (access to improved drinking water source) and socio-economic status (asset-based wealth index) are at the household level.

Minimum Detectable Effect Sizes (MDES) Contextualized for Key Individual and Household Level Outcomes[59]									
Bottom quintile of wealth index	0.21	0.16	0.14	0.13	0.12	0.10	0.09	0.09	0.08
[a] MDES = Minimum detectable effect size (calculated in units of standard deviation from the mean); Scenarios are taken from Table 6 above [b] Power calculations were re-done for these outcomes as they are binary as opposed to continuous. The FA team therefore changed the power calculations from a comparison of means to a comparison of proportions. All other parameters remain the same. Units for the effect sizes are therefore percentage points. [c] Means were calculated from 2018 DHS Zambia for the subset of provinces that overlap with the project area and potential comparison areas and adjusted as described in the footnotes.									

The FA team looks forward to discussions with USAID, TNC, and other implementing partners regarding whether these effect sizes are consistent with expectations for Kafue Activity impacts. If MDES are much larger than could be realistically expected to achieve based on programming, the potential for useful learning through an IE is at greater risk, as smaller impacts would be indistinguishable from zero.

Design and Analysis Considerations for Biophysical Outcomes

As detailed above, **tree cover loss/gain** will be a primary outcome, both as a direct measure of habitat loss and as a proxy for biodiversity. Satellite-based images are analyzed as raster data, which are comprised of pixels. For analysis, each pixel is a unit of observation. Pixel size will vary based on the satellite being used (e.g., Moderate Resolution Imaging Spectroradiometer [MODIS] has 250m resolution per pixel, whereas Hansen has 30m resolution). It is expected that analysis would occur at both the whole of project level (in line with TNC's proposed methodology) as well as around human settlements (using a buffer distance still to be determined within which impacts from threat reduction might be expected). Based on the large project area and possible comparison sites (involving sample sizes of tens of thousands of pixels per site, or millions across the project area), the FA team is confident that they will be sufficiently powered to detect realistic changes in forest cover and other remote sensing-based outcomes between treatment and control sites.

The FA team summarized forest loss in the GMAs and Kafue National Park since 2000 using Hansen data. The forest loss in GMAs is relatively high at -1.4 percent per year on average across areas. For context, an analysis of global patterns of forest loss across International Union for Conservation Nature (IUCN) categories of protected areas found that in Southern Africa, the average forest loss (+/- standard errors [SE]) with a similar category as the management areas around Kafue (IUCN category VI) is much lower, i.e. on the order of -0.3 percent per year over a 15-year period (-4.5 percent +/- 0.5

percent SE in total).[60] Highly protected areas (IUCN cat II, National Parks) had approximately a 0.23 percent loss per year.

Over shorter timeframes and smaller areas than covered by the above analysis, such as would be the case for the Kafue Activity, the FA team would expect larger SE and small more variable losses, so this is probably overly optimistic. There is also a risk of having an over-powered evaluation design for outcomes when observations are measured in pixels, based on the extremely large sample size. Analysis will need to account for spatial autocorrelation, which is the presence of systematic variation as areas or sites that are close to each other will tend to have similar values. Several standard techniques in spatial analysis are available.

In addition, the FA team considered Kays et al. (2020) evaluation of **camera trap** study design parameters to inform the proposed number of sites, duration, and season of sampling to maximize precision of estimates of species richness, occupancy, and detection rate for mammals.[61] Their overall recommendation is that each sampling bout should run for three to five weeks across 40-60 sites per array, at a minimum. However, the precision of species-level estimates of occupancy was highly sensitive to occupancy level, with more than 150 camera sites or longer time intervals, likely needed for rare species. In GKE, nocturnal predators such as lions, or trafficked species such as pangolin, will be relatively rare to capture, so if cameras are used for detecting these species, a moderate number of cameras will be needed. The study also recommends comparisons of detection rates be model-based and include environmental covariates (e.g., vegetation type) to help account for variation in detection, and that comparisons across study areas or times must account for seasonality, which could have strong impacts on mammal communities in both tropical and temperate sites.

[60] Leberger, Roxanne, Isabel M.D. Rosa, Carlos A. Guerra, Florian Wolf, Henrique M. Pereira, "Global patterns of forest loss across IUCN categories of protected areas," Biological Conservation, vol. 241 (2020), 108299, https://doi.org/10.1016/j.biocon.2019.108299,

[61] Kays, et al., "An empirical evaluation of camera trap study design: How many, how long and when?" *Methods in Ecology and Evolution* 11, no. 6 (2020): 700-713, https://doi.org/10.1111/2041-210X.13370.

V. Illustrative Evaluation Design Options and Methods

This section provides a brief overview of evaluation approaches that were considered for this assessment, followed by illustrative evaluation design options for each of the SAs. This assessment finds that the Kafue Activity is amenable to an evaluation design that includes mixed impact and performance evaluation elements. Several components of the program being implemented in Year 1 might be amenable to evaluation through a quasi-experimental DiD approach. For some individual SAs, experimental methods like a randomized lottery around eligibility cut-off for health and water activities can be further explored. Final decisions about the evaluation design and methodology for activities can only be made when the interventions and sites are determined at the end of Year 1. This section discusses these design options in further detail, including potential approaches for identifying a valid counterfactual.

Overview of Evaluation Approaches

IEs measure the causal impact of a program. In other words, the difference in outcomes caused by the program and not by other external factors. The FA team considered not only IEs, but a variety of evaluation approaches as part of the feasibility assessment. This included (1) experimental approaches, which measure the causal impact of programs through randomization, (2) quasi-experimental, which also attempt to measure causal impacts but without randomization, and (3) non-experimental approaches, which can answer descriptive questions about differences but cannot measure causality with the same degree of rigor or confidence. The latter includes PEs which generally include before-after comparisons without a rigorously defined counterfactual,[62] and case studies, which include in-depth learning from an instance through extensive description and analysis.[63] Table 8 below includes a high-level summary of different approaches that were considered:

[62] USAID, "Performance Evaluation Designs," Accessed February 17, 2022, https://www.usaid.gov/project-starter/program-cycle/project-monitor-evaluation-plan/monitor-evaluation-plan-evaluation-component/performance-evaluation-designs.
[63] USAID, "Technical Note: Evaluative Case Studies," November 2013, https://usaidlearninglab.org/sites/default/files/resource/files/usaid_case_study_tech_note_2013.pdf.

Table 8. *Overview of Evaluation Approaches*[64]

Type	Approach	Description
Experimental	Randomized Control Trial (RCT)	Random assignment (e.g., a coin toss or random number generator) determines who may participate in the program so that those assigned to participate in the program are, on average, the same as those who are not, in both observable and unobservable ways. Since the participants and nonparticipants are comparable, except that one group received the program, any differences in outcomes result from the causal effect of the program.
Quasi-Experimental	Difference-in-Difference (DiD)	Measure the before-and-after change in outcomes for the program participants, then subtract the before-and-after change in outcomes of the non-participants to find the relative change in outcomes for program participants. This methodology is only valid when if the program had not existed, the participants and non-participants would have experienced identical trajectories during the study period.
	Statistical Matching	Individuals who received a program are compared to similar individuals who did not receive it. Comparison groups can be constructed with different techniques including exact matching and propensity score matching. This methodology is only valid if characteristics that were not included in matching either do not affect outcomes or do not differ between participants and non-participants.
Non-Experimental	Performance Evaluation	Performance evaluations, as defined in ADS 201, encompass a broad range of evaluation methods. They often incorporate before–after comparisons but generally lack a rigorously defined counterfactual. Performance evaluations may address descriptive, normative, and/or cause-and-effect questions such as the following: questions about project or activity results or outcomes; implementation processes and their effectiveness; what has been sustained since a project or activity ended; how cost effective was the program compared to existing practice or another approach; was the project or activity viewed as being relevant, or given positive ratings by intended beneficiaries; were men/women, or elderly, or poor, differentially affected by the project or activity.
	Case Study	According to the widely used U.S. Government Accountability Office

[64] J-PAL, "Impact Evaluation Methods: What are they and what assumptions must hold for each to be valid," Accessed February 17, 2022, https://www.povertyactionlab.org/sites/default/files/research-resources/2016.08.31-Impact-Evaluation-Methods.pdf; USAID, "Evaluative Case Studies," 2013; USAID, "Performance Evaluation Designs," 2022.

Type	Approach	Description
		definition: "Case study as an evaluation method is a means of learning about a complex instance, based on a comprehensive understanding of that instance obtained through extensive description and analysis of that instance taken as a whole and in its context". Benefits of case studies include their flexibility of use, efficiency, dealing with multiple interventions, and addressing context. The evaluative case study is best used when the major questions are "how" or "why" questions.

A mixed-method evaluation integrates two or more evaluation methods, usually drawing on both quantitative and qualitative data.[65] Mixed-method evaluations may use multiple designs, for example incorporating both randomized control trial (RCT) experiments and case studies. They also may include different data collection techniques such as structured observations, key informant interviews, household surveys, and reviews of existing secondary data. Mixed methods designs can strengthen an evaluation by (1) using different methods to answer different evaluation questions, or (2) using different methods to answer the same questions (increasing confidence in the validity/reliability of results). Generally, mixed methods evaluations can provide a deeper understanding of why change is/not occurring and capture a wider range of perspectives.

Illustrative Evaluation Design Options

Based on the information available, this feasibility assessment finds that the Kafue Activity is amenable to an evaluation design that includes mixed impact and performance evaluation elements. Some SAs will likely be amenable to causal impact analysis. However, the Kafue Activity is not amenable to a whole of project IE—nor are all SAs amenable to evaluation through IE methods. Activities that are not amenable to an IE can mostly be considered for a mixed-methods, rigorous PE.

As noted above, decisions about the evaluation design and methodology for activities under SA3, SA4, and some components of SA1 and SA2 can only be made when the interventions and sites are determined at the end of Year 1. Several components of the program being implemented in Year 1 (SA1 resource protection and SA2 agricultural markets and out-grower support) might be amenable to evaluation through a quasi-experimental DiD approach, with matching during analysis to improve rigor. For some individual SAs, experimental methods like a randomized lottery around eligibility cut-off for health and water (SA3) activities will be explored. Based on the information received to-date, an illustrative evaluation plan for the Kafue Activity might involve the following elements:

[65] USAID, "Technical Note: Conducting Mixed-Method Evaluations," June 2013, https://www.usaid.gov/sites/default/files/documents/1870/Mixed_Methods_Evaluations_Technical_Note.pdf.

Table 9. *Overview of Potential Evaluation Methods for each Strategic Approach*

Strategic Approach	Sub approach	Potential Evaluation Method(s)
SA1: Strengthen natural resource compliance and management systems	Resource protection/ law enforcement	Mixed; Performance
	Investigation/ prosecution	Performance; Case Study
SA2: Develop inclusive ecosystem-based markets for local prosperity	Sustainable/conservation agriculture	Mixed; Performance
	Tourism	Case Study
	Conservation enterprise	Performance; Case Study
SA3: Strengthen community Maternal and Child Health (MCH) and improve access to clean water	Community health workers	Impact; Mixed
	Boreholes	Impact; Mixed
SA4: Develop effective land and resource use planning, tenure, and governance systems	Natural resource governance	Performance; Case Study

Strategic Approach 1: The **resource protection and law enforcement** component has potential to be evaluated through a mix of PE and IE methods. Given the large area and number of settlements to be covered by increased resource protection activities, the initial design under consideration is a quasi-experimental DiD methodology. For evaluating forest condition outcomes, we would use a matching approach to develop synthetic controls. For the evaluation of settlement and household level livelihood, well-being, governance, and health outcomes, we would apply matching techniques at the analysis stage to improve the balance and power of the study. Comparison areas and settlements may be identified from non-activity GMAs on the western side of the Kafue National Park, matched on key biophysical and human population characteristics. The FA team expects spillovers (i.e., impacts outside of direct project implementation areas on nearby communities or habitats/wildlife) both within GMAs as well as between the GMAs and Kafue National Park. The evaluation can be designed to measure potential spillovers although the analysis of these impacts will likely be more appropriate for PE methods (as impacts from spillovers within GMAs might be too small to statistically detect, and spillovers into Kafue National Park will be challenging to attribute solely to the Kafue Activity) rather than IE methods. PE approaches might combine descriptive analysis of spillovers with qualitative data from key stakeholders to better understand the mechanisms through which these activities might be working.

Also, we highlight that this would not be a comparison of resource protection versus *no* resource protection; there are no 'pure' control areas without any level of resource protection, although many GMAs have very small patrol units compared to the large land area. Instead, inferences about the treatment effect would center on the impact of *high/higher intensity* resource protection versus *low/lower intensity* resource protection. This would still enable the study to address valuable policy and program questions to inform conservation programming—specifically about the level or intensity of treatment needed to motivate better forest and biodiversity outcomes.

Investigation/prosecution of wildlife crime has potential to be evaluated through primarily PE methods, or potentially through case studies. It is the FA team's understanding that these interventions will not be conducted in individual communities but will be focused on judicial reforms and guidance at more centralized levels. Therefore, PE approaches that consider trends in prosecution, penalties, etc. coupled with qualitative data from key stakeholders are most suited to evaluating the effects of these activities.

Strategic Approach 2: The **sustainable/conservation agriculture component** also has potential to be evaluated through a mix of PE and IE methods. Similar quasi-experimental DiD methods may be used as described above for resource protection. Based on Amatheon's currently planned activities, they will be (1) expanding the number of beneficiary farmers in their out-grower program up to more than 2500 farmers, and (2) improving agricultural productivity through sustainable management and organic techniques on nearly 4,000 hectares of smallholder land by the end of the project period. One potentially significant limitation is that Amatheon has already contracted farmers for the most recent growing season (late 2021/early 2022), the crops from which will be harvested in March/April 2022. Therefore, the FA team is unlikely to be able to establish a pre-implementation baseline for most out grower farmers, given that the primary change in behavior (adopting quinoa or chili) as well as the associated impacts on income (via payouts from the contracts) would have occurred. We understand from planning documents that 1,725 of 2,624 out grower farmers may have already been contracted, more than half of the anticipated beneficiaries. However, as farmers increase their yield/productivity over time, for example, as they get better at growing these new crops and adopt sustainable management techniques, it is expected that improvements in yields/productivity and agricultural incomes would continue to occur. The FA team also notes that agricultural outcomes are heavily influenced by external factors, such as weather, emphasizing the importance of establishing a well-defined comparison group. Therefore, it is likely that there will be challenges in attributing change to program activities, and analysis will likely need to be supplemented with PE methods.

Based on the FA team's current understanding of the evolving implementation plans, it is not expected that the **conservation enterprise** activities will occur in a sufficient number of settlements to utilize IE methods. Instead, the FA team anticipates that this intervention set will be most viable for a primarily PE approach.[66] Methods might include interviews with participants in the enterprises, other community leaders/members, and key players in relevant markets to understand the effectiveness of the enterprises. Additionally, **tourism activities** have the most potential to be evaluated through a case study approach. This is because tourism investments and support will be targeted in only a handful of communities. Similar approaches may be taken, interviewing those who work at and visit the tourism camps, as well as key players in the industry at large and community members who live nearby camps. If possible, case studies will be identified for photographic and game drive-based tourism as well as

[66] Based on the activity start-up workshop, we understand that of the 10,000 people that should benefit from increased income under SA2, approximately 9,000 are from the agricultural activities (calculated as 3,000 farmers * ~3 people per household) and only about 1,000 (meaning, only ~300 direct beneficiaries) are from the tourism and conservation enterprise activities combined.

hunting-based tourism investments for comparison. For both conservation enterprises and tourism activities, analysis would also be conducted on secondary data as available (e.g., on trends in tourism visits, enterprise sales).

Strategic Approach 3: Community health workers and boreholes have the most potential to be evaluated through IE methods (supplemented with qualitative data sources). In Year 1, i4Life will be conducting a baseline health assessment in a subset of communities/districts in the northern part of the project area in and around the Lunga-Luswishi GMA. This assessment will include use of the CARE Community Score Card, which will assess the healthcare services for a given community. Assuming that there is a predetermined eligibility threshold, and that more communities will be eligible for the community health workers activities than resources will allow i4Life to work in, *eligible* communities can be randomized into two groups—one that receives the community health workers, and another that does not. Based on the current draft of the MERL plan, implementation plans are for a total of 50 community health workers to be trained, equipped, and linked to health facilities throughout the project, and it seems reasonable that there will be more than 50 communities with a need for better health services. Alternatively, if there are concerns of equity whereby the implementing partners want to ensure that the worst-off communities receive the health workers, two thresholds can be established—one lower threshold, with everyone below the cut-off receiving the intervention, and a second higher threshold at the eligibility cutoff. Assuming that there are enough communities between the first and second threshold, a randomized lottery could be held for this group around the thresholds.

Assuming that there are similarly more communities eligible for borehole construction based on the Year 1 site assessment than there are resources, eligible communities might be randomly assigned to receive a borehole or not. Based on the current draft of the MERL plan, implementation plans are for a total of 100 new boreholes to be newly operational throughout the project, and it seems reasonable that there will be more than 100 communities with a need for—and which meet the criteria for—borehole construction and/or repair.

Strategic Approach 4: Natural resource governance activities have potential to be evaluated through primarily PE methods. This SA is expected to establish new conservation governance structures (at least one CFMG and one CRB), improve capacity of at least 10 existing CRBs and develop/enforce land use and resource management plans. Based on the relatively small number of governance structures (i.e., less than 20) that will receive support, the FA team anticipates that this strategic approach will most likely be amenable to a performance evaluation. This may include interviews with CRB leadership and members, government stakeholders, DNPW, and others as relevant, as well as analysis of any data collected by the CRBs regarding operations. However, if the program focus shifts to a lower level, such as VAGs, then the team can consider whether it would be possible to examine outcomes through an impact evaluation design.

Impact Evaluation Methodology

Difference-in-Differences

DiD is a quasi-experimental evaluation design that estimates programmatic impact by comparing (1) changes in outcomes among program participants with (2) changes in outcomes among non-participants. This method requires four data points: participant group baseline, participant group endline, non-participant group baseline, non-participant group endline. Comparing changes over time between participant and non-participant groups helps control for unobserved and observed fixed confounding factors. The comparison group serves as a counterfactual for the treatment group, providing estimates on what would have happened to the treatment group, had they not received the program intervention.

DiD is one of the most frequently used methods for IEs. In the context of the Kafue Activity, a DiD method can be used to compare outcomes over time for human health, well-being, and livelihood indicators between settlements in districts/GMAs involved in the Kafue Activity and districts/ GMAs not involved in the activity, as well as potentially settlements within project districts that are not involved in the Kafue Activity. Given the inability to randomize implementation across these sites, a RCT or experimental design is not feasible for an evaluation of these components and DiD approaches provide a rigorous alternative.

DiD is a data-driven method which requires a large-scale data collection effort and econometric methods to minimize selection bias between treatment (Kafue Activity) and comparison (counterfactual) groups. However, DiD requires stronger assumptions than randomized selection, and there are several methodological limitations. The key identifying assumption for DiD is the parallel or **common trends assumption**, which states that the counterfactual trend behavior will be the same in treatment and control areas. This is a strong assumption and represents the key limitation of DiD—it cannot control for time-variant differences between the treatment and control groups.[67,] For example, if another organization initiated a forest management project in a control GMA/district, the DiD would not be able to control for these events as part of the regression analysis. For DiD to produce a valid counterfactual, the FA team must assume that no time varying differences exist between the treatment and control groups.

---

[67] Discussions of additional DiD limitations in the literature include:
(a) **endogeneity of interventions** (Besley, T. and A. Case, "Unnatural Experiments? Estimating the Incidence of Endogenous Policies," *The Economic Journal,* 110 (2000): 672-694, https://doi.org/10.1111/1468-0297.00578);
(b) **isolation of specific behavioral parameters** (James J Heckman, "Policies to foster human capital," *Research in Economics*, Volume 54, Issue 1 (2000): 3-56, https://doi.org/10.1006/reec.1999.0225; Richard Blundell and Thomas Macurdy, "Chapter 27 - Labor Supply: A Review of Alternative Approaches," *Handbook of Labor Economics*, *Elsevier,* Volume 3, Part A (1999): 1559-1695, https://doi.org/10.1016/S1573-4463(99)03008-4);
(c) **linearity assumption** (Athey, Susan, and Guido W. Imbens, "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica* 74, no. 2 (2006): 431–97, http://www.jstor.org/stable/3598807); and
(d) **large standard errors** (Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-In-Differences Estimates?," *The Quarterly Journal of Economics*, Volume 119, Issue 1 (2004): 249–275, https://doi.org/10.1162/003355304772839588).

Thus, the DiD strategy is valid if the Kafue Activity is the only factor that induces a deviation from common trends for well-being, livelihoods, and deforestation—including other factors of interest to the evaluation. Although the treatment and control areas can differ before the implementation of the Kafue Activity, this difference must not be reflected in different time trends for key indicators.

The FA team can also mitigate weaknesses in the DiD design with an estimation strategy that **combines matching with DiD** to improve comparability between groups. For evaluating forest condition outcomes, we would combine DiD with a matching approach to develop synthetic controls in the Kafue National Park and GMAs using pre-treatment remotely sensed data. For the evaluation of settlement and household level livelihood/well-being/governance and health outcomes, there is a lack of pre-treatment data that enables the pre-selection of specific control settlements. Instead, we would apply matching techniques at the analysis stage to improve the balance[68] and power of the study to evaluate these outcomes.

In summary, the risk to the validity of this DiD design is that evaluators may not be able to identify an ideal counterfactual group, and that the team will not be able to effectively compensate for or eliminate differences between treatment and control that change over time.[69] The treatment and counterfactual groups do not need to be identical on pre-treatment characteristics, but they need to be subject to the same time varying factors during the course of the program implementation. The greater the differences between the treatment and counterfactual at baseline, the more likely that there will be an unbalanced unobservable or observable time variant factor that influences the outcomes of interest. Thus, it is critical to identify areas and settlements that provide good matches to the treatment areas. For the analyses to be credible and robust, the control group should be as similar as possible to the treatment group across important characteristics that also shape the outcomes of interest. The strength of the evaluation ultimately depends on the validity of the control group or counterfactual.

Finding a Valid Counterfactual

Identifying a valid counterfactual for the Kafue Activity must consider both (1) the unit of treatment and (2) the key outcomes of interest, with their associated measurement approaches. Methods for identifying a counterfactual for settlement-level outcomes, such as human well-being, health, etc. will likely differ from those identifying a counterfactual for GMA or district level outcomes, such as biophysical outcomes. Some outcomes, such as forest cover loss/gain, will likely include more than one approach to answer different evaluation questions (e.g., constructing synthetic controls to measure GMA-level forest cover change from resource protection activities, and estimating change in forest cover within buffer zones around settlements to estimate more direct impacts from threat reduction behavior).

[68] Balance is defined as whether characteristics are similar between treatment and control groups. Balance tests often include means-comparisons for key covariates and outcomes between such groups.
[69] Abadie, Alberto, "Semiparametric Estimation of Instrumental Variable Models for Causal Effects," *National Bureau of Economic Research*, 2020, DOI 10.3386/t0260, https://www.nber.org/papers/t0260.

Finding valid matches for treatment areas at the **settlement level** in the Kafue project area is challenging due to several factors. First, the FA team has limited data, especially at the settlement level, across the study area. This lack of pre-treatment micro-level settlement data on ecological, social, and demographic factors makes it impossible to analyze the comparability or balance between treatment and potential comparison settlement areas at this time—baseline data collection, potentially preceded by a community listing activity, would need to occur to meet these needs. Second, within treatment GMAs and districts, as shown in Figure 8 below, the treatment settlements are clustered together and separate geographically from control settlements. However, lack of pre-treatment data and geographically clustered interventions are not unique to the Kafue Activity and are standard challenges for DiD designs for development programming.

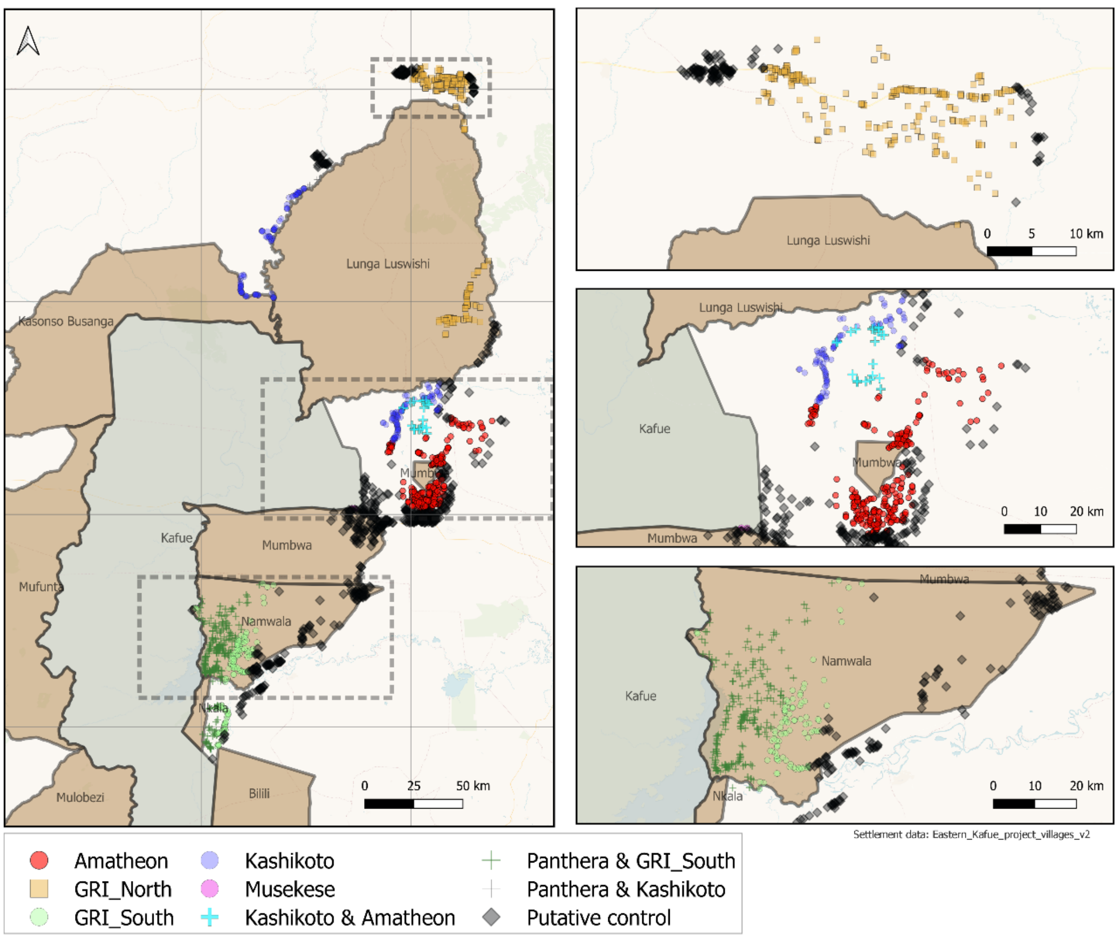


Figure 7. *Spatial Distribution of Treatment and Control Settlements*

As part of the evaluation design process, a set of priority matching characteristics should be used to guide the selection of control **districts and GMAs**. These criteria represent variables that could have an impact on our outcomes of interest, besides the Kafue Activity. The aim is to 'control' for as many of these as possible by selecting counterfactual sites that resemble treatment sites on as many of these key criteria as possible, including ecological characteristics, tenure security, livelihood security, and economic growth. Priority matching variables include the variables shown below in Table 10.

Table 10. *Possible Matching Criteria for Comparison Areas*

Matching Characteristics	Possible Variables of Interest
Ecological Zone	• Contiguity of forest blocks • Forest types • Biophysical risks (fires, pests, diseases) • Biodiversity value (presence of threatened or endemic species or habitats) • Type of current protection regime • Strength/quality of management of current protection regime
Drivers of Deforestation	• Agricultural productivity of the land • Charcoal market drivers (e.g., intensive export to Lusaka, Malawi, or Tanzania versus more localized use) • Reliance on local markets • Travel time (distance) to cities • Access to roads/road density (market integration) including distance to the nearest road

Community and Demographics	• Tenure status
	• Type and capacity of current forest management organization/structure
	• Proximity to forest
	• Population density
	• Poverty levels
	• Ethnicity
	• In or out migration
	• Livelihood systems
	    ○ Cropping systems
	    ○ Level of dependence on nearby forests
	    ○ Energy options
	    ○ Level of dependence on charcoal for livelihoods

Influences from Other Projects	• Presence of strong implementing partners • History of similar projects in the communities • Overlap with other similar or competing projects

Across the study area, the FA team emphasizes that our identification of control areas would **not be 'pure control' areas without any development programming**. Indeed, GMAs were purposefully selected for the Kafue activity, in part because of the pre-existing activities across IPs. For example, Lunga-Luswishi is an area of historical interest for TNC and Panthera and has been classified as important in terms of conservation. TNC has active community conservation projects in the Mujimanzovu chiefdom in Upper Lunga. GRI's historical area of focus has been in the eastern GMAs, initially grounded in proximity to its Elephant Release Facility, close to Nkala GMA. From this base, it has expanded into Mumbwa and Namwala, including with USAID funding through the Community Wildlife Protection (CWP) project. Additionally, the overriding influencing factor over the location of the project is the location of Mushingashi Conservancy (managed by Kashikoto Conservancy Ltd.) to the northeast of the Kafue National Park, and its critical role in the project not just as a partner providing financial leverage but also influencing conservation and livelihood outcomes across a large geography beyond the boundaries of the private property.

Thus, resource protection activities are taking place across the GMAs and Kafue National Park, although there is confirmed variation in the extent of resource protection. Similarly, Amatheon's planned activities are an extension or intensification of their prior activities in Mumbwa. The evaluation design will need to collect information about the extent of ongoing and historical development activities in the comparison GMAs/matched areas of the Kafue National Park. This has important implications for the inferences of the 'treatment effect', as well as the magnitude of detectable effects.

An additional challenge is related to finding viable comparison areas while accounting for leakage/spillover effects (discussed more below in Section VI) and spatial autocorrelation (discussed above in Section IV) for outcomes related to **forest condition and biodiversity**. It should be noted that for their internal evaluation, TNC will use matching based on NDVI and the Enhanced Vegetation Index to identify areas outside the project area to serve as a counterfactual, using DiD techniques to evaluate changes in fire frequency and tree cover. The FA team can consider identifying counterfactual settlements or comparison zones from three areas, depending on the outcome of interest:

1) Areas within Kafue National Park. The FA team can compare conservation and biodiversity outcomes in buffer areas on the eastern side of Kafue National Park to similar matched areas along the western side.
2) Areas within the four GMAs that are the focus of the Kafue Activity. However, the FA team anticipates significant spillover issues within GMAs, and analysis of these areas might be more suited to a PE versus IE approach.
3) Areas within four GMAs on the Western side[70]

The FA team conducted an initial analysis of the comparability of biophysical/climatic variables across treatment and control sites at the GMA and settlement levels (i.e., within 5km of the center point of the settlement) in the Kafue Activity project area. Annex 10 summarizes key bioclimatic variables across different management areas in the Greater Kafue Ecosystem. Variables of interest include: woody cover 2000 (sq km),[71] woody cover 2020 (sq km), percentage of woody cover loss 2000-2020, mean

[70] There are technically five non-project GMAs that border the Kafue National Park. However, the FA team received information from TNC that the land use for Billi Springs in the southeast is primarily (up to 90 percent) agricultural, and so it is not a valid comparison for the other GMAs that still have more natural environments and habitat.
[71] Hansen et al., "High-Resolution Global Maps of 21st-Century Forest Cover Change," *Science 342* (November 2013): 850-53, https://www.science.org/doi/10.1126/science.1244693.

percentage burned per year 2000-2021,[72] percent change in area burned per year 2000- 2021, cropland cover percentage,[73] human population density (sq km),[74] human population total, mean annual precipitation (mm),[75] soil fertility (CEC),[76] elevation,[77] bird species richness, and mammal species richness.[78] With the exception of the significant north-south rainfall gradient, and associated gradient in woody cover (less rainfall and woody cover in the south), and the large difference in the east versus west GMAs in extent of crop cultivation, relatively few other bioclimatic indicators are different across potentially comparison GMAs.

In addition, the team examined changes in key biodiversity drivers—forest loss, vegetative greening, and fire frequency—across the GKE over the past 20 years. Forest loss was most prominent in Eastern Lungwa-Luswishi, Eastern Mumbwa and across Namwala GMAs (Figure 9) and which corresponds with areas of relatively high human density (Table 11). Interventions in these areas may stand to produce the largest biodiversity gains, depending on success.

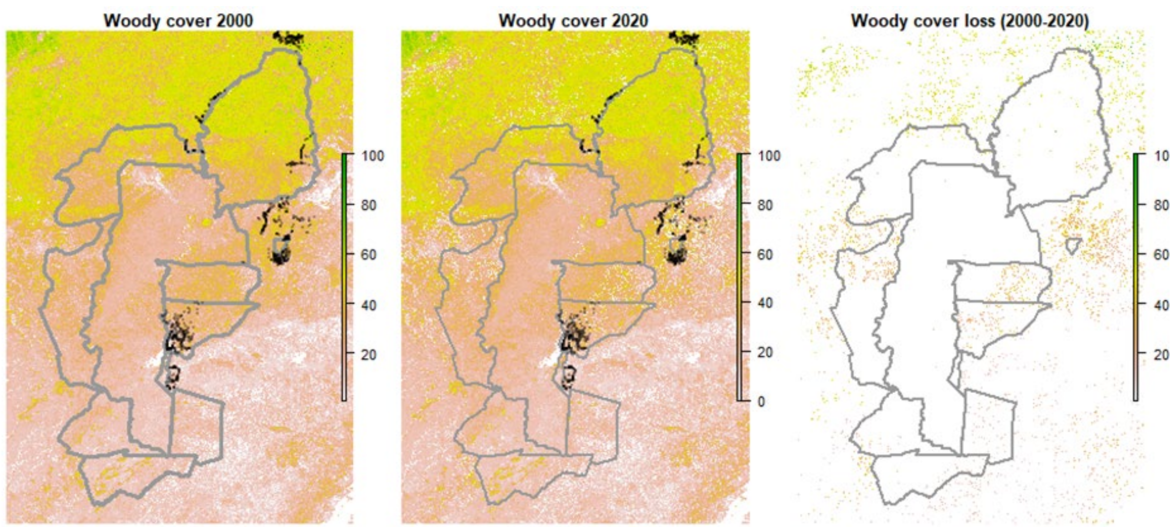


Figure 8. *Woody Cover (Percentage) in 2000 and 2020, and Loss Percentage over Two Decades*

---

[72] "MCD64A1 - Combined Level 3 Direct Broadcast Burned Area Monthly Global 500m SIN Grid," NASA, accessed February 17, 2022 https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MCD64A1#overview
[73] ESRI World Landcover Map, ArcGIS, June 19, 2021, https://www.arcgis.com/home/item.html?id=d6642f8a4f6d4685a24ae2dc0c73d4ac
[74] "Human Population Density," *WorldPop*, Accessed February 17, 2022, https://www.worldpop.org/project/categories?id=18.
[75] S.E. Fick and R.J. Hijmans, "WorldClim 2: New 1km Spatial Resolution Climate Surfaces for Global Land Areas," *International Journal of Climatology* 37, no. 12 (2017): 4302-4315, https://doi.org/10.1002/joc.5086.
[76] Cation exchange capacity (0-5cm), Source: "Africa SoilGrids - Cation exchange capacity (CEC)," ISRIC World Soil Information, Accessed February 17, 2022, https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/e0d921ff-5f7b-48e5-ae27-7c1515055e3b.
[77] This is from the Digital Elevation Model (SRTM), the data of which is provided by NASA. Source: "Digital Elevation Model - SRTM 3 (NASA)," *the United Nations*, Accessed February 17, 2022, https://www.un-spider.org/links-and-resources/data-sources/digital-elevation-model-srtm-3-nasa.
[78] Beale et al., 2018.

Vegetative greening, as measured by maximum annual NDVI, was similarly heterogenous across the ecosystem between 2000 and 2021.[79] Many areas experienced increases in maximum greenness, particularly inside the Kafue National Park and most GMAs, with the exception of Eastern Namwala and Bilili (Figure 10). Fire frequency was extremely high across the ecosystem, with many areas burning in most years.[80] Statistically significant reductions in the probability of annual burning were observed in Bilili, Namwala, Eastern Mumbwa, Western Lungwa-Luswishi (blue below), while increases in fire probability were observed in many other areas (red below).

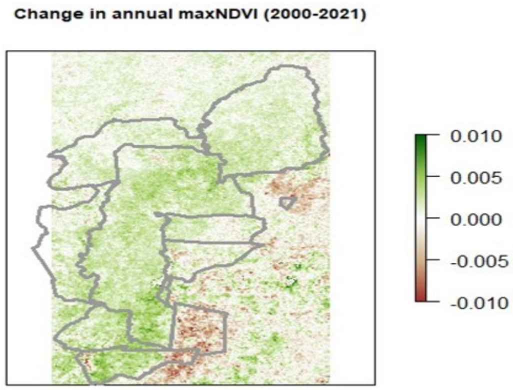


***Figure 9.** Change in Maximum Annual NDVI across the Kafue Landscape between 2000-2021*[81]

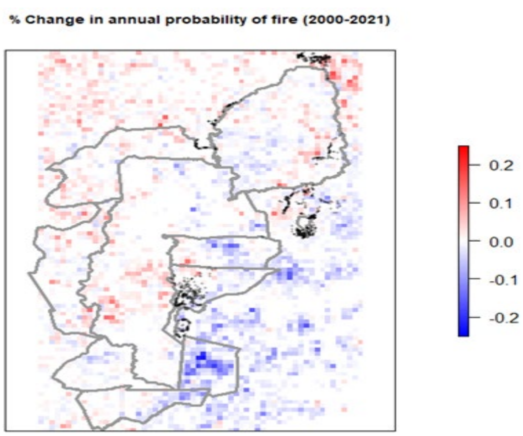


***Figure 10.** Percent Change in Annual Probability of Fire (2000-2021)*[82]

Within GMAs at the settlement level, long-term trends in biodiversity indicators (forest cover loss, change in the probability of fire and change in maximum NDVI) were also broadly similar across different IPs in treatment versus nearby potential control settlements (see Annex 11).

Further and more nuanced analysis of this data—including within GMA analysis—can inform discussions about differences between control and treatment settlements. During the evaluation design phase, a deeper examination of factors such as woody cover change at different distances to the protected area boundary will be necessary based on satellite imagery. GMA boundaries are complicated, and the FA team will need more detailed zoning boundaries or input from TNC before proceeding.

The FA team also collated socio-economic data from Game Management Plans across the nine GMAs surrounding Kafue National Park. Table 11 below also summarizes key socio-economic variables across the Kafue Activity GMAs and the potential control GMAs. In general, the livelihoods of the people in each of the GMAs are primarily agriculture based. All GMPs referenced a reliance on natural resources for livelihoods (e.g., bee keeping and the collection of

[79] Changes in greenness could be driven by an assortment of ecological and anthropogenic processes. In general, decreases in NDVI are associated with land conversion to intensive cultivation, increase in grazing intensity of livestock, and a reduction in fire frequency and extent.
[80] Fire probability is a complex trait in arid environments, and generally declines with greater grazing intensity by livestock and wildlife, lower-than-normal rainfall (causing reductions in fuel load), increases in cultivation, and reductions in prescribed burning by managers and landowners. Thus, a variety of factors could have contributed to observed changes in burning patterns.
[81] Greener areas indicate places where the maximum vegetative greenness has increased in this period, while brown areas indicate decreasing NDVI.
[82] Based on NASA's MODIS Burned Area Product aggregated to 5km resolution and fitted to logistic regression models at every location (i.e., cell). Locations showing non-significant (p>0.05) trends in fire were set to 0.0. The black dots indicate locations of nominal treatment settlements.

honey; traditional or artisan fishing; and carving, pottery, and weaving). Except for Mufunta and Sichifulo GMAs, there was also mention of tourism related business. Additionally, across all GMAs, the primary livestock are cattle, goats, and chicken. There was also consistent mention of limited access to basic services (e.g., water and sanitation) and social services (e.g., health facilities and schools).

Table 11. *Socio-economic Variables, Greater Kafue GMAs*

Socio economic Variables, Greater Kafue GMAs						
GMA	**Pop. Density (sq km)**[83]	**Pop. Total**[84]	**Tribe(s)**	**Chief(s)**	**Main Cash Crops**	**Cropland Cover Percentage**[85]
Project GMAs						
Lunga-Luswishi GMA	4	52,094	Kaonde; Luvale; Lunda; Nkoya; Lamba; Bembas	Kasempa; Mujimanzovu; Ngabwe; Shibuchinga; Kaindu	Maize; Cassava; Groundnuts	0.06%
Mumbwa GMA	6	21,059	Ila	Chibuluma; Mulendema; Kabulwebulwe	N/A	7.21%
Mumbwa Forest Reserve	N/A	N/A	N/A	N/A	N/A	8.33%
Namwala GMA	10	32,109	Ila; Lozi; Tonga	Kaingu; Shimbizhi; Chilyabufu	Maize; Cassava	6.50%
Nkala GMA	18	4,535	Musungwa; Shezongo	N/A	Maize; Cotton	0.26%
Potential Control GMAs						
Bilili Springs GMA	N/A	N/A	N/A	N/A	N/A	21.50%
Kasonso-Busanga GMA	4	26,795	Kaonde; Lundas; Lozis; Luvales; Mbundas; Tongas	Mushima Mubambe	Maize; Sweet potatoes	0.31%
Mufunta GMA	8	50,029	Nkoya; Lozi; Luvale; Mbunda	(Mwene) Kahare	Paprika; Tobacco; Cotton	1.48%

83 "Human Population Density," *WorldPop*, 2022.
84 "Population Counts," *WorldPop*, Accessed February 17, 2022, https://www.worldpop.org/project/categories?id=3.
85 ESRI World Landcover Map, 2021.

Socio economic Variables, Greater Kafue GMAs						
GMA	**Pop. Density (sq km)**[83]	**Pop. Total**[84]	**Tribe(s)**	**Chief(s)**	**Main Cash Crops**	**Cropland Cover Percentage**[85]
Mulobezi GMA	4	12,695	Nkoya; Tonga; Totela; Lumbu Ila; Tokaleya	Moomba	N/A	0.02%
Sichifulo GMA	9	30,532	Tonga; Nkoya	Nyawa; Siachitema; Moomba	N/A	0.00%

The final selection of comparison areas for a DiD design would occur during the evaluation design process in collaboration with TNC and IPs. In particular, the FA team will further explore the viability of using Mulobezi GMA which is also in TNC's historical area of interest and has also been classified as important in terms of conservation values.

Randomized Control Trial

RCTs are the most rigorous experimental evaluation method, as random assignment over a sufficiently large sample ensures that the two groups (those who receive the program, and those that do not) are on average identical on both observable and unobservable characteristics. Since the participants and non-participants are comparable, except that one group received the program, any difference in outcomes can be attributed to the causal impact of the program.

This enables evaluators to attribute any change in outcomes to the program. As discussed above, the implementation plan and sample size of treatment settlements indicate that the CHW intervention might be viable for evaluation through an RCT. TNC has expressed interest and support for randomizing this component of the programming, although the details would have to be determined in collaboration with TNC and i4Life at the end of project Year 1.

Additionally, the borehole intervention can be considered for randomization if there is a sufficiently large number of hydrologically viable areas that can be randomized between a treatment and control set. The evaluation team will not be able to determine this information until the end of project Year 1.

VI. Challenges and Limitations

There are several challenges to conducting an impact evaluation of the Kafue Activity. However, many of these challenges are not unique to the Kafue Activity nor to evaluations of biodiversity conservation programs. Instead, **design challenges** such as phased implementation and related data collection challenges, multiple, bundled interventions, long-time horizons necessary to observe changes in key outcomes, selection bias of communities, and historical legacy of prior interventions are common across impact evaluations for development projects in many sectors. Additional common **threats to the integrity of the evaluation** that will need to be mitigated include potential contamination from other organizations doing similar conservation, agriculture, and/or health interventions in the Kafue Activity area, spillover from project activities into the control group from spontaneous adoption by control group members of popular project interventions, and non-compliance from implementing partners not adhering to implementation plans (i.e., implementing activities in control areas, or not implementing activities in treatment areas).

Other challenges to the rigor of the evaluation discussed in previous sections include (1) having limited statistical power to detect effects (such as, from small sample sizes, large variances/heterogeneity in outcomes, etc.), (2) identification of valid comparison areas, especially considering that outcomes are measured for a few contiguous forest areas, (3) having sufficient evidence for the parallel paths or equal trends assumption that underpins the potential DiD design and (4) challenges estimating direct impacts for biodiversity outcomes. The below sections discuss some of these challenges specific to the Kafue Activity not already addressed in other sections of the report. This is followed by illustrative cost estimates to finalize the evaluation design, conduct data collection, and analyze and report on findings for a base scenario, with a variety of additional approaches and data collection activities. Finally, some additional information needs for the evaluation design are outlined, as well as considerations for alignment with the USAID program cycle.

Design Challenges

Phased Implementation Plan. The phased implementation rollout and uncertainty about the locations for certain activities has significant implications for the evaluation design. As described in detail above, some activities under SA1 and SA2 will begin in Year 1, but other activities for SA3, SA4, and other components of SA1 and SA2 will not begin until Year 2. Therefore, given the uncertainty about locations for the Year 2 activities, the FA team cannot determine the final evaluation design and methodology for these strategic approaches—nor the extent to which evaluation questions reliant on the overlap of strategic approaches can be answered—at this time.

The Year 2 implementation plan will, for example, determine whether RCT or other IE methods are more appropriate for the SA3 activities in particular. In addition, the MERL plan will need to be updated at the end of Year 1, and on a regular basis as implementation information becomes available. A rigorous evaluation will require detailed M&E tracking of inputs, outputs, and specific site locations, along with significant—but not overly burdensome—coordination among the IPs and between the IPs and evaluation team, to ensure that the design is appropriate as implementation plans evolve.

Implications for Data Collection Timing. The phased implementation rollout also has important implications for the timing of the baseline data collection. As detailed above, only resource protection activities for SA1 and agricultural activities under SA2 will begin in Year 1, and the locations of Year 2 activities are yet to be defined. Given this, there are two potential approaches for baseline data collection for human well-being outcomes: (1) collect baseline data in a phased approach, or (2) once at the beginning of Year 2. Given that there is additional cost but limited technical value with the phased approach, the FA team's recommendation is to conduct one comprehensive baseline; please see more details below:

1) Option 1 is to collect data in June/July 2022 in a limited number of areas, such as Mumbwa and one other biologically important area (TBD, potentially Lunga-Luswishi), and again at the beginning of Year 2 once the rest of the implementation plans are finalized. Baseline data collection in these areas in Year 1 would enable the evaluation of interventions related to resource protection (SA1) and the agricultural activities (SA2), as well as the combination of these two activities in the Kaindu area of Mumbwa where they overlap. Panthera wildlife monitoring data and remote sensing data can be used to determine the baseline for conservation and biodiversity outcomes in these areas. To understand the impact of resource protection and agricultural interventions on well-being, livelihood, poverty, and health outcomes, baseline household survey data should also be collected.
    a. However, there are important costs and limitations to this approach. First, there would be additional costs related to the two rounds of household data collection, in terms of both data collection costs and analysis/reporting. Second, is that even a baseline in June/July 2022 would not be a "true" baseline for the agricultural activities. This is because the first cohort of farmers will already have undergone a full agricultural season, including their first harvest as participants in the program.
2) Option 2 is to wait until the end of Year 1/ beginning of Year 2 to conduct one comprehensive baseline household data collection effort after all activities and locations have been determined. The FA team could collect data October - November 2022, before the rainy season begins in earnest in December.[86] This is the most efficient from a cost perspective, but from a technical perspective, it removes the potential to have a "true" baseline for the resource protection components on human well-being outcomes. While this is a limitation, the FA team does not expect significant changes in important human well-being outcomes prior to this single baseline. This is based on the limited resource protection activities that will have taken place on the ground, and the time required to have impacts on well-being.

Integrated (Bundled) Interventions. USAID is interested in studying the impact of integrated conservation, agriculture, livelihood, and health programming. As noted above, the extent of integration across the SAs is to be determined at the end of Year 1 once implementation plans are finalized in terms

[86] Note that it would be important to complete data collection before the rainy season begins in earnest. December through April is the heaviest rainy period.

of their content and geographic locations. In areas where the Kafue Activity implements a "bundle" or "package" of interventions, the evaluation cannot disentangle the causal effect of any one intervention or type of sub-activity on outcomes of interest. The evaluation will only be able to ascertain the average treatment effect for the project, potentially with some descriptive analysis comparing areas or communities that receive different bundles of interventions - unless there is a significant sample size with varying degrees of integration/bundles that could be compared against one another. It should also be emphasized that without significant overlap of interventions, the evaluation will not be able to measure the impact of an 'integrated' program, and instead will be measuring the impact of discrete interventions.

An exception might be if the FA team can randomly assign the CHW or borehole subcomponents, which would allow for an evaluation of the additional impact of these activities compared to "standard" Kafue communities receiving the other interventions. However, it should be emphasized that in order to have enough statistical power to detect the difference between sites with and without CHW and/or boreholes, the additional impact from the CHW or borehole activities would have to be very large—likely larger than realistically expected—on the shared outcomes of interest.

Longer Term Effects. Primary biophysical outcomes of interest—including impacts on forest regrowth, biodiversity and impacts on human well-being from improved watershed health—will take a longer time period to materialize than the standard USAID program cycle. In addition, long-term data on well-being, livelihoods, and health will answer important questions about the sustainability of program impacts. To capture longer-term effects that are key to USAID's learning agenda, the FA team proposes a follow-up to endline data collection about five years after the end of the activity. The evaluation can also provide value by measuring shorter-term outcomes at midline (halfway through program implementation, at about Year 3) and endline (at the end of program implementation, at the end of Year 5), which can provide early evidence regarding the potential sustainability of impacts. For example, if there is no evidence that people are changing threat reduction behavior at endline, it is unlikely that long-terms impacts would be achieved. Overall, this 10-year time period—which has long been used for health and education programming and increasingly for land and resource tenure and democracy and governance programming—will provide important value for USAID.

Selection bias of communities and historical legacy of prior interventions. Selection bias occurs when there are pre-existing, systematic differences between participants and non-participants. Regarding selection of program areas and their previous conservation activities, TNC informed the FA team that the overriding influencing factor for the project location is the location of the Mushingashi Conservancy managed by Kashikoto, and its critical role in the project not just as a partner providing financial leverage but also influencing conservation and livelihood outcomes across a large geography beyond the boundaries of the private property.[87] In addition, TNC's historical areas of interest around Kafue National Park have been the Lunga-Luswishi GMA (program area) and Mulobezi GMA (potential comparison area)—both areas having been classified as important in terms of conservation value. Panthera has also been active in the Lunga-Luswishi GMA, and TNC has active community conservation

[87] More background on the Kashikoto Conservancy can be found in the 2020 report "Community-Based Natural Resource Management in Zambia: A review of institutional reforms and lessons learned from the field" available on USAID's LandLinks (https://www.land-links.org/wp-content/uploads/2020/06/Zambia-CBNRM-Review-Report_Final-1.pdf).

projects in the Mujimanzovu chiefdom in Upper Lunga. Finally, GRI's historical area of focus has been in the eastern GMAs, initially grounded in proximity to its Elephant Release Facility, close to Nkala GMA. From this base, it has expanded into Mumbwa and Namwala, including with USAID funding through the CWP project.

Based on this, the FA team cannot assume that non-program areas are subject to the same pre-Kafue Activity conditions as program areas, and therefore a simple cross-section comparison between program and non-program areas would not be valid. In addition, the history of prior interventions complicates the determination of a baseline period and has implications for the type of evaluation questions that can be answered. For example, instead of comparing GMAs with/without scouts and anti-poaching patrols, the FA team can only answer questions related to the increased investment in scouts and anti-poaching patrols compared to "business as usual" enforcement programs. This will be mitigated by collecting more detailed information on prior activities and using this as potential matching criteria for the evaluation design.

Threats to Integrity

Contamination can occur from other organizations conducting similar conservation, agriculture, and/or health interventions in the Kafue Activity area or control areas. Generally, with a robust evaluation design, contamination is not an issue if interventions from other organizations are spread across both treatment and control areas—but they can be an issue if, for example, an organization targets control areas for implementation specifically because the Kafue Activity is not working there. To date, the FA team is aware of no specific threats of contamination, but this will need to be actively monitored throughout the evaluation.

Spillover. The evaluation design will need to take several forms of spillover into account. Spillovers are indirect effects of a program on those who have not been direct participants. For example, farmers might adopt new technologies/approaches that they see benefitting other farmers (behavioral) or households might learn about benefits of conservation or the importance of family planning from their neighbors (informational). They can be positive—such as neighboring communities hearing about a high-profile poaching arrest, leading to deterrence—or they can be negative, such as older workers losing their jobs if there are incentives to hire more youth/women at tourism enterprises. If spillovers are not considered in the evaluation design, they can become a threat and contribute to either over or under estimating program impacts.

For the Kafue Activity, there is a risk of spillover effects within program GMAs and districts. And, in cases of resource protection, spillover is expected and represents an important mechanism through which outcomes are expected to occur. This is compounded by the close proximity of treatment and control settlements—according to analysis of the settlement data provided to the FA team, the median distance between 'control' settlements and the nearest 'treatment' settlement is 6.4 km (min = 0.1km, Q1 = 3.7km, Q3 = 10.0km max = 46.3km). To mitigate this risk, for the assessment of most interventions included in the Kafue Activity, an evaluation design should (1) use the settlement or community as the unit of treatment that allows for between-household spillovers within the same community) and (2) establishing a minimum distance between treatment and control settlements to reduce any between-community spillover. Finally, the evaluation could be designed to estimate spillover effects by measuring outcomes for those that do not directly participate in the program activities.

Non-compliance or partial compliance occurs when there is a deviation from implementation plans (i.e., people in control areas receive activities, or people in treatment areas do not receive activities). This can happen for a variety of reasons—for example, perhaps someone who signs up for farmer training never actually attends, or instead of using farming inputs, decides to sell it instead (both examples of treatment not getting treated). Or perhaps IPs do not adhere to implementation plans and implement activities in control areas (an example of controls getting treated). To minimize the potential threats from non-compliance, the FA team will need to ensure strong buy-in from all IPs, as well as ensure there are robust tracking systems in place regarding who is receiving what activities throughout the program (which the FA team understands that TNC has plans in place for).

Illustrative Cost by Evaluation Design

Table 12 below provides illustrative cost estimates to finalize the evaluation design, conduct data collection, and analyze and report on findings for a "base" scenario, with a variety of "add-ons". The **base scenario** includes the minimal elements required for a mixed-methods evaluation – two rounds of data collection (a baseline and endline) with the following primary data collection activities: (1) household surveys, (2) focus group discussions, (3) key informant interviews, and (4) community leader surveys. This base scenario would also include analysis of remote sensing data for forest cover but would not include other direct biodiversity measurement. This also assumes that the mixed-methods evaluation would include a quasi-experimental DiD approach, which requires substantial time and labor effort for analysts to obtain sufficient data for matching, and construct a viable comparison group, and conduct additional robustness checks/corrections to confirm aspects of design validity.

Additional elements can be added along the following parameters:

1) **More data collection rounds:** Up to four rounds of data collection including a baseline, midline, endline, and long-term follow-up. A midline will provide important information on progress halfway through the Kafue Activity, which can focus on intermediate outcomes and be used for adaptive management as well as the evaluation. A long-term follow-up will provide answers to key learnings questions on sustainability and information on some of the outcomes that will require longer time horizons to materialize.
2) **Community listing:** Conducted prior to the baseline, a community listing over a sample larger than required for the baseline (e.g., over 400 settlements) will provide important information on community-level characteristics needed for matching, as well as ground-truthing for IP provided information on the location/existence of settlements. Based on the FA team's experience in Zambia, existing settlement or village listings are often inaccurate in rural areas. However, if the absence of a community listing, matching may still occur at the analysis phase.
3) **Women's survey:** Data collected to understand the effectiveness of development interventions on livelihood and well-being indicators has tended to focus on the household head, which largely excludes the experience of women both within programming and research.

Including a 30-minute close-ended survey for the primary female decision maker in male-headed households will facilitate a more nuanced gender analysis of outcomes for women.

4) **Direct biodiversity measurement:** The FA team currently recommends utilizing (1) camera trapping and (2) ecosystem-level aerial surveys. It should be emphasized that these estimates are conservative and assume that the FA team would need to both (1) cover costs for expanding camera traps to comparison areas, and (2) facilitate an ecosystem-level aerial survey. However, the team is still exploring options for data sharing with implementing partners to gain access to data already being collected in potential comparison areas, which could substantially decrease costs.
5) **Randomization:** If feasible, conducting an RCT will come with additional costs, particularly during the design phase (e.g., constructing a sampling frame, conducting random assignment) and due to additional implementation monitoring/coordination with the IPs (e.g., to ensure compliance with the treatment/control assignment, monitor attrition, etc.). However, these costs are marginal compared to the value-add for including this approach.

Table 12. *Illustrative Cost Estimates*

Budget Estimate by Design Option	Base: 2 rounds + No add ons	Option 1: 4 rounds + No add ons	Option 2: 2 rounds + Survey Add Ons	Option 3: 4 rounds + Survey Add Ons	Option 4: 2 rounds + Biodiversity	Option 5: 4 rounds + Biodiversity	Option 4: 2 rounds + RCT	CT
Methods	DiD + PE	DiD + PE	DiD + PE	DiD + PE	DiD + PE	DiD + PE	DiD + PE + RCT	DiD + PE + RCT
Data Collection Rounds	Baseline, Endline	Baseline, Midline, Endline, Follow-up	Baseline, Endline	Baseline, Midline, Endline, Follow-up	Baseline, Endline	Baseline, Midline, Endline, Follow-up	Baseline, Endline	Baseline, Midline, Endline, Follow-up
Biodiversity Data Collection	No	No	No	No	Yes	Yes	No	No
Community Listing and Women's Survey	No	No	Yes	Yes	No	No	No	No
Labor, Consultant, Travel, Other Direct Costs, G&A[a]	$442,264	$884,459	$472,264	$934,459	$525,634	$1,055,725	$472,264	$934,459
"Base" Data Collection[b]	$409,875	$842,950	$409,875	$842,950	$409,875	$842,950	$409,875	$842,950
Community Listing	-	-	$23,364	$23,364	-	-	-	-
Women's Survey	-	-	$44,579	$91,680	-	-	-	-
Camera Traps[c]	-	-	-	-	$280,018	$502,443	-	-
Aerial Surveys[c]	-	-	-	-	$74,298	$152,801	-	-
Total Estimated Budget	$852,139	$1,727,409	$950,081	$1,892,453	$1,289,824	$2,553,918	$882,139	$1,777,409

Notes:

[a] Conservative staffing estimates using INRM labor rates. All scenarios assume international travel for data collection and dissemination for each data collection round. Budgets are inclusive of: Team Lead; Senior Health Advisor; Biodiversity/ Geographic Information System (GIS) Expert; In-Country Coordinator; Evaluation Director, Manager/Specialist, and Assistant, U.S. university-based research assistants and PhD students; INRM Management support.

[b] Data collection costs are illustrative only and must be refined during IE design phase through a competitive bidding process to data firms.

[c] Biodiversity measurement assumes the FA team will be collecting data from comparison areas. The extent to which costs are already covered by Panthera and Kashikoto are still to be determined.

Additional Information Needs for Evaluation Design

If USAID decides to move forward with an evaluation of the Kafue Activity, it will be important for the evaluation team to continue discussions with USAID, TNC, and other IPs during planning and implementation. Additional details on specific implementation plans and work planning for the activity will enable the team to further develop and refine an evaluation design. Some outstanding points to clarify are below:

- **Final intervention plans:** As discussed at length, the FA team will need to know precise details on *what* activities will be implemented, *where*, and *when*, for all SAs. We understand that this information will not be available until the end of Year 1 when all situation analyses and needs assessments have been completed.
- **Further details on extent and content of IP biodiversity monitoring:** In particular, it is still unclear what area will be covered by which methods, and whether the FA team will need to (1) supplement planned data collection to cover the whole of project area, or (2) replicate methods in comparison areas. In particular, the FA team looks forward to exploring data sharing arrangements with Kashikoto, Panthera, and DNPW to access data already collected across GKE and reduce primary biodiversity data collection costs.

Alignment with the USAID Program Cycle

USAID's program cycle, codified in the Automated Directive Systems (ADS) 201, is the Agency's, "operational model for planning, delivering, assessing, and adapting development programming in a given region or country to achieve more effective and sustainable results."[88] According to the ADS 201, the value of an evaluation is in its use, including to inform Agency decision-making, contribute to learning, and help improve the quality of development programs.[89] There are several ways in which findings from an evaluation of the Kafue Activity can be used for adaptive programming and to inform future programming decisions:

- Baseline findings would be an important information source for overall monitoring & evaluation of the activity. This is especially true for the proposed outcomes measured through the household surveys, as to our knowledge, these outcomes are not being measured through other data sources. Baseline findings would also provide additional contextual information to help inform adaptive programming, as well as assess support for some of the key underlying assumptions of the theory of change.

[88] USAID Learning Lab. "The USAID Program Cycle," https://usaidlearninglab.org/program-cycle-overview-page.
[89] USAID Bureau for Policy, Planning, and Learning (PPL). "ADS Chapter 201: Program Cycle Operational Policy." Revised 21 September 2021. https://www.usaid.gov/sites/default/files/documents/201.pdf.

- Supplemental analysis tracking implementation fidelity and trends in monitoring data from the IPs throughout implementation would help support USAID's adaptive management, especially considering the limited resources available to the IPs for these activities. Additionally, tracking trends in biodiversity and forest conservation could provide preliminary indicators of whether improvements are being made for these outcomes. Midline data collection, focusing on intermediate or short-term outcomes important for the theory of change, but not captured through existing monitoring data, might also be useful for adaptive management if there are resources available.
- The endline analysis at the completion of the activity would provide a comprehensive analysis of performance and impact indicators, as well as an analysis of implementation issues. Depending on the timing of USAID/Zambia's program cycle, this could either inform decision making related to continuing funding for the Kafue Activity, other integrated activities being implemented by USAID/Zambia, or other activities in the HEARTH portfolio more broadly.
- Finally, the long-term follow-up evaluating impact and sustainability post-activity completion would provide important learning for other activities in the HEARTH portfolio, USAID, and the develop community at large. Assessing long term impacts would provide more accurate inputs for estimates of value for money, which would help inform current and future programming and investments more broadly for USAID.

Ultimately, use of the findings will require coordination with the USAID/Zambia Mission and awareness of their programming needs. We recommend following principles of utilization-focused evaluation to help USAID and the IPs best make use of results and data.[90] This will include determining what kinds of reporting formats, styles, and venues are most appropriate, making sure results are delivered in time to affect important decisions, and deciding if findings merit wider dissemination.[91] For example, if we know that a decision on follow-on procurement will need to be made prior to endline data collection, there might be greater value in adding midline data collection – which, would not have conclusive findings about the overall program performance and impacts, but provide some evidence for whether things are going in the right direction. Additionally, given that we are aware at the outset that many biophysical outcomes will take a longer time scale to be realized than the typical 5-year program cycle, we would recommend that USAID/Zambia take this into consideration when making programming decisions.

[90] Patton, Michael Quinn, *Essentials of utilization-focused evaluation*, Sage, 2012.
[91] Patton, Michael Quinn, "Utilization-Focused Evaluation (U-FE) Checklist," *The Evaluation Center,* 2013, https://wmich.edu/sites/default/files/attachments/u350/2014/UFE_checklist_2013.pdf.

VII. Summary and Recommendations

Viability of an evaluation: This report seeks to determine the feasibility of a rigorous evaluation for the Kafue Activity. Our analysis finds that some Kafue Activity interventions are amenable to an IE, while other interventions are amenable to a rigorous PE. Some outcomes such as well-being, health, livelihoods, and forest conservation are more amenable to analysis through IE methods than other outcomes such as governance. The Kafue Activity overall is not amenable to a whole of project IE, and final decisions about the evaluation design for many activities will need to be made when the interventions and sites are determined at the end of Year 1.

Nevertheless, the FA team finds that the Kafue Activity presents an important opportunity to improve USAID's baseline understanding of conservation and biodiversity programming through a mixed methods evaluation. Given the dearth of counterfactual-based studies on the interventions that comprise the Kafue Activity, even knowledge generated through a well-designed PE would advance USAID's and the HEARTH portfolio's learning agenda. Furthermore, an evaluation would add value by strengthening the program's theory of change and promoting a deeper understanding of where to focus on intervention integration and quality. Baseline data will provide a key source of M&E data and provide important contextual information that can be used to promote more effective, adaptive programming.

In addition, the Kafue Activity presents a unique opportunity to explore the effect of conservation programming on biodiversity outcomes. This is due to the large amount of biodiversity monitoring that will take place as part of the program implementation; this large-scale wildlife monitoring makes it feasible to pursue a cost-effective, rigorous, and long-term study of biodiversity outcomes. Extensive observation data collection through a combination of SMART monitoring, spoor surveys, and camera traps may provide the statistical power necessary to measure biodiversity outcomes in the context of an IE approach.

Timing of baseline data collection: As described above, given the phased implementation design, the FA team recommends waiting until the end of Year 1/ beginning of Year 2 to conduct one comprehensive baseline household data collection effort after all activities and locations have been determined. The FA team could collect data October through November 2022, before the rainy season begins in earnest in December. This is the most efficient from a cost perspective, but from a technical perspective, it removes the potential to have a 'true' baseline for the resource protection components on human well-being outcomes. While this is a limitation, the FA team does not expect significant changes in important human well-being outcomes prior to this single baseline.

Need for Pause and Reflect: As discussed extensively throughout this report, the site locations and content of most interventions will not be finalized until the end of Year 1. These details will evolve over the course of Year 1 as the IPs complete their situation analyses and needs assessments. The FA team recommends a series of regular coordination and information exchange as implementation information becomes available. In addition, at the end of Year 1 the MERL plan will need to be updated and there should be a Pause and Reflect of all stakeholders. It will also be important for the evaluation team to revisit the logic model and theory of change to ensure consistency with the implementation plans as they are finalized.

Biodiversity measures: To ensure a cost-effective study, the FA team recommends building off the planned biodiversity monitoring that will be conducted as part of the IP monitoring; this includes a reliance on the planned Panthera and Kashikoto data collection activities. Overall, we recommend a combination of approaches for monitoring biodiversity outcomes that leverages existing datasets and data streams from partners with new data sources. As noted, remote sensing data, particularly forest cover, provides rich and readily available proxies for biodiversity, and will be important for contextualizing any direct biodiversity indicators. Further, forest cover loss/gain is important itself as a direct measure of forest/habitat outcomes (not just as biodiversity proxy).

Based on the team's preliminary assessment, the most likely direct biodiversity indicators will involve changes in wildlife behavior or spatial distributions near treatment sites. Camera traps provide an efficient balance between cost and field effort, yielding high-quality data for a broad taxonomic diversity of large mammals, including key target species (i.e., elephants, lions). In addition, the long record of aerial censuses in GKE, and recent SMART monitoring activities make valuable baselines for understanding biodiversity outcomes for common large-bodied species (ungulates and elephants).

Long-term evaluation: The FA team recommends that this evaluation move forward if there is a commitment to a long-term evaluation, including a follow-up endline data collection about five years after the end of the activity. The primary biophysical outcomes of interest—including impacts on forest condition, biodiversity, and impacts on human well-being from improved watershed health—will take a longer time period to materialize than the standard USAID program cycle. Furthermore, since most activities will only begin in earnest in Year 2, a longer period will be required to see program impacts.

Strong coordination and collaboration are required throughout design and implementation: A rigorous evaluation will require detailed M&E tracking of inputs, outputs, and specific site locations, along with significant coordination among the IPs and between the IPs and the evaluation team, to ensure that the design is appropriate as implementation plans evolve.

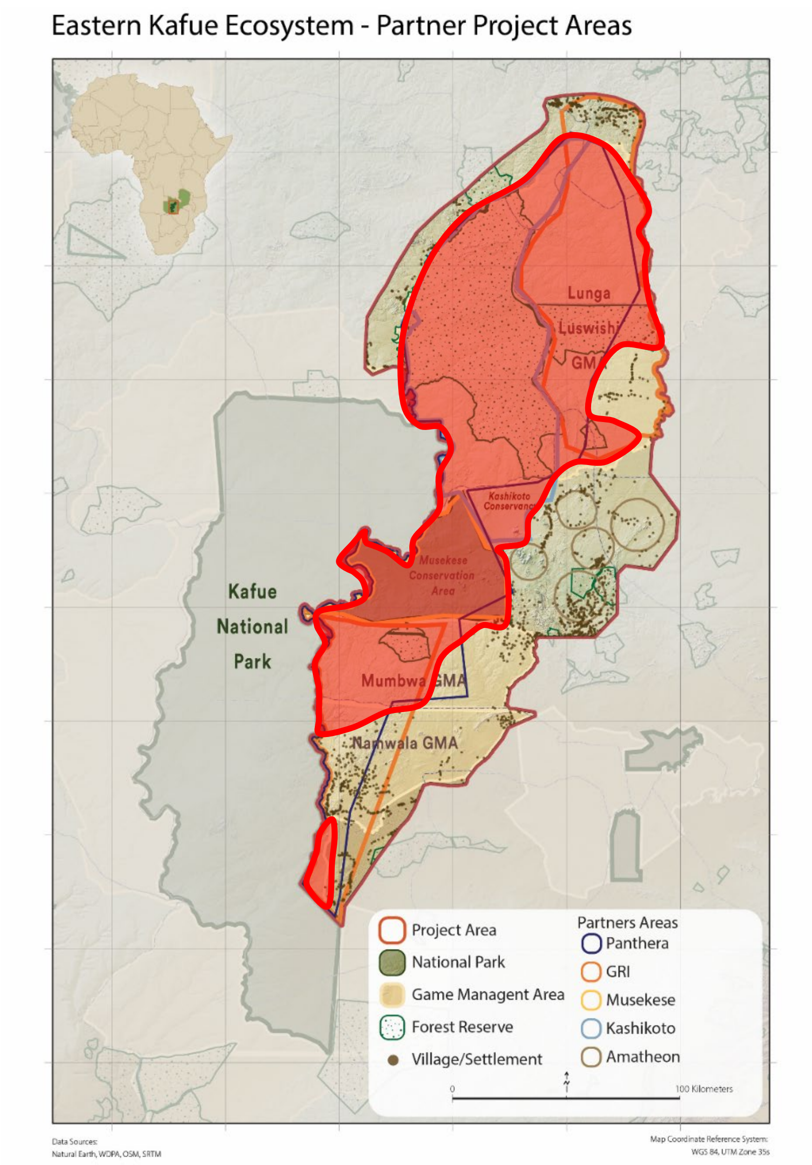
USAID and IP focus on integration and quality of programming: The FA report highlights that the extent of activity integration is unclear and will be determined at the end of Year 1. Integration is a key underlying assumption for the whole of project theory of change, as well as for the theories of change for several SAs. Thus, site selection for activities should prioritize overlapping implementation to the extent possible, to answer key learning questions related to integrated programming.

VIII. Annex 1: Kafue National Park, GMAs, and Administrative Districts

Polygons in orange with red text are administrative districts, and in dark green with black text are GMAs and the Kafue National Park.

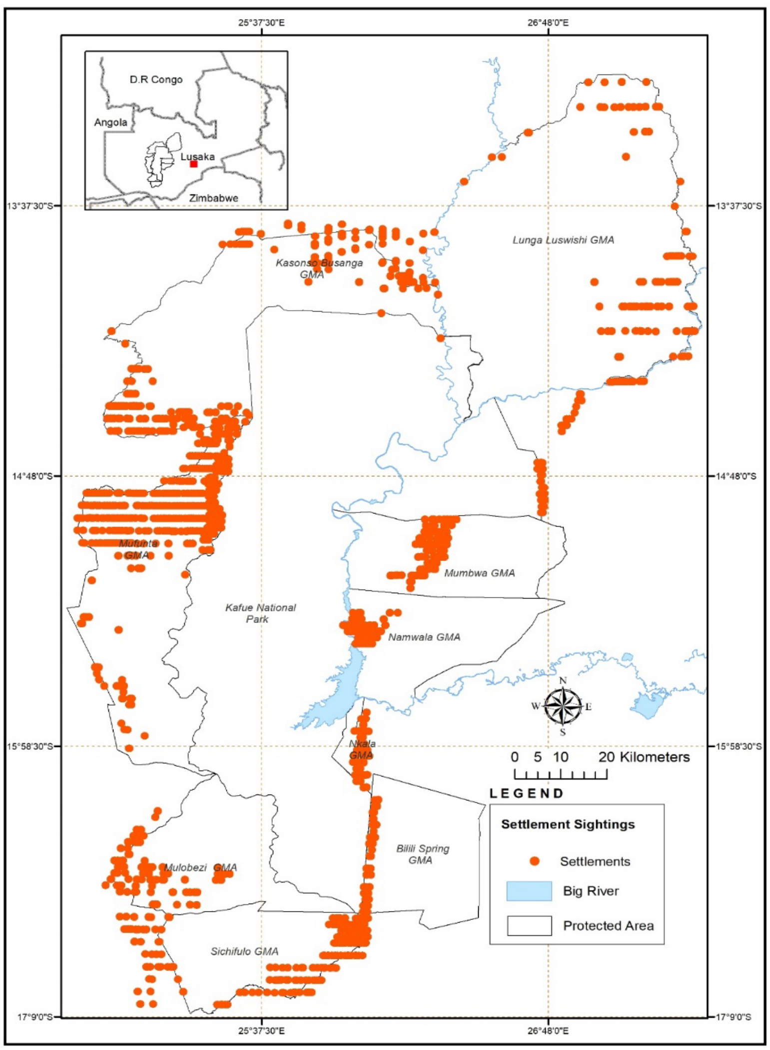
IX. Annex 2: Kafue Activity High Conservation Value Areas

The map below presents the project partners' approximate areas of operation overlaid with areas of high conservation value (see red shading).



X. Annex 3: Human Settlements in Game Management Areas surrounding Kafue National Park

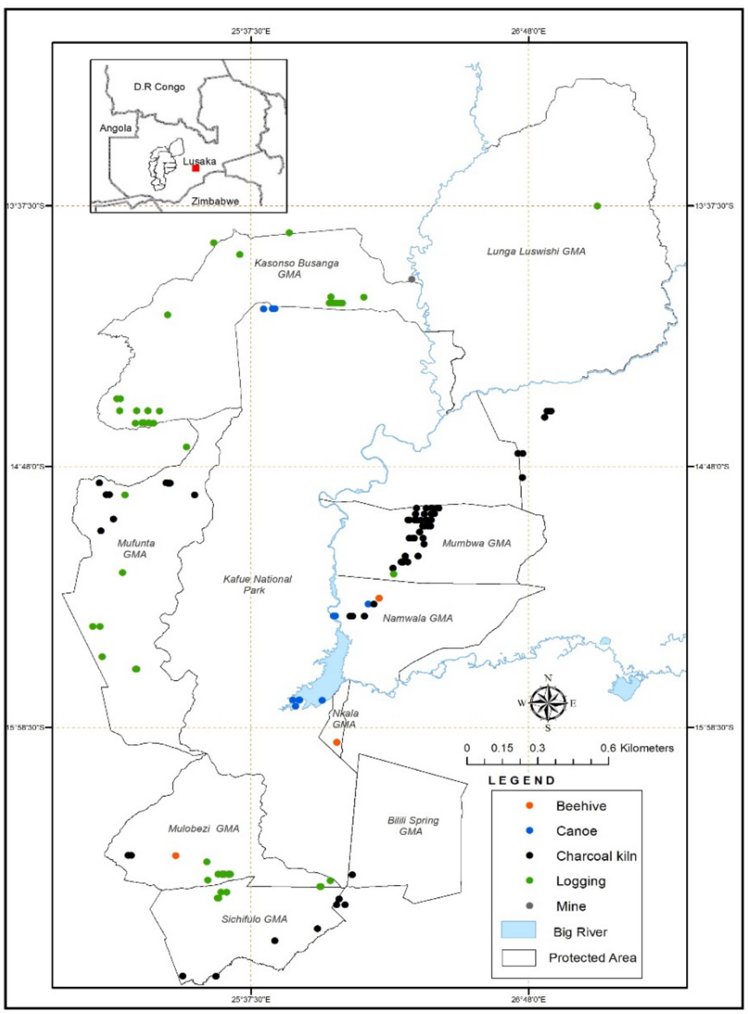
Findings provided by The Nature Conservancy.[92]



[92] DNPW (2021). Aerial Survey Report of Elephants and other Large Herbivores in the Kafue National Park and its surrounding Game Management Areas.

XI. Annex 4: Natural Resource-Based Activities in GMAs Surrounding Kafue National Park

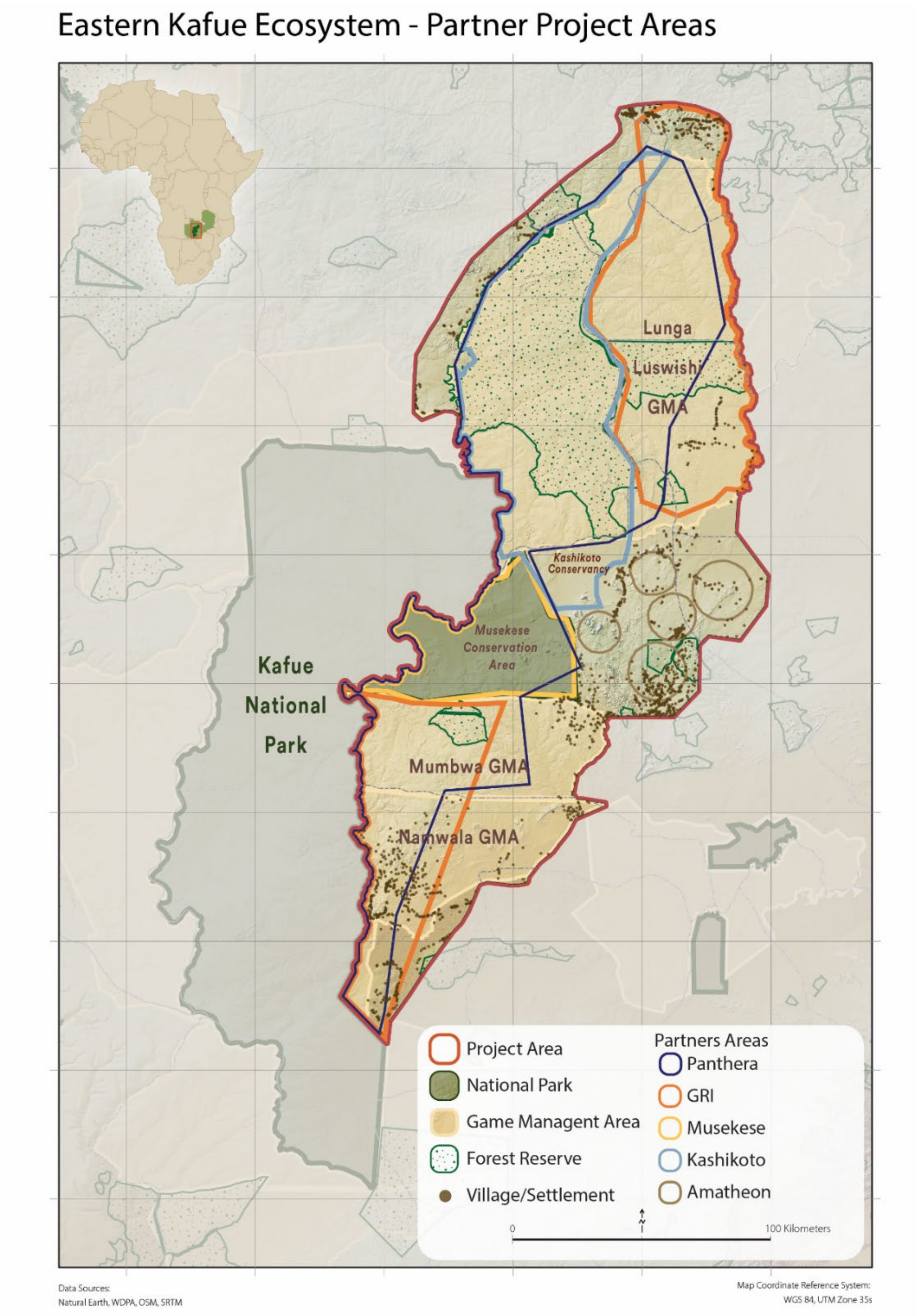
Findings provided by The Nature Conservancy.[93]



---

93 DNPW (2021). Aerial Survey Report of Elephants and other Large Herbivores in the Kafue National Park and its surrounding Game Management Areas.

XII. Annex 5: Kafue Activity Partner Project Areas



XIII. Annex 6: Simplified Year 1 Workplan

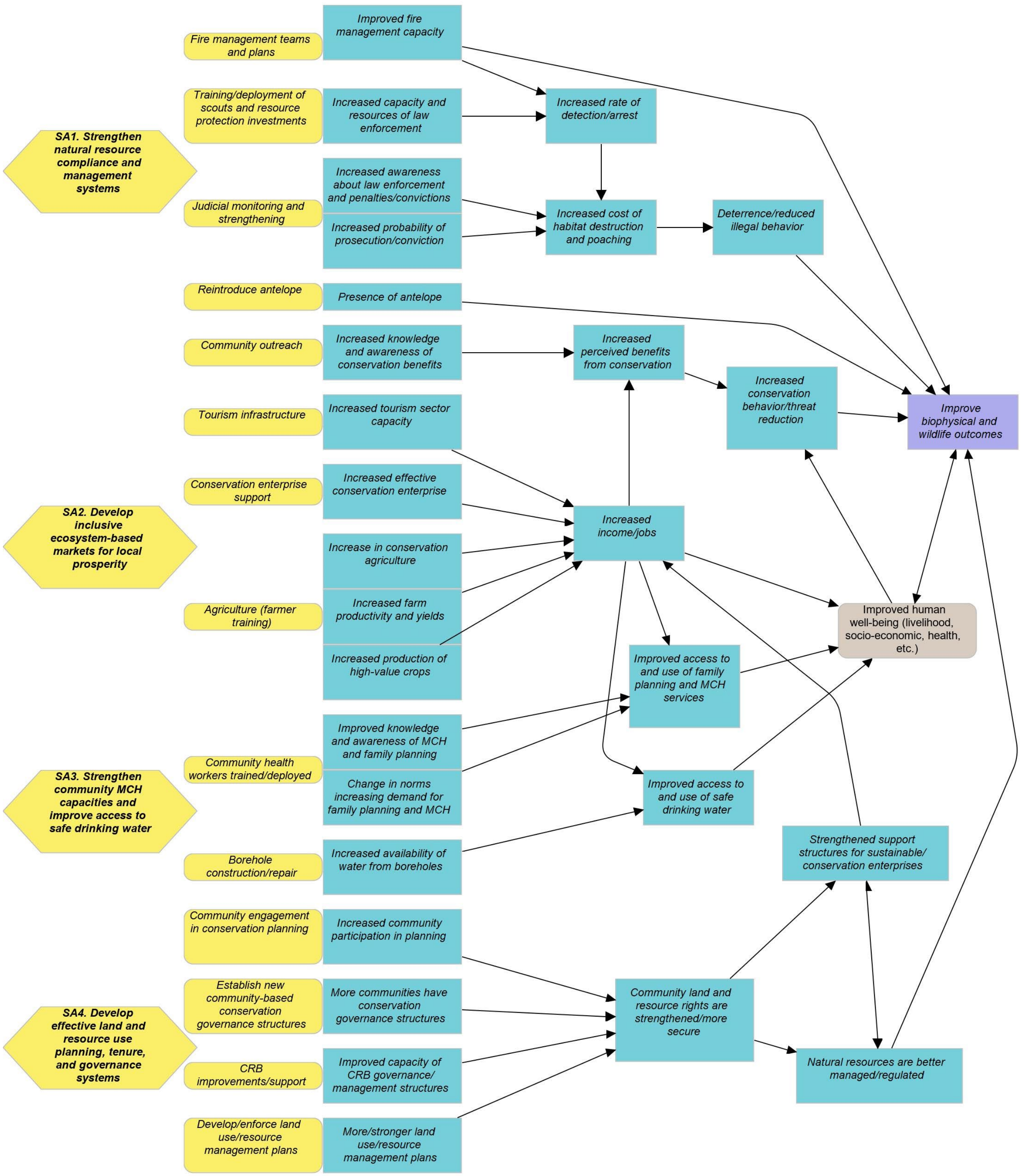
Activity start dates in red text are after the date of this feasibility assessment (and thus, the FA team could still conduct baseline data collection prior to implementation) whereas dates in black text are either on-going activities from before the Kafue activity or have already begun implementation in the first or second quarter of Year 1.

Strategic Approach	Activities	Districts	Implementing Partner	Earliest Start Date
Strategic Approach 1: Strengthen natural resource compliance and management systems in protected areas and other community conservation areas for more integrated protection and management of wildlife, forests, and fisheries	Scouts/anti-poaching patrols	Mumbwa, Kasempa	Kashikoto	On-going
		Mumbwa	Musekese	Jun-22
		Itezi Tezhi, Mumbwa, Mushindamo, Lufwanyama, Ngabwe	GRI	Jan-22
	Resource protection infrastructure/equipment	Mumbwa, Kasempa	Kashikoto	Apr-22
		Mumbwa	Musekese	Jan-22
		Itezi Tezhi, Mumbwa, Mushindamo, Lufwanyama, Ngabwe	GRI	Jan-22
	Fire management teams	Mumbwa, Kasempa	Kashikoto	May-22
		Mumbwa	Musekese	Jan-22
		Itezi Tezhi, Mumbwa, Mushindamo, Lufwanyama	GRI	Apr-22
	Courtroom and prison monitoring	Whole Project Area	WCP	Dec-21

Strategic Approach	Activities	Districts	Implementing Partner	Earliest Start Date
	Development of investigation/prosecution guidelines	Whole Project Area	WCP	On-going
	Law enforcement media/awareness	Whole Project Area	WCP	On-going
	Annual aerial survey	Mumbwa, Kasempa	Kashikoto	Oct-21
	Biodiversity assessment/monitoring	Whole Project Area	Panthera	May-22
	Focal species monitoring and protection	Whole Project Area	Panthera	Apr-22
	Reintroduction of key antelope species	Mumbwa	Kashikoto	Apr-22
Strategic Approach 2: Develop inclusive agriculture and ecosystem-based markets for local prosperity	Community outreach manager (livelihoods, conservation education, etc.)	Mumbwa, Kasempa	Kashikoto	On-going
	Explore wildlife-based business models/relationships	Whole of Project Area	TNC	Jul-22
	Farmer training and contracting	Mumbwa	Amatheon	Nov-21
	Contract smallholder farmers for the 2021-22 season	Mumbwa	Amatheon	Oct-21
	Off-take of harvested high-value crops from competing farmers	Mumbwa	Amatheon	Mar-22
	Processing and marketing of high-value crops	Mumbwa	Amatheon	May-22
	Tourism infrastructure	Mumbwa, Kasempa	Kashikoto	May-22

Strategic Approach	Activities	Districts	Implementing Partner	Earliest Start Date
	Tourism opportunity assessment/relationships	Whole of Project Area	TNC	Aug-22
Strategic Approach 3: Strengthen community MCH capacities and improve access to safe drinking water	Baseline health assessment	Kasempa, Mushindamo, Lufwanyama, Ngabwe	i4Life	Jul-22
	Develop health implementation plan	TBD	i4Life	Sep-22
	Borehole siting assessment	Whole of Project Area	TNC	Jun-22
	Establish contracts for boreholes	TBD	TNC	Sep-22
Strategic Approach 4: Develop effective land and resource use planning, tenure and community governance systems	Situation analysis	Whole of Project Area	TNC	Jun-22
	Community conservation planning	TBD	TNC	Jul-22
	New community-based conservation governance structures	Ngabwe; TBD	TNC	May-22
	CRB needs assessment	Whole of Project Area	TNC	Mar-22
	Identify land/INRM planning needs	TBD	TNC	Apr-22

XIV. Annex 7: Whole Project Theory of Change



XV. Annex 8: Choosing Target Species

True monitoring of changes in biodiversity would require accurate species identification and counting to measure changes in species abundance and indices of community diversity across all species within a community or ecosystem. However, this is typically not realistic nor feasible for most projects. Thus, choosing which species to measure is critical and should be grounded in both the desired objectives of the intervention as well as the types of management actions pursued. The HEARTH global M&E toolkit recommends that all HEARTH activities use a structured decision-making approach for identifying candidate species for monitoring. This is a systematic approach using key concepts from structured decision-making and drawing on ecological and biological knowledge and stakeholder priorities.[94] For example, the Kafue Activity may aim to reduce species decline within an area by managing threats that are thought to be acting against these species. To understand whether the activities are effective for reducing species decline, the project should identify candidate indicator species (species that are likely indicators of a biotic response to either environmental stress and/or management actions) for monitoring.

Valuable indicator species may be species that can provide early warning of biotic responses to environmental stressors, represent precursors of broader community or ecosystem-wide change, that are well-studied (clear understanding of life history and ecology), and have clear taxonomic distinctions (to avoid issues of misidentification or ongoing speciation). Candidate species for monitoring can include keystone species, area-limited umbrella species, dispersal limited species, resource-limited species, process-limited species, flagship/charismatic species, invasive species, and/or species associated with specific habitat features of interest. For example, in agricultural projects, tracking the presence of native pollinator species across the project area may be important.

Table 13. *Types of Candidate Species for Monitoring and Definitions*[95]

Type	Definition	Examples
Indicator species	Species which can indicate a response to either environmental stress and/or management actions	Water quality (macro aquatic invertebrates), illegal hunting intensity (cape buffalo), illegal logging (ebony tree), biodiversity (*Vossia* spp, an aquatic plant found on Kafue flats that

[94] P. Bal, AIT. Tullock, PF. Addison, E. McDonald-Madden, JR. Rhodes, "Selecting Indicator Species for Biodiversity Management," *Frontiers in Ecology and Evolution* 16, no. 10 (2018): 589-598, https://doi.org/10.1002/fee.1972.
[95] V. Carignan, M-A. Villard, "Selecting Indicator Species to Monitor Ecological Integrity: A Review," *Environmental Monitoring and Assessment* 78, (2002): 45-61, https://doi.org/10.1023/A:1016136723584.

Type	Definition	Examples
		provides habitat/food for grazers, insects and birds.
Keystone species	Species with disproportionally strong ecological effects, such that if this species were to decline or disappear, the ecosystem would drastically change	Predators (lions, wild dog, crocodille), prey (sable, hares), ecosystem engineers (elephants, hippos), mutualists (pollinators), plants (nitrogen fixers, like *Acacia*)
Area-limited 'umbrella' species	Species that require large areas of suitable habitat to maintain viability and whose habitat requirements also cover those for a wider array of associated species	Species with large home ranges (e.g., hooved mammals, elephants, vultures, large cats).
Dispersal-limited species	Species that are limited in their ability to move from area to area or those with high mortality risk (in moving)	Species restricted to microclimates (e.g., sky islands, humid areas, aquatic habitat) like amphibians, fish, or endemic species like Chaplin's barbet
Resource-limited species	Species requiring specific resources that might be available on a limited basis	Species that rely on specific habitats or prey species
Process-limited species	Species sensitive to ecological processes (e.g., fire, flood, grazing, competition, etc.)	Species that require fire (miombo tree species) or flood (lechwe) for survival and reproduction
Flagship species	Species that attract public support for conservation or are on priority lists (e.g., International Union for Conservation of Nature Red List species)	Sable, wild dog, elephant
Species associated with specific habitat features	Species that are strongly linked to specific habitat features, such that their persistence is closely linked to the persistence of that feature.	Bird species that are closely linked to a habitat type (e.g., ovenbirds are indicators of a closed-canopy mature forest with a sparse understory)
Invasive species	Species that are non-native to the particular area. Typically, the term 'invasive' species is given to non-native species that grow and reproduce quickly and spread	Redclaw crayfish, water hyacinth, *Mimosa pigra*

Type	Definition	Examples
	aggressively, with the potential to cause harm to native species and/or ecosystems.	

XVI. Annex 9: Supplemental Methods for Species Monitoring Considered

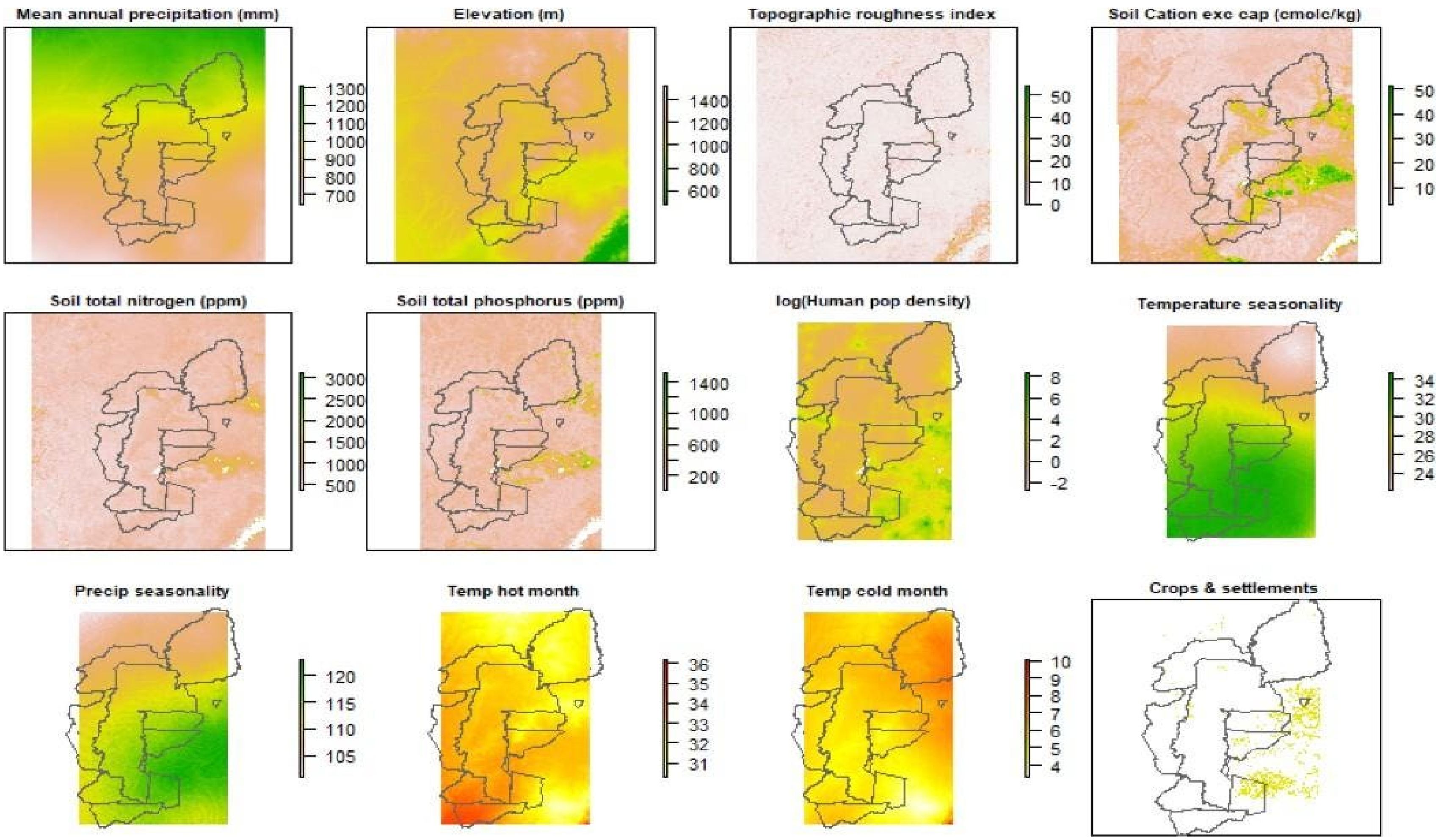
Method	Tools Required	Considerations for Monitoring
Satellite-based counts	High resolution satellite imagery (<1m), computers and software	Recent tests have shown that satellite imagery can be used to census large-bodied species (e.g., elephants) across large areas, though has not yet been widely applied as a wildlife census technique except in marine mammals. Machine learning or neural networks (similar to camera trap identification) can be used to automate the count process, though require considerable technical expertise and support from GIS. Similar body sized animals cannot be differentiated. This approach might be cost-effective if the team can access free, high-resolution satellites and low-cost technical expertise. However, ultimately, this method is not recommended because (1) it is limited to elephants, whereas the other counting methods can capture multiple species, and (2) this method works best in areas with very high densities of elephants,[96] which is not the case in GKE.
Individual identification based on unique markings	Binoculars, surveying equipment	Very resource intensive, difficult to implement over large areas and requires dedicated personnel to regularly survey individuals. Generates high quality demographic information, such as abundance, life history and movement patterns. In GKE, likely only to be relevant for target carnivore species (e.g., lion, wild dog and hyena), elephants, giraffe, and threatened reptiles or amphibians. The large amount of time required to develop a registry of individuals makes it unlikely

[96] Gill, 2021.

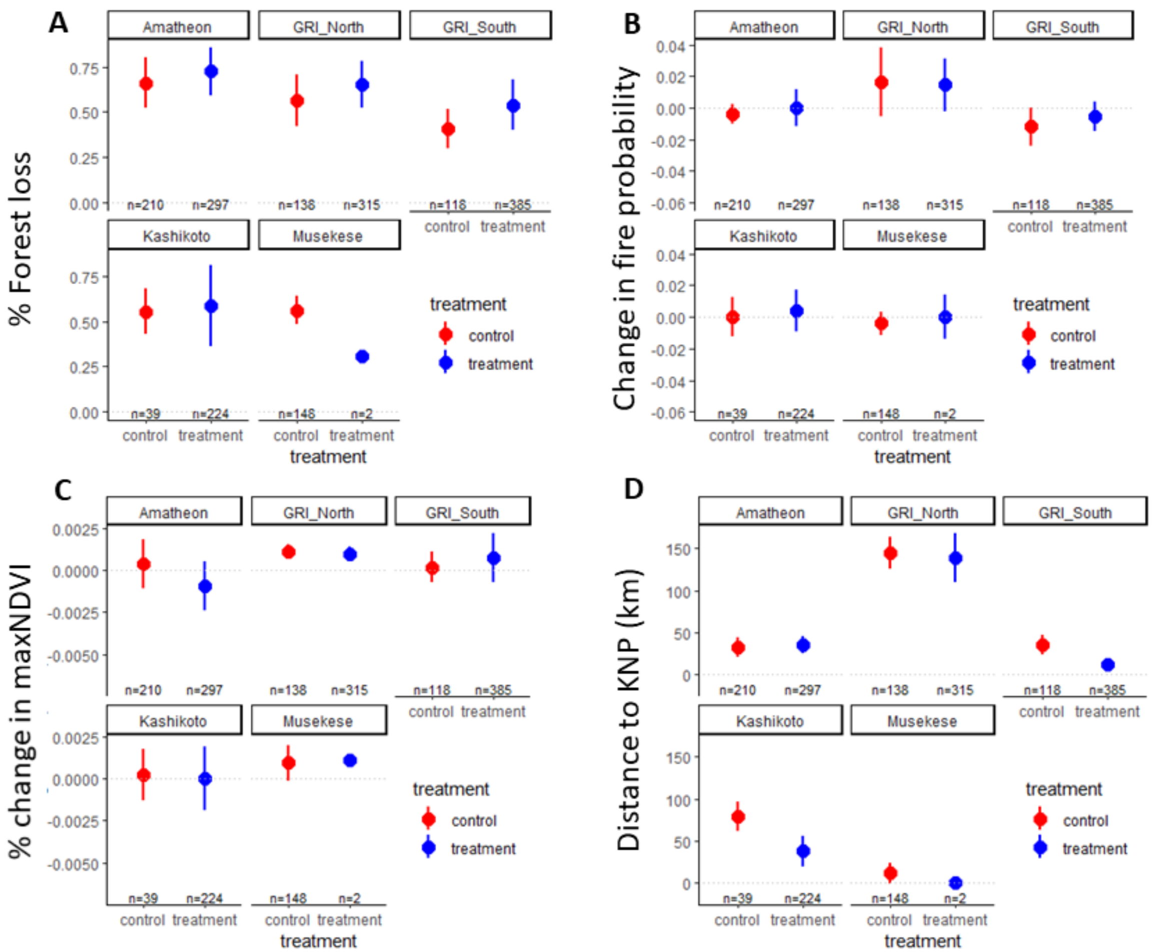
Method	Tools Required	Considerations for Monitoring
		to be a viable option for assessing impacts of Kafue Activities, unless research projects are already ongoing.
Genetic and chemical tagging: Tools to remotely track animal movements	Different types of technical options for tracking movements of wildlife across protected area boundaries. Genetic and chemical tagging can be used to determine population connectivity (e.g., dispersal, gene flow) across discrete sites.	Provides coarse information about individual-level patterns of movement (e.g., degree of connectivity between areas), changes in behavior (e.g., avoidance of human infrastructure or protected area boundaries), patterns of habitat use (e.g., preference for forest versus grassland sites), conflict behaviors (e.g., pattern of crop-raiding by elephants or livestock interaction from carnivores). Can be resource-intensive, requiring specialist instruments (genetic or chemical) and analytical skills, and are often difficult to implement over large spatial scales. May be useful in estimating genetic connectivity with other nearby ecosystems (e.g., Kavango Zambezi ecosystem), though detection of changes in movement over the course of the program would only realistically be feasible with large sample sizes and large numbers of dispersing individuals.
Environmental deoxyribonucleic acid (eDNA)	Reagents for preserving genetic material (e.g., RNAlater, ethanol, etc.), equipment for preserving deoxyribonucleic acid (DNA) (e.g., dry shippers, storage vials, other types of temperature control), molecular laboratory equipment, data storage, computing software for analyzing data	Environmental DNA (eDNA) can be useful for tracking presence/absence of a wide range of cryptic species, or early detection of invasive species, with little disturbance to habitat. eDNA is a relatively new tool for species monitoring, but has been used successfully in many different contexts, primarily aquatic ecosystems. In aquatic settings, the method is relatively cheap (in terms of catch per unit effort) compared to alternative methods (electrofishing). Soil and recently air can also be sampled. The method may not be appropriate to monitor species abundance and will likely miss very rare or ephemeral species. eDNA requires access and funding to laboratory resources and analysis and expertise for analyzing data.
DNA barcoding	Reagents for preserving genetic material (e.g., RNAlater, ethanol, etc.),	DNA barcoding can be useful for tracking species abundance and presence/absence, particularly when visual identification of species is either

Method	Tools Required	Considerations for Monitoring
	equipment for preserving DNA (e.g., dry shippers, storage vials, other types of temperature control), molecular laboratory equipment, data storage, computing software for analyzing data	difficult or not feasible. Barcoding uses a conserved sequence of DNA that is present in all species. Each species has a unique DNA fingerprint or barcode that differentiates them from other species. DNA barcoding is particularly useful in monitoring contexts where whole bodies are not present (e.g., meat and fish markets, parts of plants) or where cryptic species may be abundant (genetically distinct species that are not easily distinguished visually). Like genetic tools listed above DNA barcoding requires access to laboratory resources and analysis and expertise for analyzing data.
Acoustic monitoring	Acoustic recorders (e.g., 'Audio Moth'), batteries, computing software for analyzing data	Can survey presence-absence and relative abundance of a wide range of species with unique vocalizations, e.g., songbirds, bats, elephants and carnivores. New pipelines for automating species identification (particularly birds) using machine learning greatly reduces processing time, though requires some specialist technical skills. Devices are small and inconspicuous so can be deployed with less risk of damage or theft than camera traps.
Freshwater fish surveys	Species identification guides; data recording devices; fishing equipment	Fish species composition and abundance can be surveyed directly with low-cost fishing equipment (e.g., nets, lines), and slightly more specialized electroshock equipment that sends a mild electric field through the water to temporarily stun fish. Electroshocking is unsuitable in areas with hippos and crocodiles. Indirectly surveys include examining the harvest of fishermen returning to shore. A key indicator of harvest intensity is the size structure of harvested fish. Biases may be introduced in harvest surveys if fishermen selectively keep certain types/sizes of fish.

XVII. Annex 10: Biophysical and Climate Variables



XVIII. Annex 11: Long-term Settlement Level Trends in Biodiversity Indicators



Control and treatment settlements were largely similar in terms of (A) forest cover loss (2000-2020), (B) change in the probability of fire (2000-2021), (C) percent change in the maximum annual NDVI (2000-2021) and (D) distance to the Kafue National Park boundary across different partner organizations. Control settlements were compared to the nearest treatment settlement from a particular IP or combination of two IPs (Panthera villages not shown because of overlap with other IPs). All values are calculated as the mean (+/- Significant Difference) within 5km of the center of the settlement. This initial analysis suggests that, for these variables, control and treatment settlements have relatively similar patterns except for distance to the Kafue National Park in Kashikoto where control sites were much farther than treatment sites.