SOCIAL IMPACT

ADVANCING DEVELOPMENT EFFECTIVENESS

MCC Sierra Leone Threshold Program

Evaluation Design Report v.5

EVALUATION DESIGN REPORT

MCC Sierra Leone Threshold Program

v.5

Submitted: August 2019

Submitted to:
Millennium Challenge Corporation
875 Fifteenth Street, NW
Washington, DC 20005-2221
Contracting Officer's Representative: Shreena Patel
Contract No. 95332418A0022

Submitted by:
Social Impact, Inc.
2300 Clarendon Blvd., Suite 1000
Arlington, VA 22201
703.465.1884
www.socialimpact.com

DISCLAIMER

The authors' views expressed in this publication do not necessarily reflect the views of the Millennium Challenge Corporation (MCC) or the United States Government.

CONTENTS

1 Introduction & Background 1
1.1 Country context 1
1.2 Objectives of this Report 1
2 Overview of Threshold & Interventions 3
2.1 Overview of the Project and Implementation Plan 3
2.1.1 Original Project Description 3
2.1.2 Project Participants and Geographic Coverage 4
2.2 Theories of Change 5
2.2.1 Water Sector Reform Project 6
2.2.2 Electricity Sector Reform Project 8
2.2.3 Regulatory Reform Project 8
2.3 Literature Review 9
2.3.1 Summary of Existing Evidence 9
2.3.2 Gaps in Literature 15
2.3.3 Policy Relevance of the Evaluation 15
3 Evaluation design 17
3.1 Evaluation Questions 17
3.1.1 Electricity and Water Reform Projects and the Regulatory Strengthening Project 17
3.1.2 District Metering Area and Standpipe Demonstration (Kiosk) Activity 18
3.2 Evaluation Design Overview 19
3.3 Performance Evaluation Approach 22
3.3.1 Methodology 22
3.3.2 Timeframe of Exposure 36
3.3.3 Study Sample 37

3.3.4 Primary Data Collection 39
3.3.5 Analysis Plan 41
3.4 DMA and Standpipe Demonstration (Kiosk) Activity Evaluation Approach 41
3.4.1 Methodology 41
3.4.2 Timeframe of Exposure 52
3.4.3 Study Sample 52
3.4.4 Primary Data Collection 54
3.4.5 Analysis Plans 58
3.5 Evaluation Challenges & Limitations 59
3.5.1 Limitations of Interpretations of the Results 59
3.5.2 Risks to the Performance Evaluation Design 59
3.5.3 Risks to the DMA and Standpipe (Kiosk) Evaluation Design 59
3.5.4 EDR Modifications 60
4 Administrative 61
4.1 Summary of IRB Requirements and Clearances 61
4.2 Data Protection 61
4.3 Preparing Data for Public Use 62
4.3.1 Access and Documentation 62
4.3.2 Privacy 64
4.4 Dissemination Plan 65
4.5 Evaluation Team Roles and Responsibilities 65
4.6 Evaluation Timeline and Reporting Schedule 67
5 References 70
6 Annexes 73
6.1 Annex 1: Process mapping of Tariff setting process 73
6.1.1 Step 1: Process Overview 73
6.1.2 Step 2: Process Boundaries 74
6.1.3 Step 3. Gap identification 76

6.1.4 Step 4. Follow up Actions 77
6.2 Annex 2: Stakeholder Comments and Evaluator Responses 78
6.2.1 MCA Comments and Evaluator Responses 78
6.3 Annex 3: Evaluation Budget 100
6.4 Annex 4: WSRP Logic Model 101
6.5 Annex 5: ESRP Logic Model 102
6.6 Annex 6: Household Sample Size Calculations 103
6.7 Annex 7: Baseline Memo 107
6.8 Annex 8: Approved Evaluation Activities 154

TABLES & FIGURES

TABLES

Table 1: Implementation Phases of Electricity Sector Reform Activities in Sierra Leone 13
Table 2: Evaluation Questions, Outcomes and Methods (Design Matrix) 21
Table 3: Key Informants by Sector 24
Table 4: Process Mapping Activities 29
Table 5: PE Outcome Level Measurements 30
Table 6: PE Respondents 37
Table 7: Distribution of Kiosks 44
Table 8: DMA Outcome Level Measurements 50
Table 9: Sample size based on the EAs 53
Table 10: Summary Table for Rounds of Data Collection 57
Table 11: Documentation 62
Table 12: Evaluation Team Roles 65
Table 13: Data Collection Timeline 67
Table 14: Detailed Breakdown of Interim and Endline Activities 67
Table 15: Required sample sizes for different values of expected change overtime (Scenario 1 with matching) .. 105
Table 16: Required sample sizes for different levels of expected change overtime (Scenario 2 longitudinal study) 106

FIGURES

Figure 1: GVWC's served and unserved areas in Freetown (Source: Freetown Water Use - Annex A) 5
Figure 2: Service Delivery System Map 27
Figure 3: RBF Process Mapping 28
Figure 4: SI's EQUI® System for High-Quality Data ..... **Error! Bookmark not defined.**
Figure 5: SI's Data Quality Assurance Strategies 56

ACRONYMS

AFDB	African Development Bank
APSEB	Andhra Pradesh State Electricity Board
ASI	Adam Smith International
AWSP	Alternative Water Services Providers
CA	Constraints Analysis
CBA	Cost-Benefit Analysis
CMS	Customer Mapping Survey
DID	Difference in Difference
DMA	District Metering Area
DO	Direct Observation
DRB	Disclosure Review Board
EA	Enumeration Area
EDR	Evaluation Design Report
EDSA	Electricity Distribution and Supply Authority
EGTC	Electricity Generation and Transmission Company
EMC	Evaluation Management Committee
EQ	Evaluation Question
EQUI	Evaluation, Quality, Use, and Impact
ESRP	Electricity Sector Reform Project
ET	Evaluation Team
EWRC	Electricity and Water Regulatory Commission
FGD	Focus Group Discussion
GoSL	Government of Sierra Leone
GVWC	Guma Valley Water Company
HH	Household

ICC	Inter Cluster Correlation
IE	Impact Evaluation
IPP	Independent Power Producer
IRB	Institutional Review Board
ITSA	Interrupted Time Series Analysis
ITT	Indicator Tracking Table
JMP	Joint Monitoring Program
KAP	Knowledge, Attitude and Practice
KII	Key Informant Interview
MCC	Millennium Challenge Corporation
MCCU	Millennium Challenge Coordinating Unit
MOHS	Ministry of Health and Sanitation
MWR	Ministry of Water Resources
NPA	National Power Authority
NRW	Non-Revenue Water
OCA	Organizational Capacity Assessment
OPM	Oxford Policy Management
PE	Performance Evaluation
PII	Personally Identifiable Information
PPA	Power Purchase Agreement
PPP	Private-Public Partnerships
PSM	Propensity Score Matching
PSP	Private Sector Participation
QA	Quality Assurance
RBF	Results Based Financing
RSP	Regulatory Strengthening Project
RSTDS	Regulatory Strengthening and Tariff Development Support

SALWACO	Sierra Leone Water Company
SI	Social Impact
SI-HQ	Social Impact Headquarters
SLIHS	Sierra Leone Integrated Household Survey
SOE	State Owned Enterprise
SOW	Scope of Work
SSA	Sub-Sahara Africa
THP	Threshold Program
ToC	Theory of Change
ToR	Terms of Reference
UC	University of Colorado
WASH	Water, Sanitation, and Hygiene
WB	World Bank
WHO	World Health Organization
WSRP	Water Sector Reform Project

1 INTRODUCTION & BACKGROUND

1.1 Country context

Sierra Leone has consistently ranked in the bottom ten of all African countries in terms of infrastructure development, largely due to poor service provision in the energy and water sectors.[1] The state of infrastructure is particularly poor in Freetown, which is still suffering the consequences of physical damage incurred during the country's decade-long civil war. In addition, as a result of the Ebola epidemic in 2014, the water and electricity sectors saw a reduction in investment as government money was shifted to provide an emergency response to the outbreak.[2] In urban areas, one-quarter of the population lacks access to safe drinking water, and less than half have access to basic sanitation.[3] In the capital of Freetown, the Guma Valley Water Company (GVWC) is responsible for the provision of water, and reports suggest that a lack of distributional capacity, poorly maintained infrastructure, and financial mismanagement all play a role in its inability to provide adequate services.[4]

The situation is equally as dire in the power sector, with only 11 percent of the urban population able to access electricity.[5] Demand is more than five times the current national generation capacity. Even in urban areas, electricity supply is only available sporadically due to aging power plants and hydroelectric plants rendered ineffective during long dry seasons.[6] The two government agencies responsible for electricity provision are the Electricity Generation and Transmission Company (EGTC) and Electricity Distribution and Supply Authority (EDSA), which were created in 2011 as a result of the unbundling of the National Power Authority (NPA). While the intention of the unbundling was to improve transparency and accountability, accusations of corruption and bribery still plague the two electricity providers as energy needs remain unmet.[7]

1.2 Objectives of this Report

This report has four primary objectives. The first is to provide an overview of the Sierra Leone Threshold Program (THP) including the program design, logic, and existing research in the water and electricity sectors. The second is to communicate the purpose of the Sierra Leone THP independent evaluation by providing the guiding research questions behind the design. The third is to define the quantitative and qualitative methods Social Impact (SI) has chosen to respond to those questions. The evaluation design encompasses the chosen methodology, sampling procedure, sample size, data collection tools, and

[1] The African Development Bank Group Chief Economist Complex (2016). *The Africa Infrastructure Development Index 2016*.

[2] Howard, Marjorie. "The Other Cost of Ebola." *TuftsNow.* 20 January 2015.

[3] WHO/UNICEF JMP. Rural and urban drinking water service levels (2015), https://washdata.org/.

[4] Niedinger, Jennifer. "Mapping Freetown's Water Pipes to Improve Service Delivery." *Dipnote: U.S. Department of State Official Blog.* 22 March 2018.

[5] USAID. "POWER AFRICA IN SIERRA LEONE Fact Sheet." *Archive - U.S. Agency for International Development*, 2012-2017.usaid.gov/powerafrica/sierraleone.

[6] Ashley-Edison International. "Voltage in Sierra Leone." *Voltage in Sierra Leone- Electricity Supply and Power Quality Overview*, www.ashleyedisonuk.com/voltage-sierra-leone-907/.

[7] Thomas, Abdul Rashid. "Sierra Leone electricity supplier under corruption investigation" *Sierra Leone Telegraph.* 4 September 2016.

analysis plan. Potential challenges and risks to the research design are also explored. Finally, the report outlines SI's administrative approach to data protection and respondent privacy plans, findings dissemination plans, team roles and responsibilities, and the survey timeline. A desk review, design trip to Sierra Leone, Evaluability Assessment, Baseline Memorandum document, and follow-up consultations with Millennium Challenge Corporation (MCC) informed this report.

2 OVERVIEW OF THRESHOLD & INTERVENTIONS

2.1 Overview of the Project and Implementation Plan

To address challenges facing the water and electricity sectors in Sierra Leone, The Millennium Challenge Corporation (MCC) established a $44.4 million Threshold Program agreement with the government of Sierra Leone in 2015.

"The Sierra Leone Threshold Program (THP) focuses on two binding constraints identified in Sierra Leone's Constraints Analysis (CA): (1) lack of access to reliable and affordable electricity and (2) lack of access to clean water and sanitation. The CA revealed that Sierra Leone's limited and aging electrical grid imposes additional costs on the economy in the form of lost output due to electrical outages and higher energy costs paid by households and firms using generators, batteries, or household fuels. Similarly, weak water supply infrastructure and accompanying water-borne diseases impose a high shadow price on Sierra Leone's economy, as indicated by comparatively high out-of-pocket health expenditures for households and low labor productivity for firms. Importantly, for the purposes of MCC's Threshold Program, the Constraints Analysis also highlighted that these constraints are exacerbated by weak sector governance and institutional capacity."[8]

Three projects were designed as THP focus areas:

(1) the Regulatory Strengthening Project (RSP);

(2) the Water Sector Reform Project (WSRP); and

(3) the Electricity Sector Reform Project (ESRP).

The objective of the RSP is to support the Government of Sierra Leone (GoSL) as it works to operationalize and build capacity at the newly established regulator, the Electricity and Water Regulatory Commission (EWRC); the Water Sector Reform Project (WSRP) addresses the issue of inadequate access to clean water by focusing on activities designed to provide technical assistance, build capacity, and establish a district metering area to test new management approaches and reduce non-revenue water (NRW); and the Electricity Sector Reform Project (ESRP) is designed to address challenges associated with insufficient, unaffordable, and unreliable access to electricity by helping key institutions develop a performance roadmap and action plan while also providing these entities with technical assistance. The expected outcome of these three initiatives is a foundation for more effective and sustainable services to ultimately attract sector investment and directly benefit households in Freetown.

2.1.1 Original Project Description

From its inception, the THP sought to build a foundation for the more effective and financially sustainable provision of essential urban services – water and electricity, with a focus on greater Freetown – in order to attract and sustain increased sector investment. To do so, the THP focuses on strengthening key institutions to improve service delivery and supporting reforms to increase service provision transparency

[8] Millennium Challenge Coordinating Unit. *Monitoring and Evaluation Plan - Version 3*. MCCU, 2018, *Monitoring and Evaluation Plan - Version 3*.

and accountability to limit opportunities for corruption and petty bribery. Specifically, the Program seeks to:

> *"1. Establish effective, independent regulation of the water and electricity sectors, including a framework for transparent tariff setting, and improve sector governance by better delineating institutional roles and responsibilities;*
>
> *2. Improve commercial practices, operational independence, and planning capacity in water and electricity sector institutions through targeted technical assistance and capacity-building;*
>
> *3. Enhance transparency, accountability, and customer service practices through stakeholder (community, consumer, and customer) engagement, outreach, and communication, and the establishment and utilization of new mechanisms to fight corruption and petty bribery; and*
>
> *4. Test innovative approaches to achieve sustainable delivery of water and electricity that could serve as models of effectiveness with the potential to scale for systemic impact."* [9]

The **Regulatory Strengthening Project** aims to establish core regulatory functions and capacities at the EWRC, as well as transparent tariff setting procedures. The project involves conducting at least one full tariff case review in each sector to test new processes and coordination, build commercial discipline, and provide an independent and transparent assessments of costs. Furthermore, at least $5.5 million was earmarked for a Results Based Financing (RBF) activity as additional incentive for regulated institutions.

The **Water Sector Reform Project** aims to provide technical assistance and an institutional roadmap for urban WASH practices. It supports the establishment of a Water, Sanitation, and Hygiene (WASH) sector steering committee with guidance and coordination under WSRP. Institutional strengthening activities focus on improving core business functions and regulatory compliance for Guma Valley Water Company (GVWC), the primary supplier and distributor of water in Freetown.

In addition, the WSRP will pilot a District Metering Area (DMA) and Standpipe Demonstration activity (referenced in this report as the DMA/Kiosk activity) in two DMA's in Freetown as an innovative approach for water service provision and payment. Both the WSRP program at large and pilot activity are included in the THP evaluation.

The **Electricity Reform Project** aims to provide technical assistance and coordination among the Electricity Distribution and Supply Authority (EDSA) and the Electricity Generation and Transmission Company (EGTC), as well as the EWRC and the Ministry of Energy.

2.1.2 Project Participants and Geographic Coverage

As the THP is aimed at improving institutional operations, program participants and those covered in the scope of this evaluation are largely individuals working or connected to targeted institutions including but not limited to EWRC, EDSA, EGTC and GVWC. These individuals and institutions will receive assistance through the THP in the form of trainings, technical assistance, and coordination. Additionally, GVWC will partake in the DMA/Kiosk activity in two DMAs in Freetown. Throughout program documentation and within this report, these individuals and their places of work are often referred to as "beneficiary institutions" or "benefiting entities" primarily located in Freetown. This is not to be confused with MCC's

[9] Ibid

standard definition of beneficiary, which is limited to "those individuals who realize economic gains, preferably in the form of higher real lifetime income, that are attributable to an MCC-funded project"[10] and typically quantified through a Cost-Benefit Analysis (CBA).

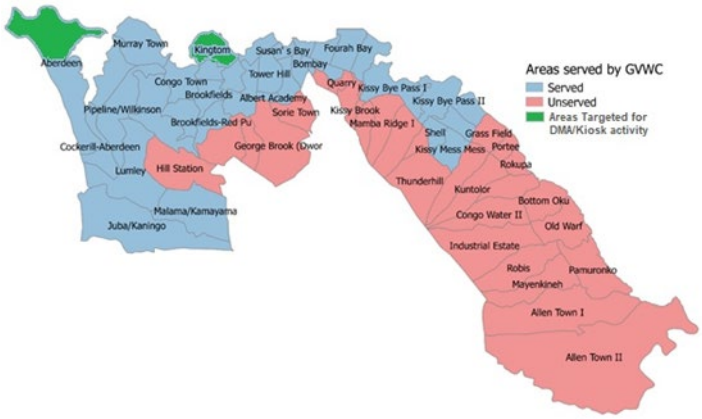


Figure 1 : GVWC's served and unserved areas in Freetown

(Source: Annex A- Freetown Water Use)

THPs are not required to produce CBAs and therefore do not quantify the number of beneficiaries or the distribution of those benefits.[11] However, it is not fair to say that the intended effects of the THP should be limited to project participants at beneficiary institutions alone. Given the nature of the THP, especially around coordination and transparency in water and electricity, the intended beneficiaries could be defined broadly as the citizens of Sierra Leone. Under the ESRP, this primarily includes households (estimated in the M&E Plan at around 58 percent in Freetown) and businesses currently connected to the electricity grid. Under the WSRP this would include households who currently source their water from GVWC in Freetown. According to the 2011 Sierra Leone Integrated Household Survey (SLIHS), 75 percent of households in Freetown receive their water from GVWC.[12] This figure should be inclusive of not just household hook-ups, but also those who receive their water through standpipes, satchels, etc. The DMA/Kiosk activity allows for an additional level of specificity focused on individuals in the targeted DMAs of Aberdeen and Kingtom.

2.2 Theories of Change

At its most basic, the Theory of Change (ToC) is consistent across all three projects and can be summarized as follows: if improvements can be made to strengthen institutional capacity, technical

[10] Millennium Challenge Corporation. "Beneficiary Analysis." Beneficiary Analysis, Millennium Challenge Corporation, www.mcc.gov/our-impact/beneficiary-analysis.

[11] Millennium Challenge Coordinating Unit. Monitoring and Evaluation Plan - Version 3. MCCU, 2018, Monitoring and Evaluation Plan - Version 3.

[12] This figure varies depending on the report referenced. Annex A of the Freetown Water Use report notes that only 63 percent of household rely on GVWC water. Source: Millennium Challenge Coordinating Unit. Annex A- Freetown Water Use. MCCU

capability, and coordination, then the generation and transmission of key services can be improved, and households and businesses can utilize those services for economic growth. Whereas compacts may be largely focused on infrastructure development, community awareness and education, and economic initiatives, the THP recognizes that sustainable change is best achieved once stable institutions and policies are in place.[13]

The logic of the overall ToC is considered valid. However, within the documentation provided, in which the ToC is disaggregated by sector, there exists a lack of assessment/self-reflection of the project's underlying assumptions, which may influence the intended results from materializing in practice, especially within the limited timeframe of the THP.

Failure to consider these assumptions makes it harder to identify and mitigate risks. For example, the ToC for each sector outlines a linear relationship among process, output, and outcome. It fails to include how non-linear effects create changes in and interactions with the influence of national and local politics on institutions, the importance of incentives in driving institutional change, donor coordination, the optimal sequencing of activities to maximize synergies, and the time required for achieving institutional reform in a complex operating environment as in Sierra Leone. In changing environments (such as government change or a change in leadership in GVWC in late-2018, for example) a large concern to SI is that these factors may affect the pathway of the logic model or ToC (at least in terms of time if not structurally) as the project progresses. This has, indeed, already been realized as the THP explores a potential time extension on several key components including the DMA/Kiosk and Results Based Financing (RBF) activities. The evaluation provides an opportunity to revisit the planned process and assess where assumptions were made, what deviations from the planned program occurred and, at least qualitatively, what was the end outcome.

2.2.1 Water Sector Reform Project

The WSRP consists of three key activities each aimed at the long-term outcomes of improved financial and operational efficiency in the sector, and at GVWC, the main supplier of water in Freetown. They include: the Urban WASH Sector Roadmap and Coordination Activity; the Guma Valley Water Company Institutional Strengthening Activity and the DMA/Kiosk Activity.[14] In theory, all these actions contribute to the intended outcomes of improved financial and operational efficiency in the sector and, eventually, improved provision of services and household welfare. However, the model, which can be viewed in Annex 4, is dependent on a number of key factors, some of which are (or are closely related to) tasks MCC has assigned to various contractors, while others relate more to external or environmental factors.

[13] Within this report, SI examines the logic model as provided in version 3 of the M&E plan. However, it is important to note that this model has been updated overtime (see 3.0 Review of Logic model in MCCU's Mid Term Review Report). As part of the evaluation, especially in regard to EQs 1 and 2, SI will revisit the original logic model and engage with stakeholders to understand the decisional factors that led to changes in the program logic overtime.

[14] The Urban WASH Sector Roadmap and Coordination Activity aimed at the development of a water sector roadmap and steering committee with the intention that improved interactions among stakeholders with clearly defined roles would lead to greater sector coordination and efficiencies. The Guma Valley Water Company Institutional Strengthening Activity aimed at the provision of technical assistance, customer and network mapping and increased customer engagement in order to improve through advanced information the efficiency, cost, billing and collection of services offered. The District Metering Area and Standpipe Demonstration Activity (DMA/Kiosk Activity) aimed at establishing DMAs within GVWC's system and exploring private-public partnerships (PPP) through the establishment of kiosks in order to isolate sources of NRW, improve water reliability and quality and to increase customer satisfaction and billings.

The evaluation team (ET) will track these factors, as failure in any one could lead to a break in the results chain.

1. **The developed roadmap will go beyond a general outline of sector priorities and provide clear and actionable steps in the immediate and long-term.** This requires not only the identification of key players who will take up the roles and responsibilities outlined in the roadmap, but also a clear picture of the current situation in the sector with data provided by invested parties.
2. **Key sector players will be available and committed to change.** It can be incredibly complex to bring together the separate interests of government entities, donors, and the private sector and ensure continued and quality engagement[15] over an extended period of time when key personnel at these institutions may undergo change.
3. **Technical assistance will be appropriately targeted and sustainable.** This can be difficult depending on the level of staff turnover in trained positions and a potentially high dependence on additional inputs in the form of hardware and/or software.[16]
4. **Capacity constraints have been appropriately identified and will be addressed through planned trainings and organizational reform**.[17]
5. **Improved customer awareness will be appropriately targeted in order to generate customer social responsibility in reporting and reduction of leaks/illegal connections. Leaks and/or illegal connections can be successfully converted.**
6. **Households will recognize the value of and improvement in water quality[18] and reliability in a manner which translates to higher payment and more productive uses of available resources.** A community outreach/behavioral change component is in place as part of the THP, so this factor is closely tied to achievement of the behavior change objectives.
7. **Private-Public Partnerships (PPP) kiosks can be enacted and sustained.** This has proven challenging with delays in a resettlement plan for affected persons and throughout the procurement process.
8. **The time is sufficient for executing project inputs and realizing subsequent outputs.**

[15] The GVWC Institutional Strengthening Report Feedback notes a lack of attendance at steering committee meetings from Ministry of Health and Sanitation (MoHS) and Sierra Leone Water Company (SALWACO) at Water Sector Steering Committee meetings.

[16] Further feedback from the Mid-Term Evaluation Report revealed that while respondents thought the technical assistance and training provided to date were good and generally improved their ability to execute their responsibilities there was still need for additional coordination among consultants, further advanced training, and support in the form of equipment (hardware and software).

[17] GVWC has past experience developing and implementing organizational improvements including its 100-day Transformation Programme in 2012 which achieved significant, but unsustainable results. Lessons learned from the 100-day Transformation Program in 2012 cited low staff morale, lack of professional development and work issues as key operational concerns within the organization impeding sustainability.

[18] Throughout this document, water quality primarily refers to water availability/continuity at the system level. However, as part of the THP, GVWC will receive technical assistance from ASI to measure water quality as defined by World Health Organization (WHO) standards on measure of hygiene, color, odor and taste. When necessary this document differentiates the two as water continuity and water hygiene respectively.

SI will work to assess the strength of these assumptions through a review of available administrative data, project documentation, direct observation, relevant focus group discussions, key informant interviews and, where available, data collected through sensor technology and household interviews.

2.2.2 Electricity Sector Reform Project

This project was designed to address the policy and institutional constraints associated with inadequate, unreliable, and unaffordable access to electricity. It consists of two primary activities: the Sector Roadmap and Coordination Activity and the Institutional Strengthening Activity[19]. The links within each activity require the following assumptions:

1. **Similar to the WSRP, the electricity sector roadmap must be actionable with stakeholders clearly engaged through the steering committee.** While this is feasible, there is a question of political willingness in terms of government perceptions of responsibility regarding bill payments, allowing a state-owned utility to enforce collections, control theft, and the institution of an unpopular tariff increase.
2. **Stakeholder roles can be clearly defined and delineated, especially between EGTC and EDSA.** Discussions on a previous Power Purchase Agreement (PPA), for example, stalled on questions of asset ownership, responsibility, and payments. Both companies face financial constraints, making compromise difficult without external financial support.
3. **Independent Power Producers (IPPs) can be responsibly introduced into the market and generation and distribution properly managed.** The ToC provides a large gap regarding the necessary steps between an IPP tenured process and the financial viability of the sector. In addition, the ToC focuses on IPPs' generation of power and fails to account for potential challenges in transmission and distribution.
4. **The appropriate technical and institutional capacity constraints have been identified to produce the intended outcomes.**

Again, SI will work to assess the strength of these assumptions through a review of available administrative data, project documentation, direct observation, and key informant interviews (KII).

2.2.3 Regulatory Reform Project

The objective of the Regulatory Strengthening Project is to support the Government in its efforts to operationalize and build core capacities at the newly established independent regulator, EWRC, to ensure transparency, support the long-term financial sustainability of the water and electricity sectors, and improve overarching sector governance. It is closely linked with the above water and electricity sector projects and its ToC is incorporated within their models through the following activities: the Electricity and Water Regulatory Commission Institutional Strengthening Activity and the Performance-Based

[19] Sector Roadmap and Coordination Activity aimed at establishing a sector steering committee and roadmap to clarify the role of sector stakeholders, including financial contributions from the GoSL as well as IPPs, which would leave to improved coordination and an open tender for increased electricity generation. Institutional Strengthening Activity aimed at technical assistance, primarily to EGTC and EDSA, to improve coordination between the two entities, provide information for improved operations and decision making all with the goal of reducing the cost of service through increased financial and technical capacity of the sector.

Regulation Sub-Activity (also known as RBF)[20]. Key assumptions underlying these activities and their intention to improve sector financial sustainability and regulation include:

1. **Data will be available and actionable to both pilot a tariff setting exercise but also to instate tariffs either during or after the THP.** A number of concerns with the potential to undermine this assumption were raised during an evaluation Scoping Trip in late 2018. For instance, data requests sent to GVWC had gone largely unanswered,[21] and there was a lack of clear delineation or ownership of assets within the power sector requiring asset value to be excluded or largely estimated when setting tariffs. Additionally, current billings and collection models in both sectors introduce complications including: (1) within GVWC, the majority of water payments are currently on a fixed fee rather than metered approach, (2) the electricity Collection Account could require EWRC to adjust tariffs in the event of a sector deficit and (3) there exists potential in which Sierra Leone integrates into a regional power network with other members of the West African Power Pool to buy or sell electricity.
2. **EWRC's capacity can be sufficiently expanded through THP activities to include regulation of IPPs, an activity which is moving forward simultaneously under the ESRP.**
3. **Stakeholders will be able to provide proper documentation for the RBF, and the extrinsic motivations provided by the program can be internalized and sustained after the program's end.**

2.3 Literature Review

2.3.1 Summary of Existing Evidence

2.3.1.1 Water Sector Reform Project

The last two decades of water sector reform have identified two of the most prevalent challenges facing the water supply sector in developing countries as (1) increasing coverage, and (2) maintaining infrastructure. To address these issues, governments began instituting water sector reforms in the 1980s with the integration of private sector participation (PSP) emerging as a key strategy in a decade characterized by a wave of privatization and deregulation policies. Between 1990 and 2005, seven developing countries began to experiment with various types of PSP in their respective water sectors

[20] The Electricity and Water Regulatory Commission Institutional Strengthening Activity aims to provide technical assistance developing hiring, tariff setting and stakeholder engagement strategies to expand EWRC's capacity to set tariffs and regulate the water and electricity industries. The Performance-Based Regulation Sub-Activity (also known as RBF) aims to promote data and information sharing among key stakeholders while incentivizing regulated utilities to improve performance. Unlike the other activities mentioned in this section this one, while planned at project conception, is not incorporated into the programs' TOCs but we have included it here for its relative importance and potential impact, both in the short and long term.

[21] Problem-Driven Iterative Adaptation (PDIA) methodology suggests a gap between 'capability' in terms of what an agency can do, and its 'capacity' observed in terms of the resources the organization has at its disposal. From this perspective, PDIA directly entails how "premature load bearing" under externally imposed 'best practices' (i.e. EWRC's instructions to GVWC) in low capacity environments (GVWC) leads to a separation between what should happen and what does happen. Our discussion with both the EWRC and GVWC suggests that there is a fundamental mismatch between what the EWRC demands and what GVWC can provide, especially in relationship to data and information sharing which is essential in both the execution of RBF and proper tariff setting. This understanding is also corroborated by Adam Smith International's (ASI) Receptive Capacity Report.[21]

with strategies ranging from build-operate-transfer models to joint public/private ownership. Outside of PSP, alternate reform strategies are defined by Schwartz as components of the 'New Public Management' model and consist of increasing autonomy levels of the utility, separating regulatory tasks from service provision, creating quasi-competition, increasing tariffs, and increasing accountability of results[22]. The literature shows that water sector reform strategies were met with varying success across country contexts; in Colombia, Gomez-Lobo and Melendez demonstrated that PSP integration tends to have a neutral to positive effect on access, especially for the poor, and a neutral impact on affordability[23]. Government investments and the creation of new public institutions in Brazil increased coverage substantially between 1970 and 1990[24]. A multitiered private management strategy in Burkina Faso resulted in an 18 percent increase in the population share with access to safe drinking water between 1994 and 2003[25].

In terms of challenges facing the water sector in Sierra Leone, there is a large and growing body of literature which documents the infrastructural and policy challenges the country continues to face. The CA conducted in 2013 by the GoSL with support from MCC identified the lack of adequate, reliable, and affordable access to clean water as one of the binding constraints to broad-based private investments and economic growth in Sierra Leone.[26] Other documents corroborate the findings of the CA related to the lack of adequacy and reliability of water and electricity supply in Sierra Leone. These issues were attributed to aging and poorly maintained infrastructure, increasing demand especially in urban areas, and system shocks including the decade-long civil war and recent Ebola epidemic in which electricity and water sectors saw investment reduction as government money was shifted to provide emergency responses to the outbreak.[27]

GVWC is a key actor in the water reform space as the main water provider in greater Freetown. The THP has produced documentation that provides a literary background of the current status of GVWC and water supply and quality in Freetown. Specifically, the GVWC Institutional Strengthening activity under the WSRP has produced the Inception and Initial Findings Report[28] and the Customer Service Status

[22] Schwartz, Klaas. "The New Public Management: The Future for Reforms in the African Water Supply and Sanitation Sector?" Science Direct, 2008.

[23] Prasad, Naren. Social Policies and Water Sector Reform. United Nations Research Institute for Social Development, 2007, Social Policies and Water Sector Reform.

[24] Ibid

[25] Ibid

[26] The CA also identified road infrastructure as one of the top three challenges facing Sierra Leone based off a stakeholder consultation process (Annex 1 of the CA). SI was informed a stakeholder consultation on the THP then narrowed the focus to water and electricity and the primary geographic scope to the Freetown area, though the exact processes is unclear. Republic of Sierra Leone. (2014). Government of Sierra Leone. Sierra Leone Constraints Analysis Report: A Diagnostic Study of the Sierra Leone Economy; Identifying Binding Constraints to Private Investments and Broad-Based Growth. GoSL, 2013

[27] Howard, Marjorie. "The Other Cost of Ebola." TuftsNow. 20 January 2015.

[28] SMEC. International Inception and Initial Findings Report. SMEC and MCCU, 2017, International Inception and Initial Findings Report.

Report at GVWC[29]. Additionally, the 2016 Desk Audit Report (AquaRating evaluation of GVWC) investigated systemic issues, customer service, and management and commercial issues in GVWC[30].

On the operation and management side, SMEC International's Inception and Initial Findings Report,[31] identifies a number of operational issues in GVWC including: (i) high technical water losses due to poorly maintained and degraded infrastructure; (ii) high commercial losses and lack of metering and collections, non-economic tariff setting, and poor customer care at GVWC; (iii) intermittent supply and leakages which negatively affect water quality; (iv) customers making illegal connections or using unsafe water sources; (v) large arrears in payment by government entities; and (vii) poor record keeping which creates serious challenges in billing and collection.

The State of Customer Service Report for GVWC estimates water coverage in Freetown at 56 percent, with 22,000 registered customers in a city of some 1.5 million inhabitants.[32] Most of GVWC's customers have limited access with water shortages/interruptions. Among barriers to improving access and services, the report includes: (i) "Spaghetti" infrastructure which makes it difficult to pinpoint leakages or restrict water cut-offs to defaulters only; (ii) Limited number of pay points; (iii) Delays in resolving customer issues, leaks, and connections due to limited mobility of area teams with one vehicle assigned to each area office; (iv) A posted communication manager position to improve communications within the company that has remained vacant; (v) Lack of staff training with all 18 interviewed frontline staff (meter readers, bill distributers, pipe fitters) claiming to have never received on-job training or orientation despite several being with the company for over 20 years.

The AquaRating Desk Audit also assesses GVWC's performance in several areas of business, operations and management. The report noted that GVWC's regulator compliance in terms of providing supporting documents is alarming. While a successful utility is expected to both document adherence to and fulfillment of financial accountability, customer care, and environmental sustainability, GVWC noted adherence with only 39 percent of project indicators and could only provide adequate documentation for 2 percent of indicators[33].

2.3.1.2 Electricity Sector Reform Project

Historically, electricity sector reform began through the improvement of State-Owned Enterprises (SOEs), which functioned as vertically integrated monopolies that owned their supply chains and impeded free competition in the open electricity marketplace. Prior to 1982, state-owned utilities across the developing world were characterized by overemployment, under-collection, prices that were unable to cover the cost of production, high losses, poor quality power and, in many countries, power delivered to only a small portion of the economy. Many national governments, private donors, and international

[29] Adam Smith International. State of Customer Service, Stakeholder Engagement and Grievance Management in GVWC. ASI, 2018, State of Customer Service, Stakeholder Engagement and Grievance Management in GVWC.

[30] Diego, Fernandez, et al. AquaRating Audit Report of Guma Valley Water Company - Sierra Leone (GVWC). Kiwa, 2016, AquaRating Audit Report of Guma Valley Water Company - Sierra Leone (GVWC).

[31] SMEC. International Inception and Initial Findings Report. SMEC and MCCU, 2017, International Inception and Initial Findings Report.

[32] Adam Smith International. State of Customer Service, Stakeholder Engagement and Grievance Management in GVWC. ASI, 2018, State of Customer Service, Stakeholder Engagement and Grievance Management in GVWC.

[33] Diego, Fernandez, et al. AquaRating Audit Report of Guma Valley Water Company - Sierra Leone (GVWC). Kiwa, 2016, AquaRating Audit Report of Guma Valley Water Company - Sierra Leone (GVWC).

financial institutions continued to fund investments in flawed utility models because of their perceived macroeconomic benefits[34].

Sierra Leone is not exempt from these energy sector issues. The GoSL has identified several shortcomings that has justified the need for sectoral reform in the country. According to the CA, generation capacity was too insufficient to meet demand for electricity in the country. The Power Africa in Sierra Leone [35] documentation reports that only 11 percent of the urban population is able to access electricity. The GoSL also identified that transmission and distribution grid capacity is inadequate in Sierra Leone. The two government agencies responsible for electricity provision, the EGTC and EDSA, were created in 2011 to improve the situation.[36] However, the rule of contract and regular payment flow between public-owned generators (EGTC) and single-buyer distributors (EDSA) were not well-defined. Additionally, these public utilities (both EGTC and EDSA) were not bankable since public utilities did not have clear ownership of their assets and their corporatization process was incomplete. Among other binding constraints, the GoSL finds that public utilities are operating under provisional licenses without any social, technical, or financial obligations. The sector also suffered from high technical and non-technical losses with poor collections from post-paying consumers[37]. The CA[38] notes that the cumulative effect of the different operating inefficiencies in the sector led to a revenue shortfall of about USD 21 million in 2009; 45 percent of the shortfall was attributed to technical line losses, 12 percent due to uncollected electricity bills, and the remaining 43 percent was lost due to underpricing. Non-cost recovery of retail tariffs and tariff structure were inadequate because industrial and commercial consumers pay more per unit than the households.

To combat the consequences of these conditions, the GoSL has taken reform measures and plans to implement a series of reform activities over a 13 year period. The EGTC and EDSA were created in 2011 as a result of the unbundling of the NPA. The ESRP, as a part of the reform process, is designed to address challenges associated with insufficient, unaffordable, and unreliable access to electricity. The main objective associated with the creation of the ESRP was to help key institutions develop a roadmap and action plan while also providing these entities with technical assistance. In terms of reform strategies, the GoSL has identified the following as its reform priorities: (i) Develop electricity supply public service where the public and economy need it most; (ii) Embrace partnerships with the private sector to create a sustainable and inclusive electricity supply public service and (iii) Focus on the complex needs of Sierra Leone's population and business community at the national and local level.

Table 1 below summarizes the reform activities' implementation phases detailed in the Electricity Sector Roadmap.

[34] Jamash, T., Mota, R., Newbery, D., & Pollitt, M. (2004). Electricity sector reform in developing countries: A survey of empirical evidence on determinants and performance. Cambridge Working Papers in Economics, CWPE 0439. Department of Applied Economics, Cambridge.

[35] USAID. "POWER AFRICA IN SIERRA LEONE Fact Sheet." Archive - U.S. Agency for International Development, 2012-2017.usaid.gov/powerafrica/sierraleone.

[36] Recent reforms are often aimed at separating functions historically consolidated within one entity with the goal of making utility functions conform to normal corporate principles, potentially allowing for competition and the addition of private sector participation.

[37] Millennium Challenge Coordinating Unit, et al. Electricity Sector Reform Roadmap (2017-2030). MCC, 2017.

[38] Government of Sierra Leone. Sierra Leone Constraints Analysis Report: A Diagnostic Study of the Sierra Leone Economy; Identifying Binding Constraints to Private Investments and Broad-Based Growth. GoSL, 2013

Table 1: Implementation Phases of Electricity Sector Reform Activities in Sierra Leone

Phase	Anticipated Timeline	Key Objectives	Relevant Policy Framework/Mechanisms
Recovery Period	2018	Address urgent viability and operationalization issues and introduce the Collection Account.	Electricity Act, EWRC Act
Transition Period	2018-2020	Prepare the sector for financial stability through the construction of an intermediate commercial framework.	EDSA, EGTC
Delivery Period	2021-2025	Institutional and organizational structures will be set for the long-term objective of universal electricity access.	GoSL policy instruments, EWRC regulatory processes

Research by Bacon (2016)[39] which included 26 studies on reform suggests that while regulation or unbundling alone as a reform strategy does not have any statistically significant impact on key impact measures tested[40], a combination of unbundling with other reforms can produce desired results. Mota found that vertical unbundling was tied to a 147 percent increase in labor productivity (MWh/employee) between 1994 and 2000 in Brazil[41]. Similarly, Andhra Pradesh State Electricity Board (APSEB), India's government-owned vertically integrated power utility, suggests that unbundling the utility into generation, transmission, distribution, and supply companies in addition to other reform measures resulted in substantial reduction in distribution losses between 1999 and 2008[42]. Additional positive outcomes of unbundling were also observed by Pombo and Toborda, who conclude that privatization with unbundling and the introduction of a bid-based pool market in Colombia resulted in a substantial reduction in electricity interruption time[43].

Akin to examples above that emphasize the value of coupling unbundling with other reforms, Zhang et al found that the co-existence of privatization and independent regulation seems to be correlated with "greater electricity availability, more generation capacity and higher labor efficiency" as well. PSP

[39] Bacon, Robert 2018. Taking Stock of the Impact of Power Utility Reform in Developing Countries A Literature Review, Policy Research Working Paper No. 8460, World Bank, Washington DC.

[40] Plant load factor is an exception. Impacts tested include: Connections per worker, Energy sold per worker, Distribution losses, Electricity generation per worker, Plant load factor, Reserve margin, Collection rate, Number of employees. Electricity generation p.c. Generation capacity p.c., Private investment in sector $, CAPEX per worker, Access, Duration of interruptions, and Frequency of interruptions.

[41] Mota, R. 2003. 'The Restructuring and Privatisation of Electricity Distribution and Supply Businesses in Brazil: A Social Cost-Benefit Analysis." Cambridge Working Papers in Economics No. 0309.

[42] Bhatia, B. and M. Gulati. 2004. "Reforming the Power Sector: Controlling Electricity Theft and Improving Revenue." The World Bank, Washington, D.C.

[43] Pombo C. and R. Taborda. 2006. "Performance and Efficiency in Colombia's Power Distribution System: Effects of the 1994 Reform." *Energy Economics* 28(3): 339-369.

introduction garnered the benefit of private management. Private sector capital has been shown to lead to significant improvements in (a) increased connections per worker, (b) increased energy sold per worker, (c) reduced distribution losses, (d) increased collection rate, and (e) reduced number of employees[44]. Izaguirre shows that, between 1990 and 1997, 62 developing countries introduced private participation in the electricity sector to varying degrees, from management contracts for the state-owned utility in Mali to the privatization of most sector operations in Argentina, Bolivia, and Hungary[45]. The dominant type of private participation varies by region. Latin America/the Caribbean and Europe and Central Asia favor divestitures and Asia at large shows a preference for greenfield projects.

Analysis of the last four decades of electricity sector reform have yielded several important lessons that countries undertaking reforms should consider. Besant-Jones[46] argues that the most important takeaway from the reformation of power markets in developing countries is that “cookbook” solutions are ruled out by the extensive range of economic and institutional endowments of these countries; there is no ‘one size fits all’ remedy for dysfunctional utilities.

2.3.1.3 Regulatory Reform Project

Historically, water and electricity service provision, given access and scale requirements, form natural monopolies, providing opportunities for overcharging or poor service provision of essential services. To avoid this, many countries, particularly developed nations, have established independent regulators charged with establishing operation rules, providing monitoring oversight, and setting and enforcing allowable tariffs. However, in countries like Sierra Leone, in which utilities were traditionally owned by government, this process has been much slower with the key assumption that governments, with the public interest in mind, can strike the correct balance between cost recovery, affordable tariffs, and levels of service.[47] Unfortunately, striking this delicate balance has typically been a challenge, creating public institutions which are largely subsidized with poor service delivery and underlining a need for increased regulation. The OECD[48] defines regulation not only as tariff setting but also as providing standards and technical assistance for improving service delivery.

The unbundling and restructuring of the water and power sectors in Sierra Leone over the past decade presented a clear need for transparent and independent utility oversight. As a response to this need, the GoSL Electricity and Water Regulatory Act of 2011 established the EWRC. Initial EWRC activities began in 2014. While intended to monitor quality and compliance and implement regulatory frameworks, the EWRC faces a number of constraints, including the absence of secondary legislation or "regulations"

[44] Zhang, YF., Parker, D. & Kirkpatrick, C. J Regul Econ (2008) 33: 159. https://doi.org/10.1007/s11149-007-9039-7

[45] Izaguirre, Ada Karina. “Private Participation in the Electricity Sector—Recent Trends.” Worldbank.org, World Bank, 1998, siteresources.worldbank.org/EXTFINANCIALSECTOR/Resources/282884-1303327122200/154izagu.pdf.

[46] Besant-Jones, John E. 2006. Reforming power markets in developing countries: what have we learned? Energy and Mining Sector Board discussion paper; no. 19. Washington, DC: World Bank.

[47] Mumssen, Yogita; Saltiel, Gustavo; Kingdom, Bill. 2018. Aligning Institutions and Incentives for Sustainable Water Supply and Sanitation Services. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/29795 License: CC BY 3.0 IGO

[48] OECD (2015), The Governance of Water Regulators, OECD Studies on Water, OECD Publishing, Paris,https://doi.org/10.1787/9789264231092-en.

under the Electricity and Water Regulatory Act, a lack of funding and resources, and a lack of procedures and methodologies for clarifying its role to utilities, customers, and EWRC staff[49]. In order to address these challenges, the GoSL and Millennium Challenge Coordinating Unit (MCCU) collaborated to develop the Regulatory Strengthening and Tariff Development Support (RSTDS) Project which targeted both the water and power sectors. Key RSTDS components are summarized below:

- Sustainable revenue stream creation
- Field equipment fund establishment
- Data and information collection
- Independent decision-making promotion
- Tariff development
- PPA adoption and implementation

2.3.2 Gaps in Literature

The major literature gaps relevant to Sierra Leone are centered around the reform activity sequencing for very small low-income countries and the empirical investigation of the relationship among sequencing, transition between reform steps, and the effectiveness of multiple reform steps. The literature above found, in the electricity sector for example, that unbundling alone produced only one desirable result. When combined with PSP, there was insufficient analysis to draw conclusions. Similarly, studies investigated unbundling and competition but not with PSP. More effort needs to be extended in this area to help guide countries like Sierra Leone in water, electric, and regulatory reforms.

2.3.3 Policy Relevance of the Evaluation

The CA conducted in 2013 by the GoSL identifies a lack of adequate, reliable, and affordable access to electricity and clean water as 2 of the top 3 key binding constraints to economic growth in Sierra Leone and therefore a focus of government policy and action.

Access to electricity in Sierra Leone can be quantified two ways: the number of consumers with power access from the grid and the amount of power supplied. While only 20 percent of the population had access in 2016,[50] even those with access did not have 24 hour supply. The CA notes that by 2009, Sierra Leone recorded an average number of 46 days of power outages over a one-year period. This is more than four times higher than the average of 11 days for low-income, fragile countries. Based on this estimated number of power outage periods, the estimated cost of these outages to the economy was approximately USD 17 million or 0.7 percent of Sierra Leone's gross domestic product for 2009.

In the water sector, the CA estimates the average household spends 21 days per year collecting water. It notes that this burden is especially acute for women and girls who bear more responsibility than men

[49] AARC Consultancy. Regulatory Strengthening and Tariff Development Support Inception Report. AARC Consulting, 2017, Regulatory Strengthening and Tariff Development Support Inception Report.

[50] World Bank. "Access to Electricity (percent of Population)." World Bank, Sustainable Energy for All (SE4ALL) Database from the SE4ALL Global Tracking Framework Led Jointly by the World Bank, International Energy Agency, and the Energy Sector Management Assistance Program., 2016, data.worldbank.org/indicator/EG.ELC.ACCS.ZS.

and boys for fetching water.[51] This impacts the ability of these groups to attend school or engage in other workforce participation activities. In addition, water-related diseases were estimated to comprise 50 percent of the average household health expenditure. Coupled together, time and money spent on collecting clean water and time lost due to water-related disease have potential large-scale economic effects.

Despite political transition in 2018, there remains a clear commitment in the GoSL to addressing these sector deficiencies. SI/MCC's Scoping Trip in 2018 provided opportunities to meet with several key government stakeholders and reinforce the GoSL THP commitment through addressing areas of inadequate policy, lack of sector regulation,[52] and institutional ineffectiveness.

As outlined in the design below, the evaluation will compile and analyze a large amount of data and information obtained from various sources and disseminate the results. This is expected to make a significant contribution by (1) addressing the present lack and limited availability and accessibility of systematic and shared data within the water and electricity sectors and (2) informing future sector investments and interventions through assessments of the operational and financial viability of benefiting institutions and the potential impacts of sector specific programs (DMA/Kiosk Activity).

It is expected that results of the evaluation will be utilized by the GoSL, benefiting institutions, other sector members, and donors and potentially inform design of a future MCC Compact should Sierra Leone become eligible.

[51] It should be noted this analysis is at the national level while much of the WSRP focuses on GVWC which services the Freetown area only. Due to shorter distances between collection points, or alternatively, larger population demand, we would expect differences in urban collection time from this nationwide estimate.

[52] New investment in power and water requires, among other things, a financially viable off-taker. Regulation can further enhance and protect financial performance and is, itself, an important contributor to investment, THP assistance may help in these areas but it not clear that the off-taker will be financially viable by the end of the THP assistance.

3 EVALUATION DESIGN

3.1 Evaluation Questions

In its entirety, the evaluation aims to validate the program logic underlying each sector's interventions and assess the sustainability of proposed and achieved outcomes after project close. It does so by focusing on several key service providers and institutions, primarily GVWC, EGTC, EDSA, and EWRC. In addition, the evaluation will use interviews outside stakeholders and independently collected data at the household and system level (in the water sector) to supplement findings.

3.1.1 Electricity and Water Reform Projects and the Regulatory Strengthening Project

To ensure learning and accountability eight overlapping evaluation questions (EQs) in ESRP and WSRP are proposed.

1. Were the Activities/Sub-Activities implemented as designed? What were the challenges and successes in implementing the reform activities in the sectors?

2. To what extent did the Projects/Activities accomplish the desired outputs and outcomes outlined in the program logics? For the ESRP this should include but is not limited to improved financial relationship between EGTC and EDSA, improved efficiency at targeted substations and generation facilities, etc.; and for the WSRP: operation and maintenance conducted more regularly and effectively, billings and collections increased, and improved responsiveness to customers.[53]

3. To what extent has coordination and planning within the water and electricity sectors improved as a result of THP activities?

4. What evidence is there that EGTC, EDSA and GVWC are becoming financially viable as a result of the THP activities? To what extent do these entities operate on principles that allow for cost recovery? Has cost of service improved?

5. What are the impacts of the THP activities on business operations and strategic planning within EGTC, EDSA and GVWC? How well did EGTC, EDSA and GVWC institutionalize (into the organization's culture and behaviors) the transformations that took place under the THP?

6. How useful was the THP's "dry run" (tariff development process) for helping to establish the tariffs in the water and electricity sectors? What were the challenges and successes in this process? To what extent does the tariff structure adopted allow for cost recovery of the utilities in the sector?

[53] GVWC and EGTC are partnering with counterparts in Ghana to learn more about their roles and better management of their systems/institutions. MCCU is planning to complete a special study on the twinning exercise, likely with support from an MCC fellow. This will be an internal assessment of the program (deep dive) while SI will be responsible for looking at the twinning exercise in the broader context of capacity building/THP. Per our understanding from discussions with MCC, SI's research plan for the special study is likely to be developed/available in September. Given that the EDR is a living document, SI will consult this plan and incorporate this topic in its PE design and explore its effects on GVWC and EGTC's performance.

7. What lessons can be learned from the implementation of the RSP? What is the impact of the results-based financing process on improving EWRC capacity to effectively monitor alternative water services providers and utilities in the water and electricity sectors? What is the impact of the results-based financing process on the operational and financial efficiency of the utilities? To what extent are the results achieved under this Sub-activity sustained beyond the THP?

8. How sustainable are the outcomes of the projects and activities and why (including around sector coordination, utility financial, commercial and operational improvements, etc.)? How was sustainability planning done during implementation and what best practices and lessons learned can be drawn from that process?

To answer the EQs, SI proposes a mixed methods approach, including both qualitative and quantitative methods through a framework developed on principles of organizational capacity assessment (OCA). SI considered conducting OCAs, but given the small number of institutions and time points, we believe the most effective and efficient approach will be to infuse our data collection with principles of OCA, rather than developing and implementing a tailored OCA for each sector. The proposed methods include: 1) document review, 2) secondary data usage, 3) key informant interviews (KIIs), 4) focus group discussions (FGDs), 5) direct observation (DO), and 6) process mapping.

3.1.2 District Metering Area and Standpipe Demonstration (Kiosk) Activity

The DMA/Kiosk Activity under WSRP incorporates its own subset of evaluation questions aimed at achieving learning and accountability in a more targeted scope of work (SOW) around the DMA/Kiosk pilot.

1. Were the activities in the DMA/Kiosk effective at reducing NRW, and if so, which activities were the most effective?

2. How did the activities pilot impact service reliability, cost of water, water quality, water collection times, consumer citizenship attitudes and behaviors, and satisfaction with water service in the targeted DMA? What evidence is there that the technical assistance provided to the GVWC resulted in increased capacity to affect outcomes related to gender and social equity in the DMA?

3. What is the impact of the DMA/Kiosk Activity on standpipe management, the levels of water service provided by the developed/rehabilitated standpipes under the THP and post-THP (e.g. hours of water, water quality, etc.)? How does the private sector approach to standpipe management (MCC-sponsored kiosks or networks) compare to other standpipe management approaches (e.g., community managed kiosks or networks) practiced in the Freetown area in terms of service reliability, maintenance, cost to households and revenues for GVWC?

4. Are there systems and results in the DMA around NRW and the standpipe pilot that can be sustainably replicated to other GVWC service areas? What factors/indicators are critical to ensure successful replication?

In addition to some of the methodologies used in the broader WSRP, ESRP, and RSP, SI will employ additional measurements including household surveys and the use of sensor technology to capture outcomes on three levels of beneficiaries. At the **system or network level**, SI will explore changes in NRW within targeted DMAs. At the **public distribution point level**, SI will explore if kiosks management and service provision improve for supported kiosks. Finally, at the **household level** SI will aim to

characterize changes in outcomes of interest including water supply, water quality, and customer satisfaction and citizenship behaviors.

Table 2 provides an overview of the data collection methods that will be used to explore each of the twelve evaluation questions.

3.2 Evaluation Design Overview

The Sierra Leone THP external evaluation contains two primary components. The first views the THP in its entirety, incorporating information and data from WSRP, ESRP and RSP into a performance evaluation (EQs 1-8). The second component provides an in-depth investigation, through the use of interrupted time-series analysis (ITSA) based on system-level data and a comparative longitudinal study based on household survey data, of one component of the WSRP, the DMA/Kiosk Activity (EQs 9-12).

For the performance evaluation (PE), SI notes that reform and capacity building plans for the water and power sector share some similarities as well as reliance on a single regulator, the EWRC. However, the functioning and operations of these sectors are quite different. Thus, within this overall evaluation of the THP, SI will utilize a PE to evaluate the effects of the THP on each sector separately before synthesizing the findings for the THP as a whole. While many of the methods will be shared between the water and power evaluations, the intended respondents and samples will vary based on the sector. Thus, though each sector will be approached separately, the final report will bring the findings together across sectors to ensure that the overall findings regarding the THP will not be lost. The final PE report will also integrate key findings from the DMA/Kiosk evaluation.

The DMA/Kiosk evaluation was originally proposed as an impact evaluation (IE) while recognizing the potential challenges and limitations of conducting an IE at the household level. Included in considerations was the inability to randomize interventions, selective targeting of DMAs, and other factors. Compounding the challenges SI's entry into the evaluation came after a baseline was conducted by another party, and targeted and comparison DMAs were selected without an independent evaluator's input.

Starting in late 2018, SI worked with MCC to explore the options for retaining the IE methodology recommended at the proposal stage, propensity score matching (PSM). To ascertain the validity of the IE design, SI completed a baseline data collection assessment and memo in December 2018. As part of the assessment SI examined the following questions:

- Are the sampling approach and sample size appropriate to employ a PSM technique?
- Are indicators identified/selected and questions asked on those indicators in the baseline questionnaires sufficient to answer all DMA/Kiosk EQs?
- Is data quality of baseline surveys reliable? Can results be replicated?
- Is Cockle Bay, the comparison DMA identified by a third party, an appropriate comparison DMA?

The complete findings and recommendations can be found in the Baseline Memo attached in the Annex 7, but in brief the team identified that the sample size provided, particularly in the Knowledge, Attitude and Practice (KAP) survey, was not sufficient for a rigorous IE. SI recommend either the execution of a new baseline survey or a change in the evaluation design. In addition, discussions with MCC articulated an additional concern in whether the focus on household level beneficiaries in the IE design was sufficient to address the system and standpipe level measurements of interest, particularly the effect of the program on NRW.

Given these factors SI, along with MCC, agreed to propose an alternative evaluation design. The evaluation design expanded upon below in *Section 3.4 DMA and Standpipe Demonstration (Kiosk) Activity* Evaluation Approach includes two components:

- ITSA of system and standpipe level data through (1) ultrasonic flow meter sensor technology installed throughout the treatment DMAs and (2) GVWC billing and customer records. SI will use an interrupted time series design and segmented regression analysis to test whether the system performance in terms of reducing NRW, improving quality, reliability and continuity have improved before and after the intervention. The end of THP intervention period will be used as the cut-off point. GVWC data will be used to address impact of the intervention on billing and collection. SI proposes to use sensor data for NRW assessment instead of GVWC administrative data since SI believes that pre-intervention data from GVWC may not be compatible to the post intervention data collected from sensors installed as part of the program.
- Household Level Analysis: A longitudinal panel regression approach of the treatment and comparison DMAs will be used to measure change experienced by households/GVWC customers between interim and endline.

In both components, the evaluation would explore changes in the targeted DMAs over time, providing valuable insight as to the correlation between the program's execution, its outputs and community perceptions. However, unlike an IE, the observed outcomes cannot be definitively attributed to the program itself.

Both evaluation components, the PE and the mixed method approach of DMA/Kiosk evaluation will be comprised of an interim and endline evaluation. Due to the timing of the evaluation start it is no longer feasible to collect independent baseline data pre-program. However, the ET will rely on a combination of administrative data, external surveys and qualitative data collected at the interim period to provide contrast before and after the program. The data collection methodology for both the PE and DMA/Kiosk Evaluation has been summarized in Table 2 below.

Table 2: Evaluation Questions, Outcomes and Methods (Design Matrix)

		Data collection method							
		Doc Review	KII	FGD	Direct obs.	Secondary admin data	Process mapping	HH questionnaire	Sensor data
	Evaluation questions for Electricity and Water sector reform project and Regulatory strengthening Project								
1	Were the Activities/Sub-Activities implemented as designed? What were the challenges and successes in implementing the reform activities in the sectors?								
2	To what extent did the Projects/Activities accomplish the desired outputs and outcomes outlined in the program logics?								
	a) ESRP: improved financial relationship between EGTC and EDSA, improved efficiency at targeted substations and generation facilities, etc.; and								
	b) WSRP: operation and maintenance conducted more regularly and effectively, billings and collections increased, and improved responsiveness to customers.								
3	To what extent has coordination and planning within the water and electricity sectors improved as a result of THP activities?								
4	Water: What evidence is there that EGTC, EDSA and GVWC are becoming financially viable as a result of the THP activities? To what extent do these entities operate on principles that allow for cost recovery? Has cost of service improved?								
5	What are the impacts of the THP activities on business operations and strategic planning within EGTC, EDSA and GVWC?								
	a) How well did EGTC, EDSA and GVWC institutionalize (into the organization's culture and behaviors) the transformations that took place under the THP?								
6	How useful was the THP's "dry run" (tariff development process) for helping to establish the tariffs in the water and electricity sectors?								
	a) What were the challenges and successes in this process?								
	b) To what extent does the tariff structure adopted allow for cost recovery of the utilities in the sector?								
7	What lessons can be learned from the implementation of the RSP?								
	a) What is the impact of the results-based financing process on improving EWRC capacity to effectively monitor alternative water services providers and utilities in the water and electricity sectors?								
	b) What is the impact of the results-based financing process on the operational and financial efficiency of the utilities?								
	c) To what extent are the results achieved under this Sub-activity sustained beyond the THP?								
8	How sustainable are the outcomes of the projects and activities and why (including around sector coordination, utility financial, commercial and operational improvements, etc.)?								
	a) How was sustainability planning done during implementation and what best practices and lessons learned can be drawn from that process?								
	Evaluation questions for the DMA Kiosk Activity								
9	Were the activities in the DMA and Standpipe Demonstration Activity effective at reducing NRW, and if so, which activities were the most effective?								
10	How did the activities piloted impact service reliability, cost of water, water quality, water collection times, consumer citizenship attitudes and behaviors, and satisfaction with water service in the targeted DMA?								
	a) What evidence is there that the technical assistance provided to the GVWC resulted in increased capacity to affect outcomes related to gender and social equity in the DMA?								
11	What is the impact of the DMA Activity on standpipe management, the levels of water service provided by the developed/rehabilitated standpipes under the THP and post-THP (e.g. hours of water, water quality, etc.)?								
	a) How does the private sector approach to standpipe management (MCC-sponsored kiosks or networks) compare to other standpipe management approaches (e.g., community managed kiosks or networks) in terms of service reliability, maintenance, cost to households and revenues for GVWC?								
12	Are there systems and results in the DMA around NRW and the standpipe pilot that can be sustainably replicated to other GVWC service areas? What factors/indicators are critical to ensure successful replication?								

3.3 Performance Evaluation Approach

3.3.1 Methodology

To answer EQs 1-8, SI will employ a mixed methods approach, including: 1) document review, 2) secondary data, 3) KIIs, 4) FGDs, 5) DO, and 6) process mapping. The analysis will focus mainly on pre-post comparison before and after the THP completion to identify changes over time. This analysis will include ITSA (if possible) of secondary data available from GVWC, EGTC and EDSA on billing, collection etc. Document review, direct observation, KII and FGDs will supplement secondary data analysis. The ET will focus mainly on (but not limited to) analysis of billing and collection; generation and distribution, change in access and coverage over time.

Some of these data collection processes will be ongoing throughout the evaluation period in order to organize and manage information as it becomes available. Since the evaluation is commencing after the program's start, it is impossible to gather true baseline measures. Therefore SI, through the assistance of MCC, MCCU, project stakeholders and consultants have an ongoing documentation review as described below. However, the majority of evaluation activities will take place at two key data collection periods, interim and endline. The interim period will be conducted shortly before or near the end of the THP and include retrospective questions to assess change over time since another baseline data collection is not feasible. Interim data collection is planned for late 2020[54] and important for capturing immediate outcomes (EQs 3, 4, 5, 6, 7) as well as ensuring institutional knowledge of the program design and implementation is not lost (EQs 1, 2, 3, 6, 7). In order to better assess long term outcomes and sustainability, the endline evaluation will take place approximately two years after the program close (EQs 3, 4, 5, 8).

3.3.1.1 Document Review

This technical assistance provided in the THP will generate a wealth of information and data that will inform the evaluation. SI has already begun the process and will retain regular contact with the M&E and technical project representatives to obtain project documents and reports.[55] SI has requested access to survey results, guidance documents, technical assistance consultant outputs, roadmaps and actions plans, business plans and other outputs that can provide insight into project progress (EQ 1) and contextual understanding. Where available, SI will also seek access to reporting (and data) from the THP partners (EGTC, EDSA, and GVWC). SI will request new documentation on a quarterly or biannual basis, in line with several THP reporting requirements including consultant outputs, due diligence reports, M&E Reports and revisions. The evaluation will also include a review of the relevant sector literature to help ensure that the evaluation design, as well as its eventual findings, are well-situated in the broader context.[56]

[54] SI is proposing late-2020 to capture extended THP programs (RBF, twinning exercise). However, it may be reasonable to conduct interim PE activities before this date in order to capture earlier THP activities.

[55] At minimum, include THP-level documents (e.g., the THP agreement, investment memo, logic models, THP M&E Plan, ITTs, etc.) and project documents (e.g., technical assistance consultant outputs, road maps documents and action plans, implementing entity progress reports, self-evaluations, training materials).

[56] At minimum, this will include relevant Government of Sierra Leone documents (laws, policies/regulations); and relevant academic and gray literature (publications by the AFDB, World Bank, DFID, UNICEF, International Water Association [AquaRating], etc.).

3.3.1.2 Secondary Data

To determine the extent to which project activities were implemented as planned and to support evaluation of key outcomes and their sustainability (EQs 2, 4, 7 and 8), SI will collect, review, and analyze secondary data. MCCU's quarterly Indicator Tracking Table (ITT) provides a high-level overview of project progress including the completion of key program outputs and key outcome indicators of interest include: operating cost-recovery ratio, revenue from sales, transmission/distribution losses, and maintenance expenditure-asset value ratio.

While the aggregated figures presented in the ITTs provide a useful program benchmark additional detailed data will be requested directly from the THP service providers: EWRC, EGTC, EDSA and GVWC. These data would include accounting information on customer billing and collections, and operational costs including production inputs and transmission/distribution losses as available, to better evaluate changes in financial viability (EQ4). The quality and quantity of this records kept by these institutions was noted as a concern at the inception of the THP due to institutional limitations in what they can collect and report. The THP is expected through multiple avenues, including technical assistance and guidelines provided as part of the tariff and service quality regulations, to improve the availability of secondary data which SI hopes to access.

In addition to secondary data and documentation collected from the benefitting agencies, SI will seek additional data from the following sources, assuming data sharing agreements can be achieved:

- Should any IPPs be attempted/implemented during or after the program data on production, costs and revenue will be requested
- Available World Health Organization (WHO) / UNICEF (United Nations Children's Fund) Joint Monitoring Program (JMP) data can also be a source of secondary data for WASH indicators.
- Data generated by the RBF activity led by Instiglio will also be requested and used in the PE unless it creates duplications in data requested directly from the service providers themselves. As per Instiglio's current plan, they will collect information on (a) indicators to assess the extent to which targets are achieved/passed, (b) actions taken to achieve targeted results, (c) and identification of external factors that may have played a role. SI will follow the same structure in the endline that Instiglio will use for midline. To maintain comparability between midline and endline, it is important that Instiglio and the ET collaborate on the preparation of interview questions. Contact with other organizations/donors such as African Development Bank (AFDB), Department for
- International Development (DFID) the World Bank (WB), and the Freetown Urban WASH Consortium were initiated during the scoping trip; these sources may provide relevant secondary information/data from independent surveys or beneficiary institutions.
- If possible, SI will purchase and/or utilize AquaRating data collected in Sierra Leone in 2016. This data will be helpful in establishing baselines of the technical, financial and commercial aspects of the GVWC's operations and to continually track the country water sector context.

Collection of secondary data and documentation will be assisted by the Local Senior Analyst in water and the Survey Specialist and evaluated by SI's Senior Analysts in water and electricity to assess the technical, operation and financial sustainability of THP activities. The project's Program Manager and Junior Analyst will contribute to the review of secondary data and documentation as well as any overall project documentation, particularly around THP monitoring and evaluation.

3.3.1.3 Key Informant Interviews

To substantiate monitoring and secondary data, collect lessons learned, address the 'why' behind any organizational changes identified, and explore sustainability of THP outcomes, SI will conduct KIIs. SI will develop and use standardized interview protocols contextualized for each stakeholder group (see below). All KII protocols will be semi-structured so that interviews collect information that is comparable across respondent groups, but also allow SI and respondents the flexibility to explore certain topics in greater depth as necessary and appropriate. For the protocols' development, SI will reference OCA approaches to identify key institutional domains and cross-reference those with project planning documents to identify focus areas for the interviews.

For the water and electricity PEs, beneficiaries are at the institution/entity level. Key informants, therefore, will largely represent employees from the EWRC, EGTC, EDSA, GVWC, National Commission for Privatization, Ministry of Water Resources (MWR), Adam Smith International (ASI) and project staff with approximately 20 KIIs planned per sector (overlapping with RSP). Following OCA principles, the KIIs will target both senior and mid-level staff to provide a more complete, accurate picture of institutional status. Similarly, engineers and frontline staff will also be interviewed for operation and management related performance. The ET's in-country Local Senior Analyst in water and Survey Specialist, and Sector Senior Analysts will be responsible for KIIs with work delineated by sector and position. Table 3 provides an overview of Key Informants by Sector.

Table 3: Key Informants by Sector [57]

Water Sector	Electricity Sector	Key Themes
MCC/MCCU (COO, CEO, M&E Director, WSRP Director, Community, Environment and Resettlement Expert)	MCC/MCCU (COO, CEO, M&E Director, ESRP Director)	• Planned Theory of change, outcomes and sustainability approach • Key challenges faced and known deviations from program's inception • Procurement processes • Location/provision of key program outputs and documentation • Mapping of key program stakeholders
WASH implementing entity staff (ASI, SMEC, Instiglio)	Electricity/power implementing entity staff (Instiglio)	• Implementation status of THP activities, including (1) perceived capabilities/knowledge transfer of technical assistance, (2) experienced issues/challenges in implementation including deviations from the original activity design, (3) perceived ongoing risks, likelihood and severity of each, (4) perceived sustainability outlook • Review of any data collected or documentation produced as part of consultancy

[57] In the event that one position crosses sectors (i.e. Director General of EWRC) only one KII will be administered per round covering both sectors.

Water Sector	Electricity Sector	Key Themes
		• Relationship with MCCU and project beneficiaries
GoSL WASH counterparts (including Ministry of Water Resources [MWR], Ministry of Health and Sanitation [MOHS], etc.), Ministry of Finance (Permanent Secretary)	Ministry of Energy (Permanent Secretary, Technical Advisors), Ministry of Finance (Permanent Secretary)	• Regulatory framework, policy context of THP • Perspective on sector coordination and planning • Plans for THP sustainability or expansion including potential stability of regulatory/policy environment with administration change • Role of IPP, Alternative Water Services Providers (AWSP) and the Economic Community of West African States - Regional Electricity Market • Amenability to proposed tariffs • Feasibility of EWRC's operational and financial independence
EWRC (Director General, Head or Legal and Consumer Services, Head of Water Division, Regulatory Economist)	EWRC (Director General, Head or Legal and Consumer Services, Head of Electricity Division, Regulatory Economist)	• Role of data transparency in regulation and oversight • Perception of tariff-run exercise, feasibility of implementation • Role of IPP, Alternative Water Services Providers (AWSP) • Operational and financial independence of EWRC • Role of RBF in the threshold program, perceived outcomes, potential sustainability of the exercise and impact on overall THP success
GVWC staff (including Managing Director, employees from the Community Water Services Department [CWSD], zonal officers, etc.)	EGTC (Director General, Select Members from the Board of Directors, Directors of: Planning, Finance, Technical, Commercial and Legal departments) EDSA (Director General, Select Members from the Board of Directors, Directors of: Planning, Finance, Technical, and Commercial departments, Secretary and Legal Advisor)	• Perceptions of successes and challenges in the THP implementation including but not limited to technical assistance, trainings, roadmap and steering committees, RBF, tariff-run, DMA/Kiosk activity • Perception of operational/financial improvements • Sector coordination and regulatory environment • Perceived sustainability and next steps • Review of available data and program documentation

Water Sector	Electricity Sector	Key Themes
WASH Sector Steering Committee members	Electricity Sector Steering Committee	• Sector Coordination improvements and challenges • Role of sector roadmaps • Sustainability and/or next steps for the sector
Community Counter part organization /local stakeholders		• Demands, bottlenecks, • Purpose and area of focus, • Partners they work/ communicate with.
Donors (AfDB, IDB, IWS, EU, WB, DFID, etc.)	Donors (EU, WB, AfDB, ECOWAS, etc.)	• Sector Coordination • Secondary Data and reports

3.3.1.4 Direct Observation

Particularly at the interim, SI will use direct observation to help answer EQs 1 and 2. SI will observe a) the performance of technical assistance and capacity building provided to water and power utilities, b) sector specific steering committee meetings (should access be feasible), and c) sector-specific operations, where relevant. This will allow the ET to understand program activities and especially outputs as they occur in their natural setting. Notably Dos will serve as opportunities to cross-check findings from other data sources. For the power sector, potential direct observations will include observing generation and substation operations including EGTC Thermal Plants and Transmissions Yards and providing verification of improved O&M practices through maintenance logs, etc. For water, observations may take place at GVWC regarding customer management and billing and collections, and at water kiosks as part of the DMA/Kiosk evaluation below. To further understand how capacity building activities are conducted, SI intends to observe capacity building training sessions (e.g. for kiosk management). Direct observations provide an opportunity to review and verify available secondary data and database management. All direct observations will use standardized tools, such as observation guides. Contingent on necessary permissions, where observation by SI might bias people's actions, visits will be unannounced and observers will remain as unobtrusive as possible. SI acknowledges that it may not be possible to observe all intended processes, but specific permissions and schedules will be determined as fieldwork approaches.

Direct observations will be conducted by the ET's Local Senior Analyst in water and the Survey Specialist but may, at times, include the Sector Senior Analysts, Program Manager and/or Junior Analyst.

3.3.1.5 Process Mapping

To help address EQs 1-3, 5-8 and 12, SI will design a longitudinal, process mapping methodology. This method allows for the tracing of steps required to complete a process in an organization or utility; the people and resources required for processes; and a variety of issues including challenges, coordination, and oversight.

Process mapping differs from process evaluation in terms of both objective and scope. Unlike process evaluation, process mapping is an 'informative tool' that helps analyze the gaps in a process within an entire service delivery system by identifying the priority areas for attention. Broadly, a process map links

supply and demand of a particular process in a system and provides information in four areas: Process overview, process boundary, identification of supply and demand gaps and measures taken (or necessary) to bridge the gap.

1. Process Overview: Describes the current situation explaining why this process is important in the wider system. Quantifies the cost, service, compliance and / or quality issues that currently exist.

2. Process Boundaries: Based on what is described above under process overview, it will outline the process, its' customer(s), and the supplier(s) along with expected and actual service delivery and outputs demanded by the customer(s).

3: Nature of the Gap: Identifies what is believed to be the 'Gap origin' in terms of resources, personnel (capacity) and management (governance) and how long it has been present.

4: Gap Consequence: Identifies/informs what is being done to minimize or eliminate this issue.

Figure 2: Service Delivery System Map

Any service delivery system may involve a series of processes. For each selected process within service delivery systems in water, electricity and institutional reforms, the process mapping tool will explore the following. In the process overview stage, the ET will review the policies and procedures providing guidance about completion of the process. In step 2, the ET will identify and compile an outline of sequence of steps along with identification of time and resource requirements to complete the process. Step 3 will provide information about human and financial resourcing challenges, gaps in coordination of individuals and resources involved in the process, and governance and management challenges. The ET will be able to determine how gaps affect process efficiency depending on the relative importance of the portion of the process chain that experiences the gap. Step 4 will detail the actions taken to overcome the challenges, changes in these factors over time, and the potential role of the THP in the process.

Figure 3 below describes each steps of RBF process mapping. In the case of GVWC, for example, the process overview stage of RBF process mapping will include review of billing and collection data (by customer type) and documents, production and distribution targets and price setting, and cost and budget from the Strategic Performance Improvement Plan for improving collections. Based on pre-intervention values of results and barriers to results, the next step will define scope of the process through verification of input, output and setting of distribution targets and prices for the total payment and prices at minimum and maximum targets. During the third stage of gap estimation, the mapping will verify model assumptions behind cost and benefit pricing and targets for connections and installments such as saddled connections, sub-main installation, the extent to which targets are reached or lagged (gaps), drivers of change, and external factors and their likely impacts. In stage 4, actions taken to bridge the gap, the mapping will provide information on actions taken such as rationing regime revision and valve regulation, the actions of the incentivized actors, and whether these actions are consistent with the drivers of change.

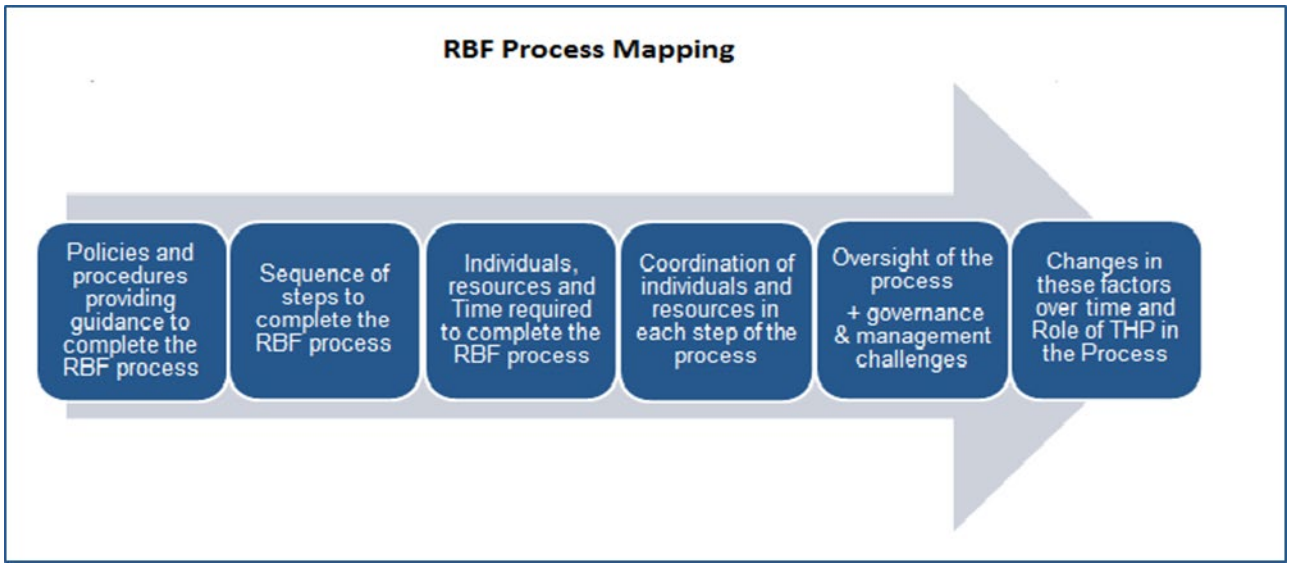


Figure 3: RBF Process Mapping

SI will use this method to trace and map processes expected to improve, including utility governance (efficiency, effectiveness, transparency, and accountability), technical and financial operations, and planning and coordination. Key institutions alongside their interactions with technical consultants, and end users will be targeted. Within the processes identified in the Project Logic[58] the evaluation team proposes focusing on those included in Table 4 below. These processes were chosen given their emphasis on improved financial and operational efficiencies. Financial and operational gains are defined in the THP logic as the intended, long-term impacts, of the program.

[58] Millennium Challenge Coordinating Unit. Monitoring and Evaluation Plan - Version 3. MCCU, 2018, Monitoring and Evaluation Plan - Version 3.

Table 4: Process Mapping Activities

Project	Process	Sub Activity to be mapped
ESRP	**Sector Framework and Coordination**	**Electricity Sector Collection Account, Billings and Payment and role of the Road Map**
ESRP	**Institutional Capacity Building, Focused on EGTC**	**Technical assistance and training in operations and management for EGTC and board**
WSRP	**Institutional Capacity Building, GVWC**	**Technical assistance and training on financial management, customer billings and collections**
WSRP	**Institutional Capacity Building, GVWC**	**Technical assistance and training in operations and management[59]**
RSP	**Regulatory Strengthening through Technical Assistance**	**Tariff dry run**
RSP	**Regulatory Strengthening through Technical Assistance**	**Results Based Financing**

The ET intends to use findings to produce visual explanations of the sub-activities listed in Table 4 alongside written analysis. The visual nature of the maps provides opportunity for the ET to efficiently identify process gaps. Furthermore, the maps are especially helpful for clearly communicating findings and takeaways to stakeholders during dissemination periods.

A detailed breakdown of tariff rate review process mapping by steps has been added in the Annex 1. Based on the structure discussed above, it outlines associated questions for illustrative purposes. This is significantly based on information gathered from discussions during the scoping trip and review of relevant documents and literature. We intend to finalize all RBF related processes through further discussion with RBF implementing and independent verification agency Instiglio and Oxford Policy Management (OPM).

These processes will be discussed and agreed upon with MCC and MCCU before data collection begins. SI will then convene teams of utility and regulatory personnel and consultants to participate in robust process-mapping sessions. During these sessions, SI will explore each process from start to finish, identifying efficiency and quality indicators, discuss bottlenecks and challenges to effectiveness, and explore changes over time and expectations for the future. Process mapping will be led by Sector Senior Analysts with oversight by SI's Program Manager.

[59] Documentation from the AquaRating evaluation of GVWC and results highlighted during the 100-day Transformation Program in 2012 as part of GVWC's Strategic Performance Improvement Plan provide a baseline of operational and management challenges, past successes and sustainability concerns.

3.3.1.6 Data Integration

Information from the process mapping, KIIs, FGDs, DOs, monitoring/utility data, and desk review will be triangulated to arrive at rigorous and nuanced findings regarding the changes that have occurred (or not occurred). Though the PEs will not be able to definitively attribute changes to the THP, SI will rely on contribution analysis to assess the chain of expected outcomes, identify alternative explanations, evaluate alternative explanations, and test underlying assumptions.

Outcome measures

The table below presents the outcome indicators and expected data sources and addresses the evaluation questions.

Table 5: PE Outcome Level Measurements

	Evaluation Question	Outcomes	Data Source	Data Type
1	Were the Activities/Sub-Activities implemented as designed?	1. Implementation Fidelity	**Desk review**: Original logic model; revised logic model; operational guidance; memos/documents related to procurement, changes in design; Technical trainings materials and reports, Full sector specific tariff case reviews, road maps, board presentations; strategic plans; GoSL planning documents. **KII/facilitated discussion** with MCCU staff, board members, contractors, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge. **Secondary information, Process Mapping** and **Direct Observation**.	Qualitative
1(a)	What were the challenges and successes in implementing the reform activities in the sectors?	1. Project risks identified, documented and mitigated. 2. Drivers of challenges and successes identified and documented including: • Acceptability • Penetration • Feasibility		

2	To what extent did the Projects/Activities accomplish the desired outputs and outcomes outlined in the program logics?	1. Project adheres/executes all outputs as defined in the Program Logics 2. Deviations in the Program Logics are well documented and justified 3. Intended outcomes are realized and successfully linked to project outputs	**Desk review**: Original logic model; revised logic model; operational guidance; memos/documents related to changes in design; EDSA financial audits, asset inventory and revaluation and power sector expansion plan, Transparent procurement process, GIS and customer mapping, and construction and operationalization of kiosks for DMA activity, and RBF indicators for payment, board presentations; strategic plans; GoSL planning documents. **KII/facilitated discussion** with MCCU staff, board members, contractors, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge. . **Secondary data** on billing and collection, consumer map**, Process Mapping** and **Direct Observation**	Qualitative and Quantitative
2(a)	ESRP: improved financial relationship between EGTC and EDSA, improved efficiency at targeted substations and generation facilities, etc.; and	1. Utilities experience improvement in • Billing and collection; • cost recovery; • efficiency gain		
2(b)	WSRP: operation and maintenance conducted more regularly and effectively, billings and collections increased, and improved responsiveness to customers.	1. Utilities experience improvement in • O&M; • billing and collection, and • customer responsiveness		

3	To what extent has coordination and planning within the water and electricity sectors improved as a result of THP activities?	1. Increased level of coordination among internal and external stakeholders achieved in water and electricity sector 2. Improvement in planning process achieved in water and electricity sector	**Desk review**: Project documents, including reports commissioned by THP; memos/documents related to coordination and planning; board meetings; strategic plans; GoSL planning documents. **KII/facilitated discussion** with MCCU staff, board members, contractors, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge. **Secondary information, Process Mapping** and **Direct Observation**.	Qualitative

4	What evidence is there that EGTC, EDSA and GVWC are becoming financially viable as a result of the THP activities? To what extent do these entities operate on principles that allow for cost recovery? Has cost of service improved?	1. Attributable evidence gathered and documented on financial viability of utilities based on- • Improved cost-recovery • Efficiency gain in Billings and Collections • Improvement in Profits of Standpipe Management • Integration of Alternative Water Service Providers • Reduction in debt held by EGTC, EDSA and GVWC • Reduction in government subsidy for program operations • Appropriate tariffs set and collected	**Desk review**: Project documents including reports commissioned by THP; board meetings; strategic plans; GoSL planning documents; financial reports; billings and collections data. **KII/facilitated discussion** with MCCU staff, board members, contractors, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge, alternative water service providers **Secondary data/information, Direct Observation**, **Process Mapping, FGD** on alternative water service consumers, independent energy suppliers	Quantitative and Qualitative

5	What are the impacts of the THP activities on business operations and strategic planning within EGTC, EDSA and GVWC?	1. Attributable evidence gathered and documented on business and strategic planning improvements based on- • Training to enhance capacity of board members at EGTC, EDSA, EWRC. • Improvement in operational efficiency of Substations and generation facilities improved • Reduction in transmission and distribution system losses • Improvement in EGTC and GVWC maintenance and its integration into operations	**Desk review**: Project documents including reports commissioned by THP; memos/documents related to coordination and planning; board meetings; strategic plans; GoSL planning documents. **KII/facilitated discussion** with MCCU staff, board members, contractors, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge **Secondary data/information, Direct Observation**, and **Process Mapping**	Quantitative and Qualitative
5(a)	How well did EGTC, EDSA and GVWC institutionalize (into the organization's culture and behaviors) the transformations that took place under the THP?	• Reduced corruption through Pay No Bribe Platform reduced • Improved Stakeholders engagements and advocacy campaign • Improved customer interactions and response times improved	**KII/facilitated discussion** with MCCU staff, board members, contractors, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge **Secondary data/information** and **Process Mapping**	

6	How useful was the THP's "dry run" (tariff development process) for helping to establish the tariffs in the water and electricity sectors?	1. Utilities gain knowledge, including the ability to identify challenges and benefits, from tariff process 2. Utilities improve knowledge and capacity to attain successful cost recovery through the tariff process 3. Improved knowledge and capacity for EDSA, EGTC and GVWC to comply with EWRC requirements 4. Protocol for utility reporting on KPI to EWRC is established and followed 5. Successful Implementation (by endline) of Tariff design 6. Improvement in billing and collection 7. Cost recovery is realized (by endline)	**Desk review**: Project documents including reports commissioned by THP on tariff development process; **KII/facilitated discussion** with MCCU staff, board members, contractors (Instiglio), MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge **Secondary data/information** and **Process Mapping**	Qualitative and Quantitative
6(a)	What were the challenges and successes in this process?			
6(b)	To what extent does the tariff structure adopted allow for cost recovery of the utilities in the sector?			
7	What lessons can be learned from the implementation of the RSP?	1. Utilities are able to articulate and synthesize the role of EWRC in their organizations 2. RBF funds are distributed and utilities can articulate the role of RBF in improving financial stability	**Desk review**: Project documents including reports commissioned by THP on RBF process; **KII/facilitated discussion** with MCCU staff, board members, contractors (Instiglio), MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge	Qualitative and Quantitative
7(a)	What is the impact of the RBF process on improving EWRC capacity to effectively monitor alternative water services providers and utilities in the water and electricity sectors?	3. Core policies and procedures for EWRC operations are developed and documented 4. Stakeholder engagement and community outreach is documented and feedback incorporated into EWRC operations 5. Capacity building successfully completed for reporting on RBF indicators & disbursement of funds		

7(b)	What is the impact of the RBF process on the operational and financial efficiency of the utilities?	6. Method of registration and oversight established for the financial and operational efficiency of alternative service providers in water and electricity sector.	**Secondary data/information** and **Process Mapping**	
7(c)	What is the impact of the RBF process on the operational and financial efficiency of the utilities?			
8	How sustainable are the outcomes of the projects and activities and why?	Knowledge and capacity gain by utilities to 1. To successfully identify barriers or opportunities 2. To provide documentation on how THP outcomes were sustained (endline) 3. Maintain Sector coordination through the steering committee or comparable meeting post-THP 4. Increase stability over time achieved and maintained for utility finance after THP 5. Achieve and maintain commercial and operational efficiency gain after THP	**Desk review**: Project documents; **KII/facilitated discussion** with MCCU staff, board members, contractors, Kiosk operators, MCC staff, and other relevant stakeholders such as EWRC, EDSA, EGTC and GVWC with recent and historical knowledge **Secondary data/information** and **Process Mapping**	Quantitative and Qualitative

3.3.2 Timeframe of Exposure

As noted in 3.3.1 above, SI will conduct interim data collection in late-2020 with endline data collection to occur two years after the close of the program. The timeline assumes an extension of the THP to March 31, 2021 and the timing of interim data collection attempts to balance ensuring the appropriate program actors are accessible and that institutional knowledge from project conception is retained and captured with the recognition that some project components (particularly the DMA/Kiosk Activity and RBF) may still be underway. Furthermore, holding interim data collection in 2020 ensures that baseline (2018), interim, and endline (2022) collection efforts are evenly spaced. At interim data collection, we expect for outputs and near-term outcomes to be completed including, for electricity, improved sector coordination, an open tender for electricity generation on the market, an agreement and improvement on EGTC's and EDSA's financial relationship, operational improvements and realized efficiencies at substation and generation facilities. For water we expect to again see improved sector coordination,

including recognition of GVWC capacity constraints and streamlining of reporting requirement/interactions with donors, as well as improved data availability on GVWC operations and customer base, improved billings and connections and increased transparency and accountability to their customer base through the Pay No Bride platform. At this time, the PE key evaluation points are planned to run concurrently with that of the DMA/Kiosk Activity (see section 3.5.2 below). However, significant delays on this (or other components of the program) may require a staggered approach between the PE and DMA/Kiosk Evaluation.

To adequately assess the sustainability and trends in changes over time, endline data collection will take place approximately two years after the THP close or in late-2022. SI is confident that two years post-interim is the appropriate time frame for endline data collection because it will allow for a sufficient, yet not superfluous, amount of time to pass for evaluation efforts to draw conclusions related to program sustainability. If Sierra Leone is granted an MCC Compact targeted at the electricity and water sectors, the endline evaluation may be moved up in order to (1) generate lessons learned and provide data useful in setting Compact goals and (2) avoid contamination from Compact activities.

3.3.3 Study Sample

3.3.3.1 Sample unit

SI's qualitative approach includes a variety of sample units with each type of stakeholder being interviewed using a distinct protocol to be developed during the data collection instrument design phase. For KIIs, the sample unit is individuals selected to represent the beneficiary institutions, government entities and grant implementers/managers as outlined in Table 6 below. Direct observation will occur primarily at points of generation, transmission and distribution of services and will include structured observational protocols focused on operational processes. Direct observations are likely to be supported with KIIs from the respective institutions. Sampling units for direct observations will also include sector steering committee meetings in which the content of the meeting and quality of stakeholder interactions is assessed. Finally, process mapping will include a group of key individuals involved in the design and execution of the identified process.

3.3.3.2 Sample size

The sample size is contingent on the method of data collection. In general, SI is proposing 15-20 KIIs per sector, 5-10 KIIs for the RSP, 3-5 for overall programming. Direct observation is expected to take place at 2-3 sites per sector (excluding the DMA/Kiosk activity) and at least 1 steering committee meeting per sector. Two process mapping exercises are expected for each project under the THP.

Table 6: PE Respondents

Method	Sector/Program	Target	Sample Size
KII	Water	Individuals from GVWC, MWR, MOHS, ASI, SMEC, Sector Steering Committee, Sector Donors	15-20 individual

Method	Sector/Program	Target	Sample Size
KII	Electricity	Individuals from Ministry of Energy, EGTC, EDSA, Sector Steering Committee, Sector Donors	15-20 individuals
KII	Regulatory	Individuals from EWRC, Instiglio	5-10 individuals
KII	Programs	MCC, MCCU, Ministry of Finance	3-5 individuals
Direct Observation	Water	GVWC Headquarters Kiosk Piloted Sites	2-3 sites
Direct Observation	Electricity	EGTC Thermal Plants EGTC Transmission Yards	2-3 sites
Direct Observation	Electricity/Water	Steering Committee	2 (1 per sector)
Process Mapping	Water	Process #1: Billings and Collections	1 group (5-10 key stakeholders)
Process Mapping	Water	Process #2: Operational Management	1 group (5-10 key stakeholders)
Process Mapping	Electricity	Process #1: Billings and Collection (Collection Account and sector coordination)	1 group (5-10 key stakeholders)
Process Mapping	Electricity	Process #2: Operational Management	1 group (5-10 key stakeholders)
Process Mapping	Regulatory	Process #1: Tariff Setting Exercise	1 group (5-10 key stakeholders)
Process Mapping	Regulatory	Process #2: Results Based Financing	1 group (5-10 key stakeholders)

3.3.3.3 Sample frame

The sample frame is composed of beneficiary entities, project stakeholders, implementing agencies and sector donors and has been developed based on a review of project documentation and consultation with MCC/MCCU.

3.3.3.4 Sampling strategy

Most key informants are selected using a **purposive** sampling technique. In some cases, there may only be one person or a few specific people who are performing the role whose perspective we require as a key informant. Prior to each round of data collection, SI will review project documentation and work with MCC and MCCU (at interim) to identify key informants. In the event that an identified informant indicates a colleague who could provide additionally illuminating information, we will attempt to contact this colleague to serve as an additional informant (**snowball** sampling).

Direct observation and process mapping with be conducted in a similar **purposive** technique with the intention of identifying observational sites and processes that can divulge the most information while representing a diverse range of project activities.

3.3.4 Primary Data Collection

3.3.4.1 Instruments

All KII protocols, direct observation reporting templates and process mapping exercises will be developed during the data collection instrument design phase drawing from SI's experience in qualitative evaluation design. All instruments will be pre-developed and when possible, field tested, prior to data collection. Care will be taken to ensure similar or overlapping themes and questions among KIIs, direct observations, process mapping and available secondary data to enable greater data triangulation. It is expected that instruments will be revised between interim and endline data collection to consider information learned from interim data collection, as part of the DMA/Kiosk evaluation and the intervening time period post-THP.

3.3.4.2 Rounds and timing

Interim data collection is anticipated to occur in late-2020 with endline data collection occurring approximately two years post-THP closure as noted in section 3.3.2 above. KIIs and direct observations of sites and steering committees (if still active) will occur at both rounds of data collection, as will process mapping activities. Finally, documentation review and secondary data collection is expected to occur through the entire evaluation period though the amount of new documentation is likely to diminish after the program close.

3.3.4.3 Respondents within the sample unit

Respondents are described in Table 6 above. SI will take care to book KIIs in advance to ensure the availability of respondents, particularly government officials. If a respondent is not available at the time of interview the team will follow pre-defined protocols to determine if the respondent can be replaced with another individual (typically from the same position) or if the meeting should be rescheduled. Direct observations are expected to be pre-scheduled as well (and in the case of steering committee meetings will follow the committee's schedule) unless SI determines that prior notice may have an external or unintended effect on the validity of the operation, in which case unannounced visits will be conducted. However, prior permission to ensure access for these visits will always be obtained in advance. Process mapping exercises will be planned in advance and require the coordination of several key stakeholders' presence. Given the potential difficulty in coordinating these meetings, should a key individual not be available, SI will request that a representative be sent in their place.

3.3.4.4 Staff

The Program Manager (Dr. Basab Dasgupta) will oversee the overall evaluation including the Sector PEs and DMA/Kiosk Evaluation. He will provide overall Quality Assurance (QA) according to MCC and SI standards. SI is composed of two overlapping teams to fully cover activities. The performance evaluation team, led by the PM, will be supported by the Senior Analyst/Electricity Sector Expert (Mr. Matthew Addison) for the electricity sector and the Senior Analyst/Water Sector Expert (Dr. Charles Pendley) and the Water Engineer (Mr. Joel Kamanda) for the water sector. The Data Collection Specialist (Mr. Peter Ghombo), will oversee data collection and quality on both teams and a Qualitative Specialist (TBD)[60] will support both sectors in content and data analysis.

The Local Senior Analysts in (Mr. Joel Kamanda), in addition to the Data Collection Specialist will play a large role in KIIs, documentation and secondary data collection and direct observations supported by Senior Analysts and the PM. In addition, the Senior Analysts for the respective sectors will take part in leading the Process Mapping Activities.

3.3.4.5 Data processing

Interview and discussion notes from qualitative data collection activities will be created during field work with daily review by the team to ensure clarity and quality. The team will also record all interviews and discussions, with respondent consent, to provide an opportunity for later reference. Qualitative data will be handled solely by the evaluation team and Social Impact Headquarters (SI-HQ) management team members.

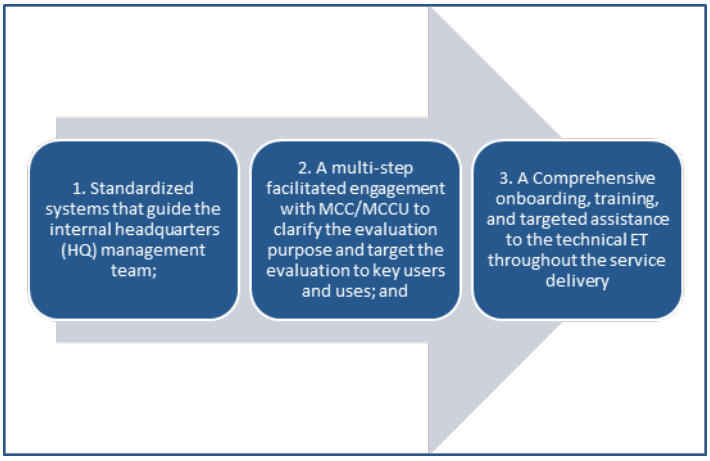
Secondary data will be reviewed on an incoming basis and digitized if needed. It will be assessed for consistency over time in order to ensure consistent naming of variables, units and measurements. Any changes in the data collection process or management system as noted by the sending agency will be documented.

3.3.4.6 Data quality

Adherence to the data processing protocol outlined in section 3.3.4.5 will lead to the collection of high-quality data. In addition, SI will employ an internally developed robust QA system called Evaluation Quality, Use, and Impact (EQUI)® that guides evaluation development, management, and implementation.

[60] Given the broad focus of the evaluation on Institutional Reform, SI is recommending changing the position originally titled “Sr. Analyst – Energy Specialist (Local)” to “Qualitative Specialist.” This individual will have experience working on past evaluations in Institutional Reform and assists both water and electricity evaluations on qualitative data analysis.

Figure 4: SI's EQUI® System for High-Quality Data



As described in Figure 4, EQUI's three-tiered approach to quality control includes detailed guidance notes, checklists, and templates for the management and evaluation teams to ensure consistency with SI's nest practices. As a part of this system, each evaluation product – from the evaluation design report to the final evaluation output undergoes a rigorous quality review, using checklists to guide development of usable, high-quality products that consider gender and social inclusion as key evaluation components. An online dashboard helps SI track project deadlines and employ tools for instituting financial management, achieving utilization checkpoints, and ensuring product quality.

3.3.5 Analysis Plan

Findings from the process mapping activities, KIIs, monitoring/utility data, and desk review will be triangulated to arrive at rigorous and nuanced findings regarding the changes that have occurred (or not occurred). Methodological triangulation also enables the team to strengthen the potential linkages and accuracy of its data if the results obtained through one method are less conclusive than another method. Though the PEs will not be able to definitively attribute changes to the THP, SI will rely on contribution analysis to assess the chain of expected outcomes, identify alternative explanations, evaluate alternative explanations, and test underlying assumptions. Evaluation findings from both the water and power sector PEs will be documented in a single report at interim and endline and will be submitted for review both by local stakeholders as well as by MCC's Evaluation Management Committee (EMC).

3.4 DMA and Standpipe Demonstration (Kiosk) Activity Evaluation Approach

3.4.1 Methodology

The ET proposes two methods for the DMA and standpipe activities: Interrupted time series analysis of data collected from the administrative sources and sensors and comparative analysis using panel regression model on household survey data. The MCCU-SL's DMA/Kiosk Activity identified two DMAs in which the MCCU can test the degree to which activities from the GVWC Institutional Strengthening Activity combined with the piloting of a PPP model for the operation and management of public standpipes can improve service provision to households, customer satisfaction, reduction of NRW, and revenue collection for the GVWC within each of the DMAs.

As described in Section 3.1.2 above, the objectives of the evaluation of the DMA/Kiosk Activity will focus on assessment of (I) reduction in NRW (EQ9), (ii) service reliability, cost of water, water quality, water collection times, consumer citizenship attitudes and behaviors, and satisfaction with water service in the targeted DMA (EQ10) (iii) the levels of water service provided by the developed/rehabilitated standpipes under the THP and post-THP (e.g. hours of water, water quality, etc.) and effectiveness of private sector approach to standpipe management (MCC-sponsored kiosks or networks) compared to other standpipe management/kiosk approaches (EQ11). Based on these assessments, the evaluation will inform whether systems and results in the DMA around NRW and the standpipe pilot can be sustainably replicated to other GVWC service areas (EQ12).

The evaluation will be comprised of two primary activities, an ITSA of administrative and sensor data primarily focused at capturing system level outcomes, and a comparative analysis inclusive of household level outcomes. These activities will be supplemented by FGDs, KIIs, direct observations and process mapping exercises. These supplementary activities will, at times, complement those taking place under the larger Performance Evaluation above. In addition, SI will work with the University of Colorado (UC) Boulder to ensure knowledge sharing and synchronization in data collection in household surveys, KIIs and FGDs when appropriate.[61]

SI recognizes that GVWC's recent rationing schedule change has and will have major effects on water supply in DMA areas. The ET will be mindful of this and continue to collect relevant information throughout data collection.

Prior to elaborating on the proposed methodology, the following two subsections describe the DMA Implementation Design in greater detail. This background information contains anticipated sources of DMA data and is crucial for understanding the proposed data collection methodology and analysis.

3.4.1.1 DMA Implementation Design: Metering Activity

MCCU's Technical Specifications Bidding Document suggests that three types of meters/sensors will be installed as part of Pipe Rehabilitation at Aberdeen and Kingtom DMAs[62]. Data from these sensors will allow comparison of the quantity of water supplied against that consumed and paid for by the community, providing the opportunity to measure, and for GVWC to react to, instances of NRW. The evaluation team understand the following will be installed:

- ***DMA Bulk Meters*** (magnetic flow meters) will be installed at four touchpoints within the targeted DMAs - (i) off Sir Samuel Lewis Road (Aberdeen), (ii) off Family Kingdom roundabout (Aberdeen), (iii) off Kallon Drive-3 (Aberdeen), and (iv) near Kingtom bridge roundabout (Kingtom). These meters will be capable of recording water flow in both directions. These will be "smart meters" providing data on water trends and measurements (on an hourly and daily basis) and capable of remote transmission to an

[61] The MCC Sierra Leone Evaluation Management Committee (EMC) did not approve all activities recommended below. Please see *Annex 8: Approved Evaluation Activities* for clarification as to which data collection activities will take place.

[62] Millennium Challenge Coordinating Unit. Bidding Document: Pipe Rehabilitation and Kiosk Installation at Aberdeen and Kingtom DMAs. MCCU

external management software capable of logging data, leak detection and error self-detection without requiring an on-site visit.

- ***Pressure loggers*** will be installed up stream of the bulk meters to measure water pressure at the meter installations. The unit will have the features of electronic type data logger configured to use 3G/GPRS (General Packet Radio Service) or SMS remote communications for periodic reporting.
- ***Domestic Meters*** (ultrasonic flow meters) will be installed at (1) all current GVWC household hook-ups with the aim of 100 percent penetration into the current customer database, (2) manifold positions on sub main pipelines within the DMA, and (3) guard boxes on constructed kiosks. Similar to the bulk meters, these domestic meters are expected to be capable of recording water flow in both directions. These will be "smart meters" providing data on water trends and measurements (on an hourly and daily basis) and capable of remote transmission to an external management software capable of logging data, leak detection and error self-detection without requiring an on-site visit. All household hook-ups are to be installed in conjunction with a representative from GVWC to ensure that the meters are properly reported in GVWC's database.

The Bidding Document notes that both bulk and domestic meters are intended for the measurement of the flow of potable water to consumers[63]. SI notes, however, that potable water, also known as water free of contamination, can only be accurately assessed at the distribution point. Rather, the installation of pressure loggers and flow meters will help measure the flow of water to the DMAs, at distribution points (i.e. kiosks) and at the households potentially allowing system level tracking of water flow and loss. The evaluation team has requested DMA mapping information to confirm the measurement points along the system.

If all bulk distribution and consumer hookup locations include meters, as suggested in the Technical Specifications, then it will be possible to use that information to conduct an NRW analysis using the interrupted time series approach. The total amount of water delivered to customers will be subtracted from the total bulk water entering the DMAs. This will indicate losses due to leakages or illegal connections. Further, if GVWC provides billing records, the fraction of water delivered to households that is paid for can also be calculated. NRW includes both unbilled / paid water, as well as water lost to leaks and illegal connections. However, one of the major limitations of using this information for NRW analysis is that in many cases GVWC uses a flat pricing (fixed tariff) system which may not provide actual loss or collection. The other limitation is that it may not be as accurate as the information collected from new installed meters for households. As a result, it may lead to less accurate measures of impact.

The pressure loggers will enable an estimate of continuity of service within DMA at the distribution points. Further, loss of water pressure will indicate higher risk of water contamination. The ET has asked if the domestic flowmeters include pressure logging capabilities, which would further allow a measure of continuity of service and potentially quality at the household level.

DMA Implementation: Water Kiosk Activity

The proposed installation guideline suggests that altogether 11 water kiosks will be constructed at sites in Aberdeen, Crabtown and Kingtom. There will be three types of kiosks, namely 2F (two faucets), 3F (three faucets) and 4F (four faucets) kiosks. Two domestic ultrasonic flowmeters will be installed within

[63] ibid

Guard boxes to measure inflow and outflow in each kiosk. The meter will be used to measure flow of potable water for domestic use in outdoor conditions. Each kiosk will have single cylinder filter installed inside kiosk to improve the quality of water exiting the kiosk. Water testing kits and reagents for water hygiene monitoring are to be supplied at Mile 13 Treatment Plant, though at the time of this document it is unclear where water testing kits will be utilized, by whom and at what frequency. At present, SI plans to use data from the ASI and GVWC water quality tests. ASI's proposed Water Quality Monitoring Plan includes the collection of 104 random samples, from hydrants, water kiosks, private yard taps, public stand taps and institutional connections, over the course of one year (one sample collected per week per treatment DMA). The data collected by ASI will create a panel dataset consisting of 52 X 2 observation. Samples are to be evaluated on taste, odor color, turbidity, iron, manganese, hardness, PH, residual chlorine, and fecal coliform against WHO guidelines by GVWC staff. Depending on the comprehensiveness of the resulting data, SI may conduct additional water hygiene tests. It is especially important to test different types of water points to capture variation in water quality within the DMA. Water hygiene variation arises due to differences in water point type, distance from the source, and slope. The name location and type of kiosks are given in Table 7.

Table 7: Distribution of Kiosks

Kiosk Number	Location	Kiosk Type
Kiosk 7	Boiling Street, Kingtom	2 Faucet
Kiosk 2, 4	White House Road, (Guard Room), Kingtom	4 Faucet
Kiosk 6	Battery Street, Kingtom	2 Faucet
Kiosk 11 and 11A	May Street, Kingtom	2 Faucet
Kiosk 8	Kallon Drive, Kingtom	4 Faucet
Kiosk 4, 5	Kincard Street (Market), Aberdeen	2 and 4 Faucet respectively
Kiosk 9	Nylander Street, Aberdeen	2 Faucet
Kiosk 15	Off Beach Road, Crabtown	3 Faucet

Source: Pipe Rehabilitation and Kiosk Installation at Aberdeen and Kingtom DMAs, Vol.2 Procurement Document

3.4.1.2 Interrupted Time Series Analysis of Administrative and Sensor Data

SI will use interrupted Time Series Analysis and proposes the use of sensor-level time series data for quality, reliability and NRW assessment before and after the completion of THP. ITSA is a robust quasi-experimental design with the ability to infer the effectiveness of an intervention that accounts for data dependency. It offers a rigorous methodology to determine the effectiveness of complex interventions on outcomes in real world settings. When RCTs or quasi-experimental designs such as matching method is not feasible or not applicable, ITSA is considered the strongest research design in the policy evaluation literature. For before-after time series data with potential to be affected by unobserved factors, ITSA

design has emerged as a quasi-experimental methodology with the strongest power to infer causality without stripping contextual and temporal factors from the analysis.[64]

For analysis, the completion of the THP intervention will be used as cut-off point for before and after ITSA to address evaluation questions around sustainability. Since high frequency data on system level outcomes of interest including NRW[65] and service reliability[66] (see Table 8 below) are expected to be available during and following the intervention from the DMA metering activity, SI will conduct ITSA to determine if discrete shifts in trends suggest a programmatic effect of the DMA/Kiosk Activity. This cutoff is appropriate because unlike other potential cutoff points, the before and after data will be most comparable as both datasets rely on sensors installed during the program itself. Furthermore, the cutoff considers the possibility of water kiosk failures that may occur within the first year after installation.[67] ITSA can identify break points statistically derived from the system level time series data. So, even if all or most outcome changes occur in the period between sensor installation and end of THP, ITSA can find a natural break from the system level data and allow the ET to compare various scenarios.

While changes over time will be informative post-intervention (post THP), we can strengthen our ITSA approach by including comparable "control" series (Linden 2015; Linden and Adams 2011; Simonton 1977). In this case, outcomes measured from two sources—treatment and comparison DMAs—can be used for the same time periods. This comparative ITSA approach can assess the post-intervention total changes by correctly capturing level and slope changes of treatment series due to intervention.

In the case of our DMA evaluation, there are two key opportunities for identifying a comparison group:

First, SI can investigate NRW, the amount of unbilled and unpaid water in the comparison DMAs along with treatment DMAs, [68] through GVWC records before and after intervention. To verify this as a feasible option SI will need to first review GVWC's billing and collection records to investigate their level of completeness and accuracy, including disaggregation at the DMA level.[69] There are currently no measurement systems in place, that SI is aware of, to capture NRW due to leaks and illegal connections prior to the meter upgrades planed for treatment DMAs. Moreover, the GVWC data on NRW may not be compatible with the sensor level system data on NRW planned for the treatment DMAs. However, in addition to GVWC billing records SI will also request access to any documentation on system leaks and repairs as well as include questions in the household survey on leaks, repairs and illegal connections to provide a qualitative narrative on NRW in the comparison area.

[64] Penfold et al, 2013; Shadish et al., 2002; Taljaard et al. (2014); Kontopantelis et al (2015); EPOC (2015)).

[65] NRW includes leaks, illegal connections, unbilled and unpaid water. Sensor data are expect to capture leaks and illegal connections (though it may not be able to differentiate between the two and would require investigation by the GVWC). Unbilled and unpaid water is best captured through GVWC administrative financial records.

[66] Measurements of service reliability, for example number of hours per day where water is available, are dependent on pressure loggers throughout the system. At this time SI is only aware that pressure loggers will be installed at DMA distribution points and has requested information as to whether or not domestic flowmeters will include pressure logging capabilities.

[67] Andres et al., 2018a; Andres et al., 2018b

[68] Including measurements at households and public standpipes.

[69] The scoping trip noted some concerns with GVWC ability to accurately maintain collections and billing records.

Second, an additional option can be through inclusion of meters in a comparison area. SI, with permission of GVWC and appropriate funding, could install sensor technology in the comparison DMA as an *optional activity*. This would allow SI to compare trends in the treatment and comparison areas overtime to tease out any system-level variation, assuming that the DMA works will include pressure logging capabilities at the household level. The most accurate way of identifying trends in water supply would be to measure it directly through a water flow meter, replicating the work planned for the treatment DMA according to MCCU's DMA Works bidding documents. However, this approach would be expensive and logistically complicated as the installation of flow meters may require cutting into existing infrastructure. As an alternative, SI is proposing the use of pressure loggers to measure water supply at the tap. Pressure loggers will allow SI to comment on service reliability, how often water is flowing through the pipes. Service reliability is a valuable measure of water quality, as intermittent or stagnant water is more susceptible to bacteria, as well as the service quality as provided by GVWC. Pressure loggers, however, cannot capture NRW and SI will continue to rely on the measurements recommended in the preceding paragraph to comment on NRW difference between treatment and comparison DMAs.

For cost effectiveness SI recommends installing pressure loggers on a subset of comparison households (n = 150) as identified for the household survey, rather than all GVWC customers in the comparison DMA. By installing on survey households SI will be able to match data produced by the pressure logger to responses, perception and recall in the household survey, providing additional data points for analysis. To ensure the best use of this investment SI recommends receiving and reviewing pressure logger data for a one-year interval at interim and endline in line with expectations for data received from the treatment DMAs. There are two options for doing so:

1. *A fixed system*: A fixed system is the preferred option in which a pressure logger is purchased for each survey household and affixed to their tap for a period of one year. During that year the data collection firm would be tasked with traveling to the field monthly to collect readings from the pressure logger devices and submit them to SI. While this option requires the purchase of more hardware, it greatly reduces the logistical challenges and staffing costs provided under point 2 below while maximizing the data available.

2. *Rotating system*: A rotating system would require installing a pressure logger on a household for a predetermined number of days before rotating it to a new location. Under this option SI would rotate pressure loggers among survey households every 7 days so that each household is visited within the span of one month. Rotations would continue throughout the planned year. This option allows for the purchase of fewer sensors but requires high staffing costs as well as logistical coordination. Based on SI field experience with this system when evaluating MCC Tanzania, one technician can install 2-3 loggers each day, requiring a fulltime team of 4-6 people for a year of installations. In addition, movement of loggers creates gaps in data at the household level and provides increase opportunity for theft or breakage. The only instance in which SI considers this a more cost-effective option is by limiting data collection to a quarterly schedule, four months of the year. While this system would allow SI to capture seasonal variation in water supply it further limits the data available for analysis.

Under either of the above options SI would take care to utilize materials and lessons learned from the use of pressure loggers in the MCC Tanzania evaluation. While in utilizing this measurement technique SI may request limiting GVWC's access to comparison metered data during the intervention period to ensure that data are not used for decision-making in the comparison area, the addition of this activity not

only provides the most rigorous measurement opportunity for system level data but also provides a future investment/technology resource for GVWC post-evaluation.

Sensor installation and data collection in comparison areas will require securing prior permissions from GVWC. This is unlikely to pose as an obstacle as SI proposes using sensors that are not invasive to the GVWC property like pipes etc. The temporary sensors will be attached to outlets.

As noted above, for several system outcomes such as cost of water, billing, and collection, SI will use time-series data available with GVWC administrative records for treatment and comparison DMAs. SI will contact GVWC through MCCU for a data sharing agreement to gain access to relevant data. In this case, SI will have access to both pre-post trends in the treatment DMAs and will also be able to compare these trends against those of the comparison DMA. Besides assessing quality and performance, this source of data will also help SI to understand changes in coverage (number of customers) resulting from the activity, assuming that with better management and quality, more households will be interested to be registered GVWC customers but may be excluded from the metering activity. Assuming a continuous improvement in the customer base during the intervention and post-intervention periods, the time-series administrative records will help SI to assess the effect of exposure (duration) on longer term outcomes.

3.4.1.3 Household Level Analysis: Longitudinal Approach

As noted, the original DMA/Kiosk evaluation was proposed as an IE using a PSM technique. During December 2018, SI prepared and submitted a memorandum to MCC describing the requirements, including an option for a separate baseline survey, in order to conduct an IE with sufficient analytical power. The evaluation team, along with MCC, considered and later rejected an IE design on the ground that in terms of value for money, the resources will be better utilized to focus on system-level improvements in outcomes. This option, though restricting a rigorous IE on household-level outcomes, does not completely dispose of household-level analysis. The decision was made to conduct a comparative study of DMAs to record changes over time before and after the threshold.

Given this conclusion, SI proposes conducting a quantitative comparative study between the treatment DMAs and a comparison DMA selected by the THP. The comparative study will focus on comparing outcomes of interest between these DMAs to establish what changes occur especially in areas of coverage and access, household satisfaction on service provided, affordability, perceived water security and awareness building. This study will also allow the ET to analyze alternative water source use trends in the DMA. The comparative analysis at the household level can still suggest the existence and direction of a program effect even if these changes cannot be causally attributed to the program due to other potentially confounding and/or unobserved factors such as community participation, awareness, or the existence of other WASH programs.

Using the ASI KAP baseline survey as the foundation (see Section 3.4.4.1), SI proposes to survey the same households from Aberdeen, Kingtom and Cockle Bay to create a panel of households. This assumes that data are available to track baseline households from the ASI KAP survey. As part of the Baseline Memo, the availability of this data was noted as a concern and it is possible it was not collected and/or stored. If that is the case, SI proposes identifying a new sample at interim and endline using the same random sampling technique employed by ASI at baseline. The overall analysis will be a three-period comparative analysis for the three DMAs. As discussed earlier, although Cockle Bay serves as a useful case study comparison, validated through SI's review of ASI's DMA selection methodology, SI does not view this "comparison DMA" as a valid counterfactual for a more rigorous evaluation approach. In SI's view, limiting the comparison group to one contiguous area selected by the contractor opens the

comparison up to endogenous differences between the treatment and control DMAs that confound the counterfactual.

In addition to the system-level measurements and household surveys mentioned above, SI will supplement the DMA/Kiosk Evaluation with the following activities.[70]

3.4.1.4 Focus Group Discussions

If MCC elects to exercise optional baseline tasks for the evaluator, FGDs with customers at baseline could investigate themes related to drinking water source selection and the potential for differential outcomes by gender or socioeconomic status. These themes could then inform interim and follow-up sampling and design. At interim and endline data collection, FGDs will be conducted following preliminary quantitative analysis to understand potential causal explanations underlying program results (e.g., perceptions of utility, perceptions of service, etc.) as well as diversity in program results based on gender or socioeconomic status. These FGDs would be conducted separately by types of customer (on-premise taps, water kiosk, and public standpipe). SI will work with local coordinators and the data collection firm to recruit participants and organize focus groups in advance. Participants will be identified through the household survey for on-premise taps and recruited directly from kiosk and standpipe customers.

3.4.1.5 Key Informant Interviews

KIIs, at interim and endline, will target the standpipe/kiosk entrepreneurs as well as other private water purveyors (bottled water companies, water delivery companies, non-PPP water kiosks, water storage vendors, etc.). KIIs with the standpipe/kiosk entrepreneurs will aim to understand dynamics that may influence the sustainability of the PPP solution and its applicability to other contexts. They would also ideally include a review of financial records to understand the revenues for GVWC and entrepreneurs and costs to customers of the water kiosks. KIIs with other private water companies would aim to understand changes in the market for drinking water over the course of the program and any perceived disruptions from improved utility service. The ET will gain essential insights about private sector preparation and operational management of public service provision.

3.4.1.6 Process Mapping

MCC's program logic requires effective implementation of an NRW strategy. To test this assumption, we propose conducting a process mapping exercise based on perspectives of utility and contractor staff at interim and endline. The Bidding Document, Vol.2 provides details of each step of the process as well as detail specifications of materials to be used, proper measures for avoiding environmental damages and rehabilitation of households. As broadly described in Section 3.3.1.5, the process mapping will link supply and demand of a particular process within a system and will provide information on whether GVWC processes for NRW have improved over time by mapping information through process overview, process boundary, identification of supply and demand gaps and measures taken to bridge the gap. It will also look into the roles that THP played in the process. Both the KIIs and process mapping exercise will be conducted alongside the PE data collection to reduce fatigue among targeted respondents. This

[70] In addition to the methods proposed here, case studies were also considered and discussed with MCC as part of the DMA/kiosk evaluation. Case studies would investigate through KIIs, observation and data when available, select situations. An example was given of a case study comparison of kiosk management under the Sierra Leone THP with other PPP kiosks that are operational in other areas of Freetown. At this time, in order to provide the most rigorous form of analysis the ET is proposing household level analysis over case study design.

method is useful descriptively to evaluate and support claims of causality in the absence of an experimental method that can establish causality mathematically.

3.4.1.7 Direct Observation

Direct observation will be made at peak periods at water kiosks, at public standpipes in the two treatment and a comparison DMA, and at public (Milla) water tanks. Selected leak repairs and capacity building trainings will also be directly observed. The DOs will allow the ET to use standardized tools as means for verify information gathered through other data collection methods about THP outputs and outcomes.

3.4.1.8 Water Hygiene Measurements

While measurements proposed under the ITSA and household survey above will assist in understanding water quality, especially as it relates to water pressure, reliability, and continuity of access and consumer perceptions, SI proposes utilizing additional water quality data as outlined in ASI's Water Quality Monitoring Plan.[71]

ASI's proposed Water Quality Monitoring Plan includes the collection of 104 random samples, from hydrants, water kiosks, private yard taps, public stand taps and institutional connections, over the course of one year (one sample collected per week per treatment DMA). The data collected by ASI will create a panel dataset consisting of 52 X 2 observation. Samples are to be evaluated on taste, odor color, turbidity, iron, manganese, hardness, PH, residual chlorine, and fecal coliform against WHO guidelines by GVWC staff. SI recommends the inclusion of water quality testing in the comparison DMA as well.

As ASI notes in their report, water quality is expected to improve for all GVWC customers in the city of Freetown due to DFID pipe rehabilitation works scheduled to start in 2019. Water quality measurements taken only from the treatment DMAs may falsely attribute any improvements in water quality to pilot DMA works rather than citywide changes. If feasible, SI recommends GVWC, with support from ASI, includes water quality testing in the comparison DMA in order to best maximize the equipment made available. If this is not feasible, SI proposes *as an optional activity* conducting our own water quality testing in the comparison area focused on the hygienic reliability of the water as noted in ASI's report, namely the presence of residual Chlorine and E.coli. Limiting these tests to focus on water hygiene rather than the full WHO parameters (appearance, staining, taste and smell) conserves costs while getting at the most important aspect of water quality, its safety for consumption. Utilizing SI experience in conducting these tests for MCC Lesotho, SI is proposing purchasing testing kits to conduct once weekly tests in the comparison areas. Test will be carried out by SI local consultants during the interim period at a random sample of hydrants, water kiosks, private yard taps and public stand taps over the course of the evaluation. At endline SI will utilize quality testing done by GVWC if it has expanded to all DMAs at that time or conduct our own tests in treatment and control areas if GVWC data are not available.

Currently SI is proposing retaining ASI's sample strategy of one water quality measurement per DMA per week. However, if funding permits SI recommends increasing testing to 4-5 samples per DMA per week to allow for weekly testing at the entry point to the DMA and several touchpoints (public and private connections) throughout.

Outcome measures

[71] Adam Smith International. Institutional Strengthening Report: Task 4.12.3 Water Quality Monitoring Plan for DMA. 2019.

The table below presents the outcome indicators, expected sources of data and frequency to address the evaluation questions.

Table 8: DMA Outcome Level Measurements

EQ	Outcome	Indicator	Frequency	Source
9	Non-revenue water	percent of water distributed that is billed	Monthly (records)	GVWC & Contractor Records
9	Non-revenue water	percent of billed water that is paid	Monthly (records)	GVWC & Contractor Records
9	Non-revenue water	percent of water distributed that is lost through leaks/illegal connections	Hourly (sensors)	Flow Meters
10	Service reliability	Hours per day of service/use	Per data collection period (survey), Hourly (sensors)	Household Survey Pressure loggers[72]
10/ 11	Cost of water to households	Leone billed per month	Monthly (records), per data collection period (survey)	GVWC billing records Household Survey
10	Water quality	WHO risk category compliance, point of consumption	Weekly	Water quality testing
10	Water collection time for household members	Minutes per day collecting water, by gender	Per data collection period	Household survey

[72] Only pressure loggers can capture service reliability, it is unclear to SI if this measurement will only be available at DMA distribution points or throughout the DMA. Clarification from the DMA upgrades procurement is expected at the end of March.

EQ	Outcome	Indicator	Frequency	Source
10	Consumer citizenship attitudes and behaviors by gender	Attitudes toward paying for water, attitudes toward those who steal water or don't pay for it, and self-reported willingness to pay, reported and total illegal connections	Per data collection period Monthly (records)	Household survey FGDs GVWC records
10	Customer satisfaction	Likert scale of satisfaction with utility or kiosks over past year, by gender	Per data collection period	Household survey FGDs
11	Standpipe management/Kiosk activity	Time between establishment and resolution of customer complaint	Monthly (records), Per data collection round (survey/FGD)	GVWC records and/or GVWC customer complaint mechanism FGDs, household surveys
11	Standpipe management/Maintenance / Kiosk activity	Standpipe "downtime" due to maintenance or disrepair over past month. Additional information collected by the ET related to standpipe management will include operation details, estimated costs, service reliability, etc.	Monthly (records), Per data collection round (survey/FGD)	GVWC records FGDs, household surveys, KIIs Flow Meter and UC Boulder sensors
11	Standpipe water service/ Kiosk activity	Hours per day of service;	Hourly;	Flow Meter and UC Boulder sensors
11	Standpipe water service/ Kiosk activity	Cubic meters provided per month, filling and use	Monthly (records) Hourly (sensors)	GVWC records Flow Meter and UC Boulder sensors
11	Standpipe water service/ Kiosk activity	WHO risk category compliance, point of collection	Weekly	Water quality testing
11	Revenues for GVWC	Leone billed to entrepreneurs per month	Monthly (records)	GVWC records

EQ	Outcome	Indicator	Frequency	Source
		for water; GVWC collection.	Per data collection period (KII)	KIIs with standpipe entrepreneurs/review of records

3.4.2 Timeframe of Exposure

Time-series data collection will be ongoing throughout the period of the THP and at least 3 years after the inauguration[73] of DMA/kiosk activities. Prior to the inclusion of system measurements as part of the DMA upgrades, SI will rely on data available through GVWC records. Once system measurements are available SI will track these measurements at least three-year post inauguration, as feasible.

For the comparative study, the baseline was implemented in mid-2018 by an external consultant, ASI. The goal of utilizing these data is to capture pre-intervention levels of outcomes and covariates of interest. During the preparation of the Baseline Memo, the ET reviewed the terms of reference and materials with the objectives of ensuring our data QA standards[74] were met and that the survey captures our key outcome and deterministic variables in a way that is useable in following analysis periods.

The follow-up household level data collection effort's goals are to estimate changes in outcomes in the short (interim) and intermediate (endline) term. The justification for these two follow up studies is that some outcomes, such as service reliability for rehabilitated standpipes, may be expected to improve almost immediately while others, such as consumer citizenship attitudes and water collection times, would be expected to take longer to manifest. In addition, the endline will allow SI to evaluate whether immediate gains have been sustained several years out from the program. Interim data collection should occur at least one year after baseline to allow for program maturation and behavior change. However, due to delays in the program implementation, SI is proposing interim data collection take place almost two years after the baseline, approximately 9-12 months after the construction and upgrading of major DMA/Kiosk infrastructure. Endline data collection should occur at least three years following the inauguration of the NRW strategy to allow for operational challenges to the program and long-term behavior change.

3.4.3 Study Sample

3.4.3.1 Sampling Approach

The sampling area, including Aberdeen and Kingtom as targeted DMAs and Cockle Bay as a comparison, for the DMA/Kiosk Activity was determined by GVWC and ASI prior to the start of the evaluation period.[75] Targeted DMAs were considered fixed as part of the program intervention and SI will continue to utilize Cockle Bay as the comparison area given available baseline data and a review of its comparability as part of the Baseline Memo.

[73] Anticipated for September 2019

[74] SI noted some concerns with data quality as part of the Baseline Memo but these are primarily restricted to the Customer Mapping survey.

[75] Adam Smith International. Guma Valley Water Company (GVWC) Institutional Strengthening, Urban WASH Sector Coordination and District Metering Area and Water Kiosk Demonstration Pilot. ASI.

Based on the discussion with MCC as a follow up to the Baseline Memo, the ET decided to use the existing baseline KAP survey data collected by a third party. From the KAP Survey Report, the ET understands the KAP sample was selected using a "cross-sectional multi-stage clustering" sample design. The first stage included the clustering of households within Enumeration Areas (EAs) based off the Sierra Leone 2015 census; 67 EAs were identified among the 3 DMAs. The KAP survey then chose to sample 10 percent of households randomly from each EA. The authors of the baseline report estimated that a sample size of 551 households was required to achieve a 95 percent confidence interval and 4 percent margin of error within the areas sampled.[76] The report argues that given there are 67 EAs, the study identifies a total sample size of 670 households, which is well above well above the estimated required sample of 551.[77] Table 9[78] shows the distribution of households across the EAs and in the total sample as reported in KAP study.

Table 9: Sample size based on the EAs

DMA	Number of EAs	Number of Households	Sample size (10 percent)
Aberdeen	20	2,000	**200**
Cockle Bay	15	1,500	**150**
Kingtom	32	3,200	**320**
TOTAL	67	6,700	670

3.4.3.2 Sample unit and Sample size

Absent the possibility of an IE, SI is proposing conducting a longitudinal analysis involving the same households over time from the treatment DMAs. For this design, comparison will be made overtime for the same households before and after the treatment. The detail methodology is given in Annex 6: Household Sample Size Calculations. It uses information on desired level of change, and Intra-Class Correlation (ICC) across clusters to have a cluster (EA) level analysis based on household level outcomes. The advantage of this longitudinal approach is that it requires a smaller sample size compared to a rigorous IE. However, it is impossible to attribute causal effects to the program using this design.

Table 15 in Annex 6 suggests that a sample of around 1,000 households overall is needed to assure a minimum detectable change in access, including (i) 30 percent in access to piped water in premise, (ii) 15 percent in access to piped water in own premise; (iii) 30 percent in access to tap water in yard; (iv) 20 percent in access to tap water outside yard and (v) 35 percent in access to water in neighbors' yard with

[76] Adam Smith International. Knowledge, Attitudes, and Practice (KAP) Survey Report. ASI, 2018, p.14.

[77] The EAs were evenly distributed by Statistic Sierra Leone (SSL) in 2015 with an equal number of households of 100 for each EA. The report informs that it uses the same EAs that SSL has constructed for the 2015 Census. The three DMAs have altogether 67 EAs. As a result, a random selection of 10 percent of households from each EA leads to 670 households total from the three DMA areas for the KAP survey.

[78] Adam Smith International. Knowledge, Attitudes, and Practice (KAP) Survey Report. ASI, 2018.

80 percent analytical power. Note that (as shown in 14 in Annex 6) the required sample size is almost half of that needed for the originally proposed IE but still greater than what was included in the baseline KAP survey and accessible for SI's proposed evaluation. This may prevent the evaluation from confirming observed differences with statistical certainty but will still provide an opportunity to observe trends in treatment and comparison DMAs overtime.

3.4.4 Primary Data Collection

3.4.4.1 Instruments

All Household Surveys, KIIs, FGDs, direct observation reporting templates and process mapping exercises will be developed during the data collection instrument design phase drawing from SI's experience in quantitative and qualitative evaluation design. All instruments will be pre-developed and when possible, field tested, prior to data collection. Care will be taken to ensure similar or overlapping themes and questions to enable greater data triangulation. It is expected that instruments will be revised between interim and endline data collection to consider information learned from interim data collection.

The household survey will largely follow the KAP survey administered by ASI at the baseline including sections on Household:

- Socio-Economic Status
- Water Sources and Perceived Supply
- WASH Practices
- Cost of Water and Payment Schemes
- Social Responsibility around NRW and Perceptions of GVWC
- Gender and Disability (disaggregates)

Questions may be updated and additional questions, especially around perceived water security,[79] may be added. The household (HH) survey will allow the ET to analyze gender issues through the incorporation of gender focused questions, such as those about fetching water. Household surveys will be administered electronically.

KIIs, FGDs, direct observations and process mapping exercises will largely focus on the operations, management and customer service of water kiosks though may include measurements aimed at capturing community perceptions of GVWC and NRW. The HH survey, KIIs, and FGDs will each offer opportunities to capture Gender Based Violence (GBV) issue data.

3.4.4.2 Rounds and timing

Household interim data collection is anticipated for late-2020 with endline data collection occurring approximately two years post-THP close as noted in section 3.4.2 above. Interim data collection was originally planned one year after baseline data collection. However, project workplans shared in early 2019 do not have DMA/Kiosk upgrades commencing until June 2019 with finalization by November 2019.[80] Once finalized (i.e. inaugurated as referred to previously in this document), it should be possible

[79] SI is exploring using the household water insecurity survey for international comparison: "Household Water Insecurity Experiences (HWISE) Scale." LIBRARIES Blog, © 2018 Northwestern University

[80] At the time of EDR revisions in August 2019, DMA works are yet to commence and the timeline is uncertain.

to start retrieval of system-level sensor data in late-2019. Retrieval and analysis of this data will continue for at least two years post-inauguration and be incorporated into interim and endline reporting.

At the household level we expect few benefits/outputs to be immediately realized and are therefore proposing interim data collection 9-12 months post-inauguration. It is SI's expectation that while immediate benefits may be realized through pipe upgrades, the main benefit of the DMA/kiosks activities will be the use of real-time data for improved decision making in areas of NRW, water supply and quality, and customer complaints and outreach. We expect this decision-making process to evolve over time, especially in the first few months as GVWC and kiosk operators familiarize themselves with the system and household beneficiaries are oriented on kiosk operations and costs. Since the provided work plan notes that operationalism of DMA works is a procedure lasting through January 2021 SI has used that timeline to benchmark interim data collection.

Endline data collection is then planned to take place two years after interim data collection (three years post-inauguration) in order to capture continue learning, decision making, and sustainability of DMA upgrades and data usage.

3.4.4.3 Respondents within the sample unit

The desired respondent for the household survey is the person most responsible for decisions related to water collection and expenditures, likely the household head. If this person is unavailable at the time of survey administration, we would permit the survey to be conducted with another adult household member who is involved in and informed of decisions related to household water use. We envision maintaining the same respondent for the entire questionnaire, but may allow respondents to refer questions to household members that are better informed based on the specifics of the question. SI intends to use the same households from ASI's baseline. If this is not feasible, new households will be selected from the same communities employing the sampling strategy ASI used at baseline.

3.4.4.4 Staff

SI's Program Manager, Dr. Basab Dasgupta, will lead the quantitative data collection effort, including potentially participating in instrument piloting and enumerator training. He will be supported by Junior Analyst, Carly Farver (SI-HQ) and a local survey specialist, Peter Ghombo, who will provide field monitoring of data collection in both locations under the direct guidance of Dr. Dasgupta and Ms. Farver. The local senior analyst for water, Mr. Joel Kamanda, may also participate in qualitative data collection. Dr. Charles Pendley will advise on quantitative and qualitative instruments and analysis remotely. Dr. Evan Thomas will advise on sensor data and analysis. Ms. Farver will participate on both the quantitative and qualitative field team at interim and endline data collection and work with Dr. Basab to ensure there are synergies between the quantitative and qualitative efforts.

SI understands that a local data collection company will be procured through MCCU at interim data collection. SI will work with MCCU to ensure the development of proper Terms of Reference (TOR) and oversight of the data collection company is in place. If feasible, this same data collection company will be used at endline, this time procured by SI directly as it will take place after the THP has ended. At both interim and endline SI will provide expert guidance in a comprehensive training, at least five days in duration and including field practice, to all field staff employed by the data collection company. SI will also be responsible for, and employ the staff noted above, for data quality oversight during interim and endline data collection. The budget and corresponding budget narrative in Annex 3 provide additional detail into staffing and data collection costs for optional and required activities.

3.4.4.5 Data processing

Since we intend to conduct electronic data collection, we expect to receive data regularly throughout field work, which we will import into Stata and using the SI-developed errout Stata command will check for a variety of common logic, range, missing value, skip, and outlier errors. This can be conducted in near real time and generates a log of errors for discussion and verification with the data collection partner, as well as for further training of staff on common errors. Once SI receives the final dataset, we will conduct data cleaning, again checking for missing data; logic, range, and skip errors; and outliers, using Stata .do and log files. Identified issues will be discussed with the data collection partner for verification and any changes will be entered into .do files with notes explaining the change. Relevant variables will be transformed for analysis. All data cleaning, management, and analysis will be conducted through Stata .do files to ensure transparency and reproducibility of results.

3.4.4.6 Data quality

For every IE in SI's portfolio, a systematic QA framework for survey data collection is utilized. SI's system aggregates learning from nearly a decade of experience conducting IEs worldwide. This QA framework can be adapted for non-IE quantitative analysis, as is the case in this evaluation. It includes comprehensive guidance documents for planning data collection and monitoring data quality in order to ensure these activities are carried out with the highest level of integrity. SI will use checklists that accompany guidance on high-quality data collection to ensure compliance with all minimum requirements and best practices. SI also uses tools and procedures outlined in our guidance in working with our local data collection partners, which allows SI to monitor performance throughout data collection, and contributes to building local capacity for high quality data collection. SI's data quality approach is three-fold and is summarized in Figure 5 below.

(a) Error Prevention		(b) Field-based Quality Control		(c) Independent Monitoring	
QC Measure	***Responsible***	**QC Measure**	***Responsible***	**QC Measure**	***Responsible***
Data collection team structure	*SI requires in procurement; local data collection partner implements*	**Supervisor observation**	*Partners implement; SI monitors observation forms*	**High frequency checks**	*SI conducts, reports issues to partners; partners reconcile*
Field-team training, pre-testing, and piloting	*SI and partners collaborate*	**Co-enumeration**	*Partners implement; SI analyzes forms and IRR*	**Monitoring electronic data collection controls**	*SI monitors and analyzes; reports issues to partners; partners reconcile*
Electronic data collection and build-in quality controls	*SI codes the forms; partners tests and provides feedback*	**Supervisor form checks**	*Partners implement, report major issues to SI; SI reconciles major issues*	**Back-checks**	*SI monitors and analyzes; reports issues to partners; partners reconcile*
Field manuals and protocols	*SI sets required content; partners develop; SI reviews and contributes*	**Team debriefs**	*Partners implement, report major issues to SI; SI reconciles major issues*		

Figure 5: SI's Data Quality Assurance Strategies

3.4.4.7 Summary Table

A summary of events as related to SI's data collection rounds are included in Table 10 below.

Table 10: Summary Table for Rounds of Data Collection

Round/Type	Timing	Respondents	Sample Size
Interim/HH Survey	September – October 2020	Households in targeted and comparison DMAs	670 Households Total
Interim/FGD	September – October 2020	GVWC Customers (disaggregation by gender/water source) Kiosk customers disaggregated by gender.	3-4 FGDs per DMA
Interim/KII	September – October 2020	Kiosk Operators, Other Private Water Sellers (AWSPs)	5-10 individuals
Interim/Direct Observations	September – October 2020	Kiosks	5 sites
Interim/Process Mapping	September – October 2020	NRW Strategy	1 group (5-10 GVWC employees)
Endline/HH Survey	September – October 2022	Households in targeted and comparison DMAs	670 Households Total
Endline/FGD	September – October 2022	GVWC Customers (disaggregation by gender/water source). Kiosk customers disaggregated by gender.	3-4 FGDs per DMA
Endline/KII	September – October 2022	Kiosk Operators, Other Private Water Sellers (AWSPs)	5-10 individuals
Endline/Direct Observations	September – October 2022	Kiosks	5 sites
Endline/Process Mapping	September – October 2022	NRW Strategy	1 group (5-10 GVWC employees)

3.4.5 Analysis Plans

To analyze the project's impact on key outcomes of interest, we must first verify that the project achieved its intended *outputs*. Prior to conducting analyses, we will use a combination of project monitoring data, GVWC records (as available), information gathered from key informant interviews and FGDs, and household survey data to determine whether the intended DMA upgrades were successfully installed. Separate FGDs will be conducted for households and the water collectors from kiosks will be disaggregated by gender to capture information on waiting time, affordability, quality and availability, and gender-based violence-related issues. The ET intends to conduct FGD for business owners to understand their point of view on quality of service delivery.

We will then review intended outputs to capture any potential relationships between project activities and NRW reduction, household access, behavior and practices and GVWC billings and collections. Specifically, this will include:

1. Descriptive statistics to compare outcome variables across key covariates, such as socioeconomic status and in some cases by age or gender. Also, SI will compare the standardized mean difference between comparison and treatment group over time to identify if any meaningful difference can be detected.

2. Analysis of determinants of outcomes: In the presence of comparison DMA, the SI proposes to use difference in difference (DID) method to compare changes between the treatment and comparison DMAs over time. Although the identification of a comparison DMA by a third party assures that the DMAs were similar in several aspects (population composition, socio-economic status etc.) at baseline, a parallel trend assumption is hard to establish due to absence of historical data. In addition to DID analysis, SI also proposes to conduct a decomposition analysis to identify the rate of returns of household and individual level demographic and economic variables on select outcomes of interest, such as water source, WASH practices and customer behaviors in changes over time. While the relationships cannot be confidently considered causal, they may be instructive in identifying additional questions and research in household water consumption and use.

3. Trend Analysis*:* For system level outcomes*,* SI will analyze high frequency water supply and quality data from different types of sensor and kiosk operations. The trend analysis will include daily information of water pressure (from pressure loggers as feasible), and daily water consumption information (from smart meters) to identify shift in trends in outcome indicators. Since the system-level interventions will not have a corresponding baseline and it has not yet been determined if employing the same technology in a comparison DMA is feasible, SI will instead investigate changes in trends overtime and at key touchpoints (interim, which occurs at the end of the THP, and endline, which capture post-intervention outcomes). This analysis will be supplemented by KIIs. We hope to establish whether the intervention was effective in improving the quantity and quality of water delivery in targeted DMAs during the intervention and is also sustainable in the post-intervention period. If we find, for example, that the DMA upgrades, including kiosks, are non-operational (pipes, meters, etc. are broken) and/or water supply is not changing/improving over time or that the comparison areas have registered improvements due to other THP activities or outside interventions, null or limited results on key outcomes may be expected.

3.5 Evaluation Challenges & Limitations

3.5.1 Limitations of Interpretations of the Results

With a rigorous mixed methods approach, we will be able to speak to which outcomes have changed as well as how and in what ways they have changed as part of the THP. We will also be able to uncover the nuanced ways in which capacities and processes have (or have not) changed. However, except in the cases of the DMA/Kiosk evaluation where household, sensor and quantitative secondary data are available, we may not be able to precisely measure the extent to which outcomes have changed. Additionally, while the evaluation is focused on the contribution that the THP has had on outcomes, without a counterfactual, we will be unable to causally attribute changes to the THP alone.

3.5.2 Risks to the Performance Evaluation Design

At the start of SI's involvement it was already too late to collect baseline data for the power and water PEs. While SI will collect as much documentation as possible and will include retrospective questions in the interim and endline data collection rounds, these mitigation efforts will not be a perfect solution to the lack of baseline data. This will weaken SI's ability to measure the extent to which changes may have occurred.

Additional potential risks for the PEs include reluctance by partners to share data that are not directly required to be shared under the THP, data quality concerns with secondary data, and challenges to collecting data (both primary and secondary) after then end of the THP when data sharing is no longer required and some staff may have moved on. All potential risks will be actively managed by SI, in collaboration with MCC and the MCCU. Regular communication regarding risks and potential mitigation efforts will help minimize any effects on the evaluation.

3.5.3 Risks to the DMA and Standpipe (Kiosk) Evaluation Design

Several large risks to the DMA/Kiosk Evaluation Design, including identification of a comparison group and appropriate baseline data, were already noted and addressed as part of the baseline memo and subsequent review process with MCC. From these discussions the decision was made to move away from an IE design and instead focus on system-level and household measurements which could provide informative, though not causal, relationships as part of the program.

Moving forward, the largest risks to the DMA/Kiosk Design include:

- *Further delay of DMA/Kiosk upgrades.* While data collection at the interim can continue to be postponed doing so would (1) increase the time period between the baseline and interim, potentially introducing confounding factors (political, economic, etc.) at the local and national level and (2) would likely require the DMA/Kiosk evaluation to be conducted on a different time schedule than that of the PE, assuming PE activities proceed as planned, making it more difficult to integrate results between the two evaluations.
- *Poor quality and availability of sensor data*. This includes the fact that (1) SI has noted several concerns in identifying a pre-intervention or simultaneous comparison data source with which to benchmark treatment DMA sensor data against. If it is not feasible to identify a reasonable comparison point in this way, we will still be able to use treatment DMA sensor data to explore trend changes over time. Risks here also include (2) quality of treatment sensor data, including ensuring installations are appropriate to capture planned measurements in NRW, initial access

to data for the evaluation team and appropriate upkeep of data at least two years post THP in order to measure effects over time.

- *Inability to identify baseline households for follow-up surveys*. There has been discussion that identifying information for baseline KAP households was not captured. SI would like to confirm this is the case and if so will proceed in with the alternative option of drawing a new random sample of households for each survey round.

3.5.4 EDR Modifications

Based on discussions with MCC, the ET understands MCC's preference of keeping this design report as an open, editable document in order to incorporate changes due to unforeseen challenges in the future. SI will be responsible for paying special attention to risks to the evaluation design posed by changes in program implementation. We will use interim data collection to update the status of implementation and verify the ongoing validity of our methodologies. Overall, the Evaluation Design Report (EDR) will be a dynamic document, open for required adjustment in peripheral indicators throughout the design phase. However, any modifications to the design or proposed indicators will be closely monitored, tracked and noted in reports should their modifications be a reaction to changes in program design.

4 ADMINISTRATIVE

4.1 Summary of IRB Requirements and Clearances

The Sierra Leone THP evaluation will be conducted in line with human subjects research guidelines both in the United States and Sierra Leone, as well as in accordance with MCC's policies and procedures. SI has an independent and fully functional Institutional Review Board (IRB) with established protocols for gathering informed consent, protecting anonymity and identifying information, and ensuring ethical data collection—including from children and other vulnerable populations. To ensure compliance with our high ethical standards, all studies involving vulnerable populations must pass through formal IRB review prior to data collection and all research staff must complete a certified course in Protecting Human Research Participants through the National Institutes of Health (NIH) or Collaborative Institutional Training Initiative (CITI). SI's internal IRB is registered with the U.S. Department of Health & Human Service's Office for Human Research Protections. SI will use MCC's standard template for informed consent.

In addition, SI closely monitors and adheres to human subjects research regulations in its countries of operation to ensure all studies are registered and fully compliant with local law. Preliminary research using the International Compilation of Human Research Standards, 2019 edition, from the U.S. Department of Health and Human Services has yielded no local IRB requirements in Sierra Leone for non-health related research. While data will be captured for all members of selected households, the information will be provided by the household head or other knowledgeable person who is over the age of 18. No minors will be interviewed in this study.

Prior to the commencement of data collection, SI will ensure that all protocols, survey instruments and informed consent procedures to be used as part of the research have been reviewed and approved by SI's IRB. SI will provide evidence of such approval to MCC.

4.2 Data Protection

SI's process for respecting privacy of respondents during data collection, transfer, storage, analysis, disposal and dissemination is governed by SI's data security guidelines, which are aligned with MCC's microdata guidelines.

The rights and privacy of all participants who take part in data collection will be respected throughout the study. Ethics training will be given to all enumerators prior to data collection addressing issues on participants' right to know what the research is about; the right to choose whether to participate; the right to privacy of responses; and the right to have no harm done to them.

All persons engaged in this contract will not divulge any information, whether obtained orally or in writing from, to any unauthorized person for any purpose. Enumerators will be trained to keep all data and particularly identifying information confidential. They will be instructed not to discuss responses with anyone outside of the study. All relevant stakeholders will sign non-disclosure agreements. Additionally, persons engaged in this study will not directly or indirectly use or allow the use of confidential information (including personally identifiable information) for any other purpose other than that directly associated with the purposes of this study. During reporting and dissemination of findings, only aggregate results will be shown and results will not identify individual

respondents. Should a KII provide an illustrative quote that may identify an individual respondent given their position, permission will be sought from that individual prior to publication.

As standard practice, any data with personally identifiable information (PII) collected on paper or electronically will be stored on password-protected electronic devices or in a locked room. SI will ensure all hardware and software meet federal standards for PII data, in accordance with OMB 06-16 and NIST (National Institute of Standards and Technology) requirements. Confidential information (including PII) that is collected with the survey will immediately be separated from the rest of the dataset. Only a small number of approved researchers will be able to link responses to the individual who provided them. Upon conclusion of the study, all PII will be destroyed. All data collected will be stored on a secure server that only authorized personnel will have access to. When sharing data files, data handlers will use a secure file transfer system.

In accordance with MCC's protocol, SI has provided non-disclosure agreement letters signed by all personnel working on this task.

4.3 Preparing Data for Public Use

SI will adhere to MCC's open data policy with regard to preparing data for publication. All primary quantitative data collected by the evaluation will be prepared and submitted to MCC according to the most updated version of the Disclosure Review Board (DRB) guidelines available at the time of data collection. On an instrument-by-instrument basis, SI and MCC will weigh the utility of publishing primary qualitative data (even in a restricted-access database) against (i) the risks of respondent re-identification and (ii) the risks of adverse effects on data quality from disclosure. In the event that the utility of this data outweighs the risk of re-identification, and that respondents can be adequately informed via a consent script as to the data's intended use without jeopardizing their willingness to be forthcoming with interviewers, SI will submit this primary qualitative data to MCC as part of the DRB process.

4.3.1 Access and Documentation

All datasets submitted to MCC will be accompanied with completed documentation in the form of standardized metadata. SI will use the MCC Evaluation Metadata Template, Nesstar, which specifies the required metadata elements for documentation purposes. This method is compliant with the international Data Documentation Initiative (DDI) 3 and Dublin Core Metadata Initiative (DCMI) 4 standards, enabling compatibility with various data archiving systems. In addition to completing the Metadata Template, SI will submit the following necessary materials as listed in Table 11.

Table 11: Documentation

Element	Requested Format	Notes
Metadata File	Nesstar	The metadata file is used to populate a public catalog entry. The metadata should be updated upon the approval of an updated design report or final report. Do not attach any datasets or related documents under the "other materials" or "external resources" sections.

Element	Requested Format	Notes
Cover Sheet	Word	The first page of the document *DRB Cover Sheet and Anonymization Worksheet. The cover sheet* briefly outlines the contents of the survey package, listing all included files and their formats and purposes.
Enumerator and trainer manuals	Word, searchable PDF	Guides for survey enumerators, supervisors, and trainers, as well as manuals used to train each of these individuals
Questionnaires	Word, Excel, or other as appropriate	Survey instruments/questionnaires
Original, Raw Data	STATA (or other format agreed with MCC)	This is the complete data file(s) submitted by the survey firm with appropriate and logical variable names and labels, and including any necessary personally identifiable information of survey respondents. Data files should be submitted in Stata version 11.0 or higher.
Public Use Data	STATA (or other format agreed with MCC)	This is anonymized data following the Anonymization Worksheet (section 3). The public use data must include all non-sensitive data collected for the evaluation, not just the data used for the formal analysis. Data files should be submitted in Stata version 11.0 or higher.
Public Use Data Codebook	PDF	Codebook of public use data files. The codebook should include a labelbook as well as basic summary statistics including frequency and distribution information.
Analysis Data	STATA (or other format agreed with MCC)	MCC is specifically committed to enabling the re-creation of results produced by independent evaluators. In order to do so, consultants must either provide full STATA do files used to produce analysis files from public use files, or submit anonymized analysis files as well. Submission of anonymized analysis files only is insufficient considering the extent to which evaluator's judgment and quality control can influence the variables constructed. Therefore, sufficient documentation is required in order to define how analysis files are produced from public use data.

Element	Requested Format	Notes
Evaluation Design Report, Baseline Memo, and Evaluability Assessment	Word or searchable PDF	These documents (deliverables required by the contract terms of reference) provide additional useful design and analytical information for users of the data – particularly if this information cannot be reasonably included in the metadata. Evaluators should ensure that all public use documents/reports have been reviewed and edited to remove any references, such as geographic locations, that may threaten or undo anonymization efforts. In these cases, the evaluator should also provid**e internal use only documents/reports that include all removed information.**
Analysis Programs & Command Files	**STATA do files (or other format agreed with MCC)**	Programs, command files, and/or ".do" files used in the preparation of the data and the analyses presented in the Final Report. These programs can be provided in any format using any widely used statistical package or software, although Stata is preferred. They should include data merging, imputing, and other preparation work, key summary tabulations**.**

4.3.2 Privacy

Only aggregate results will be shown during reporting and dissemination of findings to protect the privacy of respondents. SI will be responsible for ensuring that all public release documentation is reviewed to minimize any risk to respondent confidentiality. This may require submission of internal use only and public release reports.

All public use data will be fully anonymized (free of Personally Identifiable Information, including geographical identifiers and variables that allow others to deduce the identity of individual subjects) to the extent possible. SI will submit the following anonymization package to the MCC Disclosure Review Board (DRB) for review of any public use data:

- Completed Anonymization Worksheet
- Metadata
- Codebooks of public use data
- Public use data files
- Informed consent statement
- Questionnaires

SI holds the right to determine if it is more appropriate to submit a proposal for the anonymization efforts for review and discussion with the DRB prior to official submission. In this case, SI will submit:

- Completed Anonymization Worksheet
- Metadata
- Informed Consent statement

- Questionnaire

4.4 Dissemination Plan

It is important that the evaluation answers the EQs and that those findings translate into policy actions by MCC, MCCU, GoSL, and other stakeholders. SI's dissemination and utilization plans for the THP evaluation will articulate an understanding of the specific context and target audience and how to reach them. These may vary between interim and endline dissemination given the changing landscape and program close. Dissemination will be targeted to the needs of primary users (MCC, MCCU, and the GoSL) but also consider the interests of other primary stakeholders including GVWC, EWRC, EDSA, EGTC, and key development partners including the World Bank, academics, and civil society groups.

The THP evaluation will include two interim and final evaluation reports (PEs and a separate DMA/kiosk evaluation report). Following the approval of these reports, particularly the final evaluation reports, SI will develop briefs that synthesize findings across the evaluations. SI will determine with MCC/MCCU the specific audience and content/focus of these briefs, but they will likely cover policy relevant findings for the water and power sectors, and implementation findings. These briefs will seek to combine relative narrative threads between the evaluations, such that stakeholders have immediate and intuitive access to the connection between findings regarding the GVWC Institutional Strengthening Activity and changes that its efforts yield within the DMAs on service provision and customer satisfaction, for example.

Following the approval of all reports and dissemination materials, SI's key personnel will lead results dissemination and utilization workshops, working collaboratively to facilitate these events with MCC/MCCU. SI will conduct two main dissemination events – one findings presentation in Washington, DC at MCC HQ, and at least one Sierra Leone-based dissemination workshop for MCCU and the GoSL. Rather than provide direct recommendations, evaluators will aim to help MCC, local stakeholders, and other interested parties to interpret how evaluation findings and conclusions are relevant to their respective interests and uses.

4.5 Evaluation Team Roles and Responsibilities

The evaluation team will be comprised of a field evaluation team and support staff at SI headquarters. In some cases, evaluation team members will have a role both as field evaluators and management support staff. The evaluation team includes all personnel described in Table 12.

Table 12: Evaluation Team Roles

Personnel	Role	Technical/Support	Responsibility
Dr. Basab Dasgupta (SI)	Program Manager/Team Leader	Both	Dr. Dasgupta is the primary responsible party for administrative coordination, deliverables, and ensuring the technical success of the evaluation.
Mike Duthie (SI)	Senior Analyst	Technical	Mr. Duthie provides technical guidance related to evaluation methodology as needed throughout the contract.

Personnel	Role	Technical/Support	Responsibility
Carly Farver (SI)	Junior Analyst	Both	Ms. Farver assists in deliverable review and submission, oversees data quality adherence, facilitates coordination with MCC and other stakeholders, and leads all aspects of HQ support.
Anna-Karin Hess (SI)	Administrative Assistant	Both	Ms. Hess provides evaluation team administrative and logistical support.
Dr. Charles Pendley	Senior Analyst/Water Sector Expert	Technical	Dr. Pendley provides technical expertise in the water sector and contributes to the design and implementation of the evaluation's water-related components.
Matthew Addison	Senior Analyst/Electricity Sector Expert	Technical	Mr. Addison is responsible for technical and methodological leadership of the Electricity Sector Reform Project and Electricity Regulatory Strengthening Project evaluations.
Dr. Evan Thomas	Senior Analyst	Technical	Dr. Thomas is responsible for advising the evaluation team on the sensor data administration and use.
Joel Kamanda	Senior Analyst/Water Engineer (Local)	Technical	Mr. Kamanda supports Dr. Pendley in the technical and methodological leadership of the Water Sector Reform Project evaluation activities and DMA and Standpipe Demonstration Activity.
TBD	Qualitative Specialist	Technical	TBD will support qualitative data analysis for both sectors in the Performance Evaluation.
Peter Ghombo	Survey Specialist (Local)	Both	Mr. Ghombo provides technical expertise and contributes to the design and implementation of the electricity-related components of the evaluation.

4.6 Evaluation Timeline and Reporting Schedule

Table 13 displays the overall schedule for the evaluation, while Table 14 provides a detailed breakout for interim data collection. This timeline assumes an extension of the Sierra Leone THP until March 31st, 2021. If an extension is not granted, the EDR will be updated accordingly.

Table 13: Data Collection Timeline

Name of Round	Data Collection	Data Cleaning & Analysis	First Draft Report Expected	Final Report Expected
Interim – PEs	September – October 2020 (Desk Review and Secondary data from July 2019 – October 2020)	November 2020 – January 2021	March 2021	May 2021
Interim – DMA/Kiosk	Time Series from September 2019 – December 2020 Household Survey September 2020	November 2020 – January 2021	March 2021	May 2021
Endline - PEs	September – October 2022 (Desk Review and Secondary data from July 2021 – Feb 2022)	November 2022 – January 2023	March 2023	May 2023
Endline – DMA/Kiosk	Time Series from January 2021 – December 2022 Household Survey September 2022	November 2022 – January 2023	March 2023	May 2023

Table 14: Detailed Breakdown of Interim and Endline Activities

Period	Task	Deadline
Interim	Draft Evaluation Design Report Submission	March 8, 2019
	Local Stakeholder and MCC response	March 30, 2019
	Final Evaluation Design Report	June 21, 2019
	Nesstar Metadata Template for Evaluation Catalog Entry	July 15, 2019
	Draft English Questionnaires	April 1, 2020
	Draft TOR for Data Collection Firm	April 1, 2020
	Summary of pre-test/Review of Back translation if relevant	June 1, 2020
	Final English and local language questionnaires and enumerator training manuals	August 1, 2020
	IRB Package	July 15, 2020
	SOW Trip Report for Interim Data Collection	August 1, 2020 September 15, 2020
	Summary of pilot test (DMA/Kiosk)	September 15, 2020
	Data Collection	September-October, 2020
	Minutes of meetings with data collection firm (DMA/Kiosk)	November 15, 2020
	Written Summary of Quality Control Checks (DMA/Kiosk)	November 30, 2020
	Draft Evaluation Report	March 1, 2021
	Final Interim Evaluation Report	May 1, 2021
	SOW Trip Report Results Dissemination	April 1, 2021 May 30, 2021
Endline	Revise Evaluation Design Report (if needed)	February 1, 2022
	Data Collection Firm TOR	April 1, 2022
	Nesstar Metadata Template for Evaluation Catalog Entry	May 15, 2022
	Update/Revise English Questionnaires and Enumerator Manuals (if needed)	April 1, 2022

Period	Task	Deadline
Endline	Summary of pre-test/Review of Back translation	June 1, 2022
	Final English and local language questionnaires and enumerator training manuals	August 1, 2022
	IRB Package Amendment	July 15, 2022
	SOW Trip Report for Data Collection	August 1, 2022 September 15, 2022
	Summary of pilot test	September 15, 2022
	Data Collection	September-October, 2022
	Minutes of meetings with data collection firm	November 15, 2022
	Written Summary of Quality Control Checks	November 30, 2022
	Draft Evaluation Report	March 1, 2023
	Final Evaluation Report	May 1, 2023
	SOW Trip Report Results Dissemination	April 1, 2023 May 30, 2023

5 REFERENCES

AARC Consultancy. Regulatory Strengthening and Tariff Development Support Inception Report. AARC Consulting, 2017, Regulatory Strengthening and Tariff Development Support Inception Report.

Adam Smith International. Guma Valley Water Company (GVWC) Institutional Strengthening, Urban WASH Sector Coordination and District Metering Area and Water Kiosk Demonstration Pilot. ASI.

Adam Smith International. Knowledge, Attitudes, and Practice (KAP) Survey Report. ASI, 2018.

Adam Smith International. *State of Customer Service, Stakeholder Engagement and Grievance Management in GVWC*. ASI, 2018, *State of Customer Service, Stakeholder Engagement and Grievance Management in GVWC*.

Andres,L., Chellaraj, G., Dasgupta, B., Grabinsky, J., and Joseph, G. (2018a): An Evaluation of the Contributing Factors of Water Scheme Failures in Nigeria, World Bank Working Paper Series, No. WPS 8317.

Andres, L. Grabinsky, J. Dasgupta, B., Chellaraj, G. and Joseph, G. (2018b). "An Empirical Evaluation of the Causes of Water Points Non-Functionality in Nigeria." Global Water Programs GP, World Bank, Washington DC.

Ashley-Edison International. "Voltage in Sierra Leone." *Voltage in Sierra Leone- Electricity Supply and Power Quality Overview*, www.ashleyedisonuk.com/voltage-sierra-leone-907/.

Bacon, Robert 2018. Taking Stock of the Impact of Power Utility Reform in Developing Countries A Literature Review, Policy Research Working Paper No. 8460, World Bank, Washington DC.

Besant-Jones, John E. 2006. *Reforming power markets in developing countries: what have we learned?* Energy and Mining Sector Board discussion paper; no. 19. Washington, DC: World Bank.

Bhatia, B. and M. Gulati. 2004. "Reforming the Power Sector: Controlling Electricity Theft and Improving Revenue." The World Bank, Washington, D.C.

Diego, Fernandez, et al. *AquaRating Audit Report of Guma Valley Water Company - Sierra Leone (GVWC)*. Kiwa, 2016, *AquaRating Audit Report of Guma Valley Water Company - Sierra Leone (GVWC)*.

Effective Practice and Organisation of Care (EPOC). Interrupted time series (its) analyses. EPOC Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services, 2015.

Howard, Marjorie. "The Other Cost of Ebola." TuftsNow. 20 January 2015.

Izaguirre, Ada Karina. "Private Participation in the Electricity Sector—Recent Trends." *Worldbank.org*, World Bank, 1998, siteresources.worldbank.org/EXTFINANCIALSECTOR/Resources/282884-1303327122200/154izagu.pdf.

Jamash, T., Mota, R., Newbery, D., & Pollitt, M. (2004). *Electricity sector reform in developing countries: A survey of empirical evidence on determinants and performance.* Cambridge Working Papers in Economics, CWPE 0439. Department of Applied Economics, Cambridge.

Kontopantelis E, Doran T, Springate DA, Buchan I, and Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. British medical journal, 350:h2750, 2015.

Linden A, (2015): Conducting interrupted time-series analysis for single- and multiple-group comparisons, The Stata Journal (2015)15, Number 2, pp. 480–500.

Linden, A., and J. L. Adams. 2011. Applying a propensity-score based weighting model to interrupted time series data: Improving causal inference in program evaluation. Journal of Evaluation in Clinical Practice17: 1231–1238. Millennium Challenge Coordinating Unit, et al. Electricity Sector Reform Roadmap (2017-2030). MCC, 2017.

Millennium Challenge Coordinating Unit. Annex A- Freetown Water Use. MCCU.

Millennium Challenge Coordinating Unit. Bidding Document: Pipe Rehabilitation and Kiosk Installation at Aberdeen and Kingtom DMAs. MCCU

Millennium Challenge Coordinating Unit. *Monitoring and Evaluation Plan - Version 3*. MCCU, 2018, *Monitoring and Evaluation Plan - Version 3*.

Millennium Challenge Corporation. “Beneficiary Analysis.” *Beneficiary Analysis*, Millennium Challenge Corporation, www.mcc.gov/our-impact/beneficiary-analysis.

Mota, R. 2003. ‘The Restructuring and Privatization of Electricity Distribution and Supply Businesses in Brazil: A Social Cost-Benefit Analysis.” Cambridge Working Papers in Economics No. 0309.

Mumssen, Yogita; Saltiel, Gustavo; Kingdom, Bill. 2018. Aligning Institutions and Incentives for Sustainable Water Supply and Sanitation Services. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/29795 License: CC BY 3.0 IGO.

Niedinger, Jennifer. “Mapping Freetown’s Water Pipes to Improve Service Delivery.” *Dipnote: U.S. Department of State Official Blog.* 22 March 2018.

Northwestern University. “Household Water Insecurity Experiences (HWISE) Scale.” LIBRARIES Blog, © 2018 Northwestern University.

OECD (2015), The Governance of Water Regulators, OECD Studies on Water, OECD Publishing, Paris,https://doi.org/10.1787/9789264231092-en.

Pombo C. and R. Taborda. 2006. "Performance and Efficiency in Colombia's Power Distribution System: Effects of the 1994 Reform." *Energy Economics* 28(3): 339-369.

Prasad, Naren. *Social Policies and Water Sector Reform*. United Nations Research Institute for Social Development , 2007, *Social Policies and Water Sector Reform*.

Penfold RB and Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Academic Pediatric Association, 13(6): S38-S44, 2013.

Schwartz, Klaas. “The New Public Management: The Future for Reforms in the African Water Supply and Sanitation Sector?” *Science Direct*, 2008.

Sierra Leone Constraints Analysis Report: A Diagnostic Study of the Sierra Leone Economy; Identifying Binding Constraints to Private Investments and Broad-Based Growth. GoSL, 2013.

Shadish WR, Cook TD, and Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Houghton, Miin and Company, 2002.

Simonton, D. K. 1977a. Cross-sectional time-series experiments: Some suggested statistical analyses. Psychological Bulletin 84: 489–502.SMEC. *International Inception and Initial Findings Report*. SMEC and MCCU, 2017, *International Inception and Initial Findings Report*.

Taljaard M, McKenzie JE, Ramsay CR, and Grimshaw JM. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. Implementation Science, 9(1):1, 2014.

The African Development Bank Group Chief Economist Complex (2016). *The Africa Infrastructure Development Index 2016*.

Thomas, Abdul Rashid. “Sierra Leone electricity supplier under corruption investigation” *Sierra Leone Telegraph.* 4 September 2016.

USAID. “POWER AFRICA IN SIERRA LEONE Fact Sheet.” *Archive - U.S. Agency for International Development*, 2012-2017.usaid.gov/powerafrica/sierraleone.

WHO/UNICEF JMP. Rural and urban drinking water service levels (2015), https://washdata.org/.

World Bank. “Access to Electricity (percent of Population).” *World Bank, Sustainable Energy for All (SE4ALL) Database from the SE4ALL Global Tracking Framework Led Jointly by the World Bank, International Energy Agency, and the Energy Sector Management Assistance Program.*, 2016, data.worldbank.org/indicator/EG.ELC.ACCS.ZS.

Zhang, YF., Parker, D. & Kirkpatrick, C. J Regul Econ (2008) 33: 159. https://doi.org/10.1007/s11149-007-9039-7

6 ANNEXES

6.1 Annex 1: Process mapping of Tariff setting process

6.1.1 Step 1: Process Overview

The financial health of a generation company depends critically on the being able to sell its power at a price that is above cost. At the system level, EGTC's suppliers include creditors, the National Petroleum Company, a privately-owned fuel supplier, and parts manufacturers and suppliers that provide spare parts. From buyer's side, EGTC's customers are wholesale electricity buyers which at this time is only EDSA. EWRC is the regulator and is one of the key player in the tariff setting process. EGTC owns and operates two fossil fuel power plants located in the capital area, Kingtom and Blackhall road, and it owns but does not operate the Bumbuna hydro power plant. Kingtom and Blackhall road are fully operational during the dry season (January – July) and then are shutdown or operate occasionally from August through December. This particular section describes the current situation explaining why this process is important in the wider system. Quantifies the cost, service, compliance and / or quality issues that currently exist.

Since 2015 EGTC has not had a clear tariff set for the sale of electricity to EDSA and EDSA disputes the prices EGTC wants to charge. As a result, EDSA owes EGTC over US$ 10 million and EGTC is not able to meet all its financial obligations. EGTC's current tariff of US$ 0.10/kWh does not fully cover operating costs. It is incumbent upon EGTC to submit a thorough tariff application to EWRC as part of its plan for financial health. Given this regulatory setup, the process mapping will first look into the following areas for a complete overview of the process. In the process overview stage, the ET will review the policies and procedures providing guidance about completion of the process. In addition, the team will consult OPM, Instiglio and other relevant parties involved in RBF related works to find out if any intermediate steps are missing before finalizing the detailed outline for process mapping of tariff setting in the electricity sector.

6.1.1.1 Policies and procedures providing guidance about completion of the process

- What types of general and technical regulations are involved in the process?
- How are the regulations set? What are the guiding policies and procedures?
- What is EWRC's role and responsibility in the process? Who from EWRC is responsible for this task?
- Are representatives from EGTC and EDSA part of the regulation setting process? If yes, who represents EGTC and EDSA?
- Who else (external agencies) is involved in deciding these policies and guidelines? What is their role?
- Are these policies susceptible to changes in external factors?

- Do current GoSL accounting requirements support EWRC tariff making accounting requirements?
- To what extent do political, financial and other pressures play into tariff setting and therefore, EGTC's ability to obtain cost-recovery tariffs?
- Will the tariff mechanism include automatic pass-through items?

6.1.2 Step 2: Process Boundaries

In step 2, the ET will identify and compile an outline of the sequence of steps along with identification of time and resource requirements to complete the process. In the overall process, this particular step helps us conduct a gap analysis by mapping the resource requirement, availability and gaps at each node of the process in terms of finance and personnel.

The tariff process includes a detailed accounting of all costs using utility accounting methods and following the chart of accounts. From our scoping trip and review of documents we understand that an application is made to the EWRC to justify the rate and is based upon a guaranteed rate of return. In terms of generation, EGTC includes, besides its own generation, the private company that operates the Bumbuna hydro plant. EGTC also owns and operates the 131 kv transmission system. EGTC owns the system control center but still has not taken over operation from EDSA. Together EDSA and EGTC still have common assets that have not yet been transferred to their owners. These common assets set challenges to identify the full cost accounting and recovery that have not yet been fully addressed.

"To trigger a tariff review process, EGTC will need to provide the information set out in nine forms to EWRC. The forms have been developed by the Commission and are available to the Utilities upon request. The 9 forms require information on the proposed tariff, the proposed revenue requirement of the utility in the tariff year, historical investments, capital expenditure plan, debt obligations, electricity or water purchases, O&M expenses, billing and operational efficiency, and technical and operational details."[81]

Based on our scoping trip and survey of documents[82], we understand that the tariff process includes a detailed accounting of all costs using utility accounting methods and may follow a regulatory chart of accounts. An application is made to the EWRC to justify the rate and is based upon a guaranteed rate of return. The following questions will be explored prior to and during the tariff setting process.

6.1.2.1 Sequence of steps to complete the process

a. Prepare tariff application.

- What is the current state of asset transfer between EGTC and EDSA?
- For commonly held assets, has an allocation method been determined and approved by EWRC?

[81] Instiglio (2018), Final Results-Based Financing Design Report, page 64.

[82] SI has repeatedly requested documents that have not yet been provided (eg., work plans) that would have allowed for a more detailed description.

- Does EGTC currently have all the necessary financial data? What is the data quality?
- Following the chart of accounts, what specific costs are included in utility accounting? Who is responsible for preparing it? How is justification of rate setting made?
- Does it require a priori estimation of guaranteed rate of return?
- What is the method of estimation? Who decides the method (the utility or the regulator?)
- If regulator, is there any guideline available with and provided by EWRC to EGTC?
- What are the potential risks (internal and external) that can influence estimated rate?
- Organized tender for generation. How is organized tender process for generation taken to the market prepared and published?

b. Launch of tariff review process.

- How is initial regulatory functions for tariff review process launched? Was it launched in scheduled time? When was it launched?
- How are plans for tariff review process, timeline, subsidies and financial exposure made? When were they made? Were they made in scheduled time? If not then why not?
- Who (agencies and staff) are involved in these planning process? What is their role in the planning process? Are they trained processional? Does the agency has sufficient resources and personnel to undertake this task?
- What are the nine-forms required to be filled by EGTC? Do information required for these forms to be filled out come directly from EGTC records, or need some specialization to prepare these information? If so, then does EGTC have sufficient resources (person power and financial) to complete these forms?

c. Decision on tariff rate.

- How is tariff rate decided? Who are involved in the process? What does the process require for rate justification?
- Is willingness to pay/affordability assessed to decide tariff rate? Who does it and how does it influence the tariff setting process?
- Does the methodology allow for automatic pass through items like fuel costs?
- What other costs does it take into account for tariff calculation?
- Is past financial deficit incorporated/adjected in tariff setting?
- How often can the tariff be updated?

6.1.2.2 Individuals, resources, and time required to complete the process

a. Capacity assessment.

- How much time does it take from start to finish of the tariff setting process? Does EGTC and EWRC has sufficient financial and personnel capacity to complete it within time? If not, then how was it affected by lack financial and personnel capacity of the EWRC and EGTC?
- Does the process involve any assessment of capacity in terms of personnel and resources, timeline to complete the process?
- How are subsidies and financial exposure made to complete the process? Who are involved in this assessment?

- Does the agency have sufficient resource and personnel to complete the tariff setting process in timely manner?

b. Capacity building.

- How is Tariff and Accounting/Finance training needs assessment made?
- What are the steps taken to step up capacity to match demand for tariff setting process in timely manner?
- What are the steps taken to step up capacity to match demand for tender process in timely manner?

6.1.3 Step 3. Gap identification

Step 3 provides information about human and financial resourcing challenges, gaps in coordination of individuals and resources involved in the process, and governance and management challenges. The gap analysis based on information on resource gap at each node along with governance and management challenges provides useful information about at what exact node in the process chain the problem is occurring or the chain is broken.

6.1.3.1 Gap in capacity of and coordination of individuals and resources

- How are financial and personnel managed and allocated in tariff setting process?
- What is the existing gap between availability and requirement of staff and financial resources for tariff setting process?
- In which way did capacity constraints in terms of resource and personnel affect (quality and quantity) of setting and generation and tariff?

6.1.3.2 Governance & management challenges.

There are challenges that face EGTC and cost recovery. For example, past tariffs have not been reflective of full costs and EGTC is operating at a loss. Will future tariffs attempt to recover this or will the GoSL subsidizes these losses. Similarly, together EDSA and EGTC may still have common assets that have not yet been transferred to their respective owners. According to the SMEC Business Plan, as a result, EGTC does not have legal ownership of some of its key assets. These common assets set challenges to full cost accounting and recovery that have not yet been fully addressed. For example, EGTC owns the system control center but still has not taken over operation from EDSA. Based on this information, the process mapping will explore

- What are the governance and management challenges envisioned before the start?
- What are the governance and management challenges actually faced?
- Audit / Cost of Service Study / Complete financial statements

- Did EGTC file new tariff application on the basis of complete financial records and asset transfer? If not, why not? How does it affect EGTC tariff, cost recovery, and financial performance?
- How will past financial deficits be addressed in the tariff methodology?

6.1.4 Step 4. Follow up Actions

Step 4 will detail the actions taken to overcome the challenges, changes in these factors over time, and the potential role of the THP in the process. This section will help us identifying the follow up actions taken by and recommended for each involved parties.

- What steps are taken by EGTC, EDSA or EWRC to bridge the gap?
- Is utility performance monitored and if so, how is it linked to gap in tariff setting process?
- Changes in these factors over time and Role of THP in the Process?

6.2 Annex 2: Stakeholder Comments and Evaluator Responses

6.2.1 MCA Comments and Evaluator Responses

Date	Reviewer Name/ Institution	Page Number	MCC Comments	SI Response
06/10/2019	M&E	21	There is mention of the Baseline Memo- include a copy of the memo in the EDR Annexes.	We will add it in the Annex
06/10/2019	M&E	24	Data collected by Instiglio will contribute to the evaluation at midline but please clarify what information will be collected related to the RBF after the THP ends (at endline).	SI will follow the same structure as Instiglio will use for midline and collect the same information in endline. This decision makes it important that Instiglio and SI prepare the questions together for interview so that all information needed for evaluation are also collected even in midline. The text is updated.

06/10/2019	M&E	27	Observations are proposed as a strategy for understanding certain aspects of performance. Please explain further how this method will be used specifically to understand the capacity building that is provided to utilities. Also, I'm not sure about the practicalities of unannounced observations.	SI has written "All direct observations will use standardized tools and, where observation by SI might bias people's actions, visits will be unannounced and observers will remain as unobtrusive as possible." SI proposes to attend meetings (if agreed upon), training sessions with focus towards capacity building. SI's local consultants (Senior Analyst, water sector and Data specialist) are based in Freetown and as per our discussion with MCCU, they can attend such trainings. The unannounced appearance in meetings was proposed to avoid any influence on the meeting agenda or responses from the parties in the meeting. However, if that is not feasible or SI needs to be there with prior information, SI is open to it as well.
06/10/2019	M&E	28	It is proposed that a process map will be done for RBF. Please explain further what aspects of it will be mapped. Will this just examine the design and implementation at a high level or will it delve further into specific processes related to outcomes that were targeted by the RBF?	The text is revised with an example of what the process map will include. For clarification, it will not just examine the design and implementation rather it will delve deeper to identify gaps in the process in terms of resource availability, capacity and management issues involved in the process.
06/10/2019	M&E	38	Please revise this statement. "Case study methodology was rejected in favor of the household level analysis." My recollection is that we didn't outright reject this approach and were open to case studies if it's appropriate.	Revised text.

06/10/2019	M&E	40	Footnote 67: "The DMA Bidding Document notes the provision of water testing kits by the supplier but does not include the location or frequency of testing. If testing does not prove applicable to the DMA/Kiosk evaluation SI would propose collection point testing using DelAgua or WagTech kits, both of which are designed to give WHO risk category results." Please clarify what is being suggested here.	Based on the ASI document shared 4/2 on water quality testing planned for DMAs, the text/footnote is revised. Through discussion with MCC, SI will decide the next steps.
06/10/2019	M&E	41	For the Standpipe management/Kiosk activity what data collection will be done with standpipe managers? I see mention of KIIs with the entrepreneurs in relation to Revenues for GVWC but I think there are other outcomes that info from the standpipe management should inform ex. how exactly do they operate, how much are they charging, service reliability, etc. There are references to the indicator of Leone billed to entrepreneurs but nothing about how much GVWC collects. That is important.	Point well noted. SI intends to collect this information as well as a part of the secondary data. We will clarify it in the revised text.
06/10/2019	M&E	42	"For the comparative case studies…" This is unclear. Please explain what case studies this is referring to.	This should be comparative study – not comparative case study. Text revised.
06/10/2019	M&E	44	"Retrieval and analysis of this data will continue for at least years…" I think a word is missing.	Text revised (now in pp 45)

06/10/2019	M&E	45	"October 2020" this needs to be updated as some activities of the Threshold will go until January 2021.	Text updated and revised
06/10/2019	M&E	47	Please provide additional details or examples of some of the criteria that will determine who will participate in various FGD or which kiosk sites will be selected for observation.	Added in the table
06/10/2019	M&E	57	I think it would be beneficial if the interim report could be done before the end of MCCU's closure period.	Yes. That's our plan.
06/10/2019	M&E	58	Please update the timeline for reporting—The draft evaluation and final report and dissemination should be in 2023.	Noted and revised.
06/10/2019	Human and Community Development (Scott)	8	Is there confusion between an "underlying assumption" and a contractor action item/deliverable? numbers 1, 3, 4, 5, 6, 7 seem to be more along the lines of things that we are tasking the contractor to address – not that we are hoping/assuming that the external environments will "provide".	Yes- we are hoping contractors will address this, however the EE will measure whether or not they were able to do so. If they are not able to do this appropriately a link in the TOC breaks. Revised text based on our response added here.
06/10/2019	HCD	19	Number 3 – comparing the private sector approach with community management approaches – are the community approaches from Freetown or other cities?	We prefer within Freetown. Added text to clarify.

06/10/2019	HCD	22	Table 2… subset questions include data sources not noted in the superset e.g. 10) a) , 11) a) include data sources not noted in 10) and 11) respectively. maybe clean that up. 11)a) and 12) - perhaps include household surveys to inform this as well?	Noted and revised accordingly in the text.
06/10/2019	HCD	25	TABLE 3. Is there value to including representatives in key informant list from local community counterpart organizations like FCC?	Revised text by adding this group in Table-3.
06/10/2019	HCD	27	Is process mapping something that ASI is already doing? Is ASI work in process mapping used in developing the SI process maps?	No. The PM from our evaluation would provide a different value added by examining the capacity and resource gaps at different nodes of a process and how it is going to affect the efficiency of the process depending upon relative importance of the node in the process chain.
06/10/2019	HCD	41	Water citizenship… can we use billing/payment records, consumer reports as indicators - not just customer interviews about attitudes? e.g. reported illegal connections vs total illegal connections discovered.	We proposed to use billing/payment records, consumer reports as indicators in addition to the customer interviews about attitudes. This is integral part of our proposed secondary data analysis. We will clarify in the text if its not clear.
06/10/2019	HCD	42	Indicators… Are we able to look at gender issues related to service issues – long queues? it looks like the HH surveys described in 3.4.4.1 will address this, maybe discuss this? Yes.	Although gender issue is not prime focus of the THP, we will look into gender aspect by analyzing, time to fetch water, waiting time etc., gender violence or conflict at the outlets etc. The HH survey includes questions around gender issues (who fetches water, how long does it take, measures of safety)

06/10/2019	HCD		It sounds like we might be examining use of alternative water sources in the DMA, but can this be discussed more explicitly? is it possible to look at use of different sources over the water year over time?	It is possible and one of our objectives is to analyze trend in use of different sources over time in DMA. We will discuss it explicitly in the evaluation report.
06/10/2019	GSI	8	When listing overarching assumptions of the ToC for WSRP, there are inherent assumptions about behavior change. I think point 6 gets at this perhaps but I wonder if it wouldn't be stronger to explicitly state our assumptions about consumer/HH behavior change itself presents a risk if we are unable to achieve our BC objectives.	Revised text

06/10/2019	MCC/EA – Epley	18 Section 3.1	Comment: *"It does so by focusing on several key **beneficiary** institutions, primarily…"* MCC reserves the term "beneficiaries" to indicate private sector households and business that benefit from its programs (e.g. from improved service delivery). To avoid confusion, please refer to these entities as service providers, implementing partners, recipient institutions or by some other similar moniker.	Thanks. We changed the text accordingly.
06/10/2019	GSI	19	Point 2: glad to see the inclusion of a question looking at how the TA provided affected outcomes in gender and social equity in the DMA and in Point 3 looking at the DMA pilot and how it impacts cost of water to households.	Thanks.
06/10/2019	GSI	19	Also glad to see they hope to lift out the replicable lessons learned from the DMA experience.	Thanks again.
06/10/2019	GSI	20	I would encourage SI to delve deeply into the assessments of technical assistance provided at the public distribution point level. I think there might be a wealth of institutional learning that can be culled from the DMA pilot, especially with regards to preparing and managing private sector provision of a public good (water), in terms of quality of customer service, operational management as well as ultimate cost for households.	Point well noted. We can expand further on KIIs/Direct Observations to collect/information data from kiosks

06/10/2019	GSI	38	While I do understand the limits to HH level analysis due to the limits of the study, I do hope that SI is able to design the proposed comparative analysis in such a way that there is a rigorous and thorough study of the kiosk model and its impact and level of satisfaction on households. This analysis should look at the effectiveness of any training offered to kiosk operators, especially with regards to customer access and satisfaction.	Point well noted. We will include visit to training program in our Direct observation and KIIs to supplement our HH level analysis (for satisfaction and customer level access).
06/10/2019	GSI	50	As a GBV situational analysis has been conducted and learnings from the findings used to influence operational considerations like Kiosk design and operations, I wonder if SI plans to include an analysis of this in their Kiosk study. While the GBV indicator is not on the ITT, we did spend threshold resources on this situational analysis. If SI does intend to somehow address this issue that has been correlated in SL to water service delivery, I wonder how they plan to collect information regarding this- through FGDs, or key informant interviews? In HH level questionnaires? If the latter, IRB considerations might need to include speaking to vulnerable populations like GBV victims or children.	Although GVB/Gender analysis is not prime focus of the THP, SI intend to capture GBV through HH survey, KII and FGDs. SI is also planning to get approval from local and SI's own IRB approval before conducting the survey.

06/10/2019	M&E	20	In Sections 3.2, 3.3.1, and 3.4.1, the analytical approach should be clearly stated. We categorize our evaluations by methodology and that's one reason why we ask evaluators to name the methodology. We consider analytical methods to be separate from data collection methods, and the former is somewhat missing (it's there throughout the sections, but we need a clear statement at the start).	Revised text
06/10/2019			I recognize that the PE employs a variety of methods, but I think the core approach for identifying results is a pre-post comparison. Whether it's the qualitative work or the process mapping, my understanding is that you will be contrasting the situation before and after the interventions (to the best of your ability, since you've noted that we've missed the time for a baseline).	This is correct. In absence of a baseline, SI will do a pre-post (THP intervention) analysis to capture sustainability. That will be established through our interim and endline data collection.
06/10/2019			Related to the DMA study, is the time series analysis envisioned to be different from an interrupted time series analysis?	At the system level the time series analysis will be interrupted time series to capture the before and after of the completion of the THP. Revised text in page 21.

06/10/2019			We consider the latter to be an impact evaluation method because, with enough pre-intervention data, you're essentially establishing the counterfactual trend.	SI will use interrupted time series analysis as mentioned above and proposes to use sensor level time series data for quality, reliability and NRW etc. For information on billing, and collection SI will use GVWC data to construct pre-intervention data. For analysis, the completion of the THP intervention will be used as cut-off point for before and after.
06/10/2019	M&E	22	Table 2. I believe this table is responding to the table requested in section 3b of the EDR template. However, it's missing the key outcomes associated with each evaluation question. The outcomes were more clearly outlined for the DMA evaluation, but I would like to see more detail on the key measures that will be the focus of the PE.	Added text in section 3.3.1
06/10/2019	M&E	29	The justification for the timing of interim data collection is articulated, but I don't see a justification for the endline, i.e. why is 2 years the right timing?	Our main goal is to provide sufficient time to capture evidence to address questions around sustainability (after intervention is completed). Sustainability is a long term concept and SI believes a gap of 1 year may not be sufficient to address the evaluation questions around sustainability.

06/10/2019	M&E	42	Is there evidence that can be cited to justify the proposed exposure periods? i.e. why we think 3 years is the appropriate time to wait to see effects. Ideally this would be discussed as part of the project design (i.e. by MCC/MCCU), but if not, the EDR should ground our proposed timings in the evidence around these types of programs/results.	Three years period is actually considered for two periods to conduct a before and after analysis-- one year during the interim (intervention period) and two years for post intervention period. Two years post intervention is considered to allow sufficient time to respond to the sustainability questions.
06/10/2019	MCCU-Urban WASH, Social and Community Development Consultant	General	I only have one comment on the proposed methodology - it would be useful to know a bit more on the purpose and process for the process mapping described under 3.4.1.6. I am not sure I understood what was described in the report, it seems the intent is to assess whether GVWC and ASI understood the technical specifications and processes outlined in the bidding documents? However, these specific processes are especially for DMA works which is different to the process for NRW implementation.	A detail description is added in Section 3.3.1.5 under process mapping which is also supplement description of process mapping under section 3.4.1.6. The proposed PM is not to assess whether GVWC and ASI understood the technical specifications and processes outlined in the bidding documents. Rather, it will inform the evaluation about gaps in a particular process in focus such as water service delivery process through DMAs.

06/10/2019	MCCU-Urban WASH, Social and Community Development Consultant	General	GVWC has rescheduled their rationing schedule early this year, this has had a major affect on water supply in Aberdeen and Kingdom, resulting in water being supplied at night and less often. Therefore, the service in these areas has reduced. The rescheduling of the rationing schedule isn't a result of the DMA pilot, it was an initiative taken on by GVWC in an attempt to more equitably share water throughout the city. This may have a impact of the results of the impact evaluation of the pilot.	This is a very important piece of information and is going to have its effect on DMA performance. SI will take proper care during assessment of impact and will discuss with relevant stakeholders to gather information during the process through KIIs, consultation of available documents.

06/10/2019	MCCU-Urban WASH, Social and Community Development Consultant	General	3.4.1.2, Time Series Analysis of Administrative and Sensor Data: We are currently planning the establishment of a control room in GVWC - the bidder documents have some simple items included in this such as computer and tables, etc. However, since the bidding document have been developed the scope of the control room has increased due to GVWC interest. The current idea is that the control room will be where data from the DMA (NRW, billing, etc.), transmission, production, service reservoirs, etc. will be sent and monitored. This might be useful for SI to know, as it could provide the information required by SI and depending on the additional scope to the control room it might affect how SI can analysis sensor data over time. The timing of data becoming available from to DMA relative to the DMA works may also determine if the data can be used as pre-intervention baseline. If SI want more information on the control room we can follow up with ASI.	This is a great news. We followed up with MCC and came to know that this may also include information on the comparison DMA. We will follow up with MCC, MCCU and other relevant stakeholders involved into this.
06/10/2019				
06/10/2019		Process Mapping	Ensure distinction between process mapping and process evaluation - How does PM feed into MCC's understanding of results? What is the value added?	Text is revised under section 3.3.1.5. to explain it in detail.

06/10/2019		PM - RBF	What is the focus on this? Are there two levels (GVWC and how they are meeting targets or the RBF process as a whole)? How does the design of the process (which was a great challenge) factor into the PM? Would be interesting to see how MCC had to insert itself into the process and that effect.	The focus here is to identify gaps in a particular process, their ranking in terms of priority and if any measures have been taken to overcome the challenges. It will also look into the roles that MCC had to play into the process. A detail description is added in the text under section 3.3.1.5.
06/10/2019		Timeline	PE interim data collection could occur in June 2020, prior to the DMA evaluation due to the sequencing of activities.	SI adjusted the evaluation timeline based on our discussion with MCC. The extension for the THP will end on January 31, 2021 with a 3-month closure period (Feb – Apr 2021).
06/10/2019		Twinning Program	SI to incorporate/mention the forthcoming twinning program and how that information will be used in the evaluation.	At present SI does not have sufficient information on twining program to design its evaluation plan. What we came to know is that GVWC and EGTC are partnering with counterparts in Ghana to learn more about their roles and better management of their systems/institutions. MCCU is planning to do a special study on the twinning exercise, likely with support from an MCC fellow. This will be an internal assessment of the program (deep dive) while SI will be responsible for looking at the twinning exercise in the broader context of capacity building/THP. As per our understanding from discussion, research plan is likely to be developed/available in September for the special study. Given that this EDR is a live document, SI will outline a plan how to incorporate it the evaluation once the research plan is available.

06/10/2019		APS Activity	Include/note the APS activity for DMA kiosk sensors(?)	Discussion is included in report text; we are open to adding additional detail following Discussion with Univ. colorado and MCC's final review.
06/10/2019		Standpipe Management	Is there a process in here we can map? What data collection will come from standpipe managers?	This process can also be mapped. Due to limited time and resources the ET needed to be selective. However, based on further discussion with MCC we can replace any process we mentioned in the list by this.
06/10/2019		NRW vs water quality	NRW is of more importance for MCC but would be interested in chlorine testing if available	This is an area that depends on feasibility and relative importance. That said SI is proposing NRW since its one of the main objectives. SI is also proposing water quality testing based on sensor data. However, SI is also interested to conduct chlorine testing with further consultation with MCC and other stakeholders .
06/10/2019		HH Survey	MCC not clear on the value added for sampling in the comparison DMA	The ET plans to collect HH level data from comparison group to examine what would have happened to the HH from treatment DMAs if the intervention was not there. In other words, this will help create a comparison group to represent the treatment households without intervention. The main value addition of collecting HH data from the comparison DMAs is to tease out change that can be attributed to the program
06/10/2019		HH Survey	What is the value added in addition to what we can get through system data?	System data will provide us the change in service delivery over time. It is sufficient to measure the benefits of improved service delivery before and after the intervention.

06/10/2019		Comparison Sensor Data	Permissions/Feasibility of collecting this information with GVWC?	Yes, permission from GVWC needs to be taken although SI is proposing sensors that are not invasive to the GVWC property like pipes etc. We propose temporary sensors attachable to the outlets. However, even for that SI's plan is to take prior permission from GWVC.
06/10/2019		Aggregation at the EA Level	If it proves not feasible to track BL households how does aggregating up at the EA level (with a random sample each round) impact the evaluation?	That's correct. ET's plan is to randomly sample the same number of households from selected EAs in BL in each round of data collection.
06/10/2019		Budget	Update budget	Budget has been updated and will be sent to MCC externally.
6/24/2019	MCC M&E	18	By “sensors” do you mean “meters”? Though not all sensors are meters I just want to make sure what’s intended is clear in each use of this word.	Yes, we meant ultrasonic flow meters. This is based on our understanding that domestic meters that will be installed at all current GVWC household hook-ups are ultrasonic flow meters. Also, these meters will be used on sub main pipelines within the DMA and in guard boxes on constructed kiosks.
6/24/2019	MCC M&E	22	There is a comment above on row 37 that indicates that MCC considers ITSA as an impact evaluation so specifically, this document should specify that this is a quasi-experimental design for the DMA. This could also be articulated in section 3.1/3.2.	Thanks for this comment. Interrupted Time Series (ITSA) is a robust quasi-experimental design with the ability to infer the effectiveness of an intervention that accounts for data dependency. It offers a rigorous methodology to determine the effectiveness of complex interventions on outcomes in real world settings. We revised the text by including references documenting ITSA as a quasi-experimental design. The references are included below as well.

6/24/2019	MCC M&E	26 Direct Observation	Direct observation should only be used when it is practical and will provide added value in learning that cannot be gained from the process mapping or qualitative/quantitative data“. So as far as “a) the performance of technical assistance and capacity building to water and power utilities, b) meetings of the sector specific steering committees (should access be feasible), and c) sector-specific operations” I question the appropriateness of this method for a and c. Also considering the proposed timeline a lot of these activities may already be completed by late 2020.	To clarify, the proposal of direct observation (sometimes also referred to as structure observation) is to have a standardized tool used during site visits and observations of project activities or outputs that can help verify information through observation. For example, DO was proposed for observing generation and substation operations and verification of improved O&M practices. To be clear, we are not proposing to directly observe the technical assistance ‘process’. Instead we are proposing to observe the tangible outputs of that process. We agree that it should be used when it is practical, and we continue to believe it is valid for situations (a) and (c). If MCC does not agree, can you please provide additional information on why it is not believed to be appropriate in these cases? On the second point, SI acknowledges that at the point of data collection many of the project activities may be completed, but those are not the primary target of our proposed direct observation. It is also true that during the period of field data collection, some of the process or operations may not be available to be observed, but this will need to be assessed closer to the time of field work.

6/24/2019	MCC M&E	Budget	On the budget, is it correct to say that the data collection (KIIs, FGD, process mapping) for the performance evaluations will all be done by Social Impact team members or the two local consultants in Sierra Leone? Please clarify. With the total data collection plug included under the DMA evaluation can you please clarify if any of the plug will be used to pay for any data collection for the WSRP and ESRP performance evaluation.	It is correct to say that the data collection (KIIs, FGD, process mapping) for the performance evaluations will all be done by Social Impact team members or the two local consultants in Sierra Leone. No survey firm will be contracted by SI for the PE. SI believes that that way we can have a grip over the quality of the data/information. Since MCCU is present, SI's understanding is that the Household survey firm will be contracted by MCCU. SI would however, like to propose a part of the data collection plug in to use for the qualitative components related to DMA evaluation (such as FGDs). The reason for this proposal is to maintain an uniformity of data quality since some of the KIIs and FGD may be overlapping.
6/24/2019	MCC M&E	29	Process mapping of RBF still seems very broad as there are a number of indictors implicated and it seems like perhaps there should be a prioritization of which "process" or KPIs are mapped. This also seems very broad and may need to consider what exactly about this will be mapped: "Technical assistance and training in operations and management." We will expect more specificity before data collection begins.	We have added a detailed breakdown of tariff rate review process mapping by steps and added associated questions for illustrative purpose. This is, however, significantly based on information gathered from discussions during the scoping trip and review of relevant documents and literature. We intend to finalize it through further discussion with relevant partners involved in the RBF implementing and independent verification agencies such as Instiglio and OPM.

6/24/2019	MCC M&E	30/43 Time frame of exposure	"SI is proposing interim data collection take place almost two years after the baseline, approximately 9-12 months after the construction and upgrading of major DMA/Kiosk infrastructure." "Should an extension to the THP until early 2021 be granted, …Likewise, if Sierra Leone is granted an MCC Compact targeted at the electricity and water sectors, the endline evaluation may be moved up in order to (1) generate lessons learned and provide data useful in setting Compact goals and (2) avoid contamination from Compact activities." The justification and/or the milestones around which data collection timing would be based should be clarified. We understand there are contractual and project timeline constraints but is there any literature or other experiences that could be cited to support these judgment calls?	1. SI's initial proposal was end of 2019 as the interim data collection period. However, as correctly noted, that due to contractual and project timeline constraints, such as, shifting dates for procurement, implementation completion, and late start of important components, SI proposed late 2020 as the interim data collection time. The idea was to keep some leeway to accommodate further changes. This was discussed regularly during bi-weekly meetings with MCC, and occasional discussions with MCCU. 2. There is another reason that motivated us to prose this shift in data collection phase. The baseline was implemented in mid-2018 and the endline will be conducted in Sept.-Oct. 2022, in order to complete the evaluation by May 2023. Given this schedule, we think Sept. Oct. 2020 (for interim) falls right in the middle and breaks the entire period of performance (of the evaluation) in two equal halves for a before and after analysis. Additionally, by allowing this extra time before the interim will also help some delayed implemented project components to mature before the second phase kicks in. 3. Finally, one of the key evaluation questions is around the effect of the THP on the sustainability of water and electricity outcomes. This will help us understand whether the sustainability

				goals are achieved after the THP was completed.
6/24/2019	MCC M&E	41 Water quality	In footnote 68 there is a plan for water quality testing in the DMA. What's the rationale for the different types of water points proposed and the number? The project kiosks is more immediately apparent but why these other water points.	To a large extent, water quality at the end point depends on the speed/flow of water and types of water points. Since it is expected that water flow may not be the same everywhere due to slope, distance from the source, and other reasons, we proposed water quality testing for different types of water points.

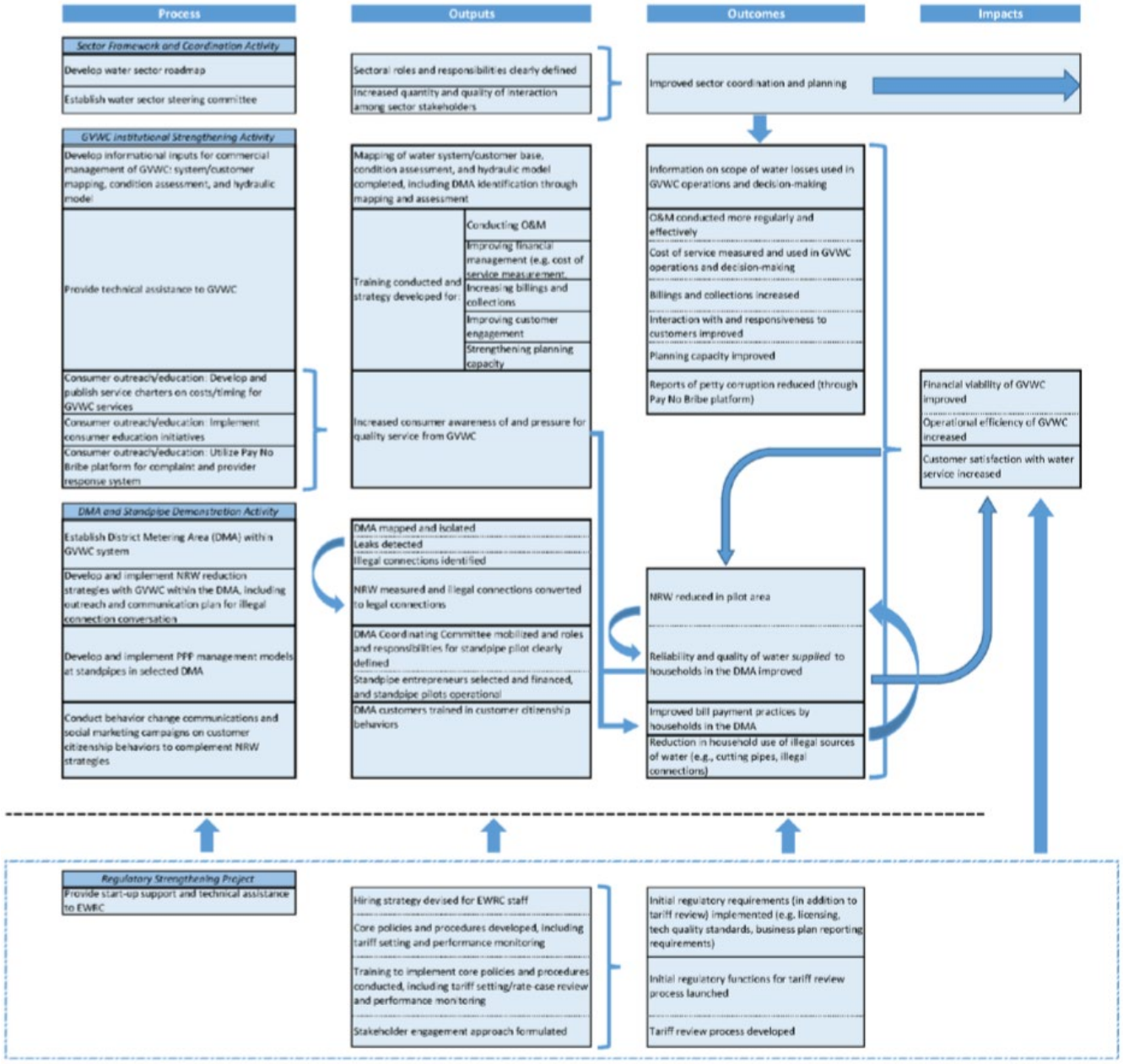
6/27/2019	EA	36-38	SI is proposing an ITSA approach to studying the sustainability of the THP activities. This choice needs to be further explained. Why is the end of the THP an appropriate cutoff? Why are the pros/cons of alternatives (there are several alternatives: completion of works, payout of RBF since RBF activities may affect sustainability, and perhaps other options). It is understood that there may not be an ideal option, but the tradeoffs need to be better understood. One concern about the current choice is that there are system level exogenous shocks that affect the comparison vs. DMA unequally – one obvious example being supply shortages that affect "downstream" DMAs more acutely. The interrupted time-series approach will erroneously pick up these differences. The other concern are that there are additional sources of time-domain variability in outcomes, such as reform efforts that end before the end of the THP, such as come of the RBF activities. The utility may prioritize the DMAs to meet the RBF targets. Additional sources of variation in the cutoff/time series could mitigate these concerns.	Please see our above comment on the relevance of ITSA approach as a quasi-experimental method to measure impact from before and after time series data. We also revised the design report by adding sources from the literature documenting use of this approach. There are several reasons behind not choosing 'Completion of work'. First of all, regular supply starts after the installation is completed. As a result, comparable 'before data' would be hard to find. A cut-off at the end of the THP will allow us to gather sufficient 'before and after' system level data from the same sources. Moreover, ITSA can identify break points statistically derived from the system level time series data. So, even if all or most of the change in outcomes occur in the period between installation and end of THP, ITSA can find a natural break from the system level data and help us compare the two scenarios. Second, two World Bank studies (2018a and 2018b) for Nigeria (same region) show that there is a 20 percent to 30 percent likelihood of failure among all small- and medium-sized schemes within the first couple of years before it peaks at about 45–55 percent when they are 8–10 years of age (World Bank 2018a). The other study for water points shows the similar picture- around 25–30 percent likelihood of failure within the very first year after installation (World Bank 2018b). Given the

				relevance of Nigerian experience in Sierra Leone's context, we decided to propose a break point that covers this period of high risk. The other reason in favor of 'end of THP' as the cut-off is also that it helps us address the sustainability question before and after the THP as a whole. So, it provides a clear change of management regime in terms of personnel and financial capacity and governance. System level exogenous shocks can be a good supporting example (rather than a concern) that strengthens the choice of ITSA. ITSA uses random effect modeling to handle influence of shocks and other unobserved factors (that affect change in slope or trend). Moreover, the policy evaluation literature suggests that ITSA assesses the effect size of an intervention via level change (change in intercept) and trend change (change in slopes). The former identifies the size of intervention's effect, while the latter quantifies the impact of the intervention (EPOC, 2015). Additionally, ITSA can be used without the use of comparison group.

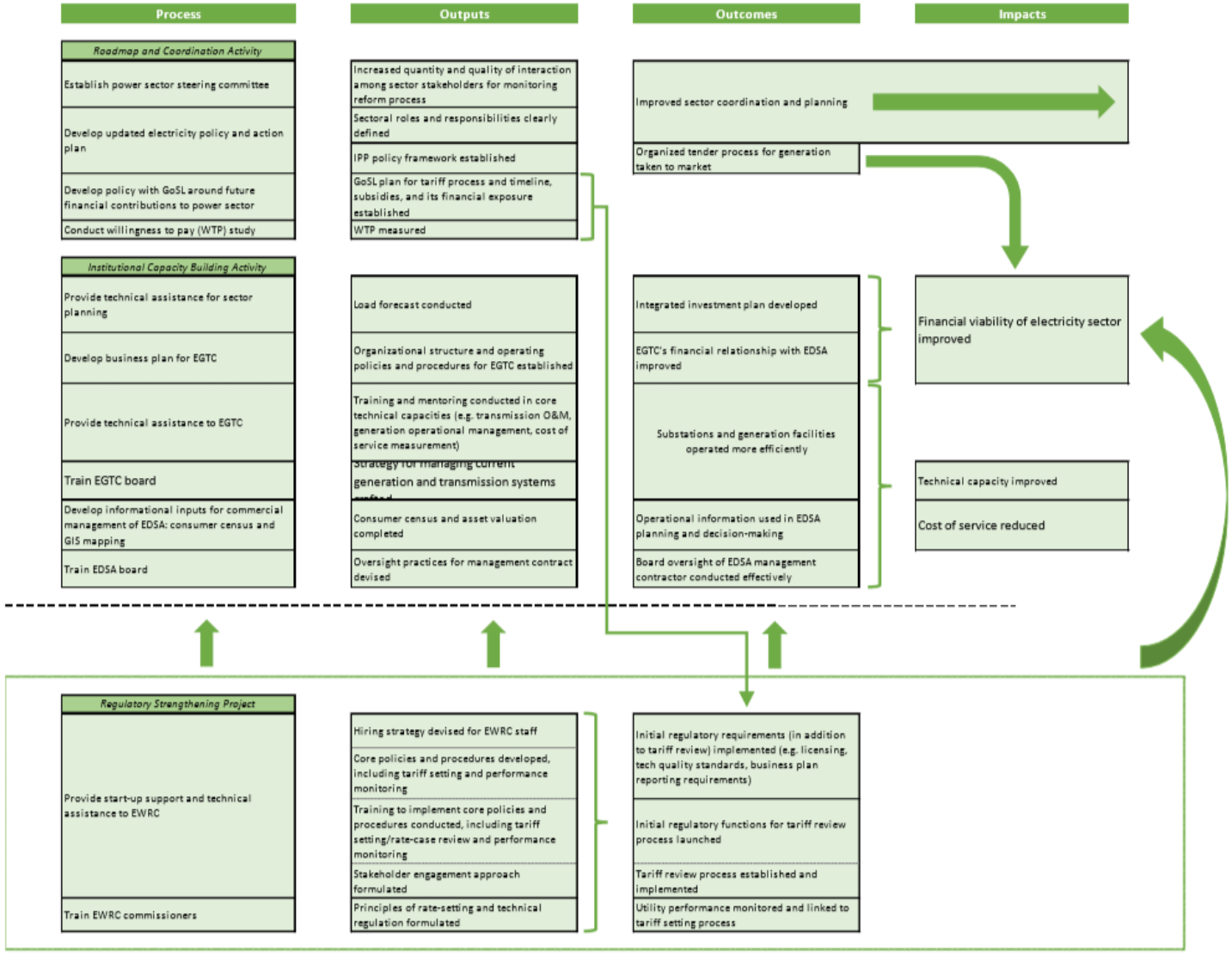
6.3 Annex 3: Evaluation Budget

Per MCC's instructions regarding sensitivities around future procurements, the evaluation budget and budget narrative corresponding to this Evaluation Design Report has been provided to MCC separately.

6.4 Annex 4: WSRP Logic Model



6.5 Annex 5: ESRP Logic Model



ES

6.6 Annex 6: Household Sample Size Calculations

Comparative study to capture changes over time

Scenario 1: Sample Size Calculation with Matching at base line (IE Design)

For binary outcomes, the number of households per arm, assuming a cluster-level analysis, is calculated as[83]

$$n_{initial} = Deff * \left[\frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 + [P_1[1-P_1] + P_2(1-P_2)]}{\delta^2}\right]$$

where:

δ represents the meaningful change (minimum detectable effect size) to be achieved over the time frame ($\delta \neq 0$)

P_1 is the estimated baseline prevalence value

P_2 is the expected ending prevalence value, equal to $P_1 \pm \delta$

$z_{1-\alpha}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−α. For 1−α = 0.95, the corresponding value is $z_{0.95} = 1.64$

$z_{1-\beta}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−β which is also known as analytical power. Here in our case (1−β = 0.80, with the corresponding value of $z_{0.80} = 0.84$.

Deff = 1+(n-1) *ICC -is the estimated design effect of the survey. Note that the sample size calculations are quite sensitive to this parameter since it depends of ICC and number of households per cluster (EAs in this case).

Scenario 2: Sample size calculation for longitudinal study (Without matching – EDR Proposed Design)

Because these indicators are assumed to be proportions (prevalence), the appropriate formula to calculate the required sample size is as follows:

$$n_{initial} = Deff * \left[\frac{z_{1-\alpha}\sqrt{2\bar{P}(1-\bar{P})} + z_{1-\beta}\sqrt{P_1(1-P_1)) + P_2(1-P_2))}}{[\delta]}\right]^2$$

where:

[83] Rutterford, C; A. Copus and S. Eldridge (2015): Methods for sample size determination in cluster randomized trials, International Journal of Epidemiology, 2015, 1–17, Advance Access Publish in July,13 2015.

δ represents the meaningful change (minimum detectable effect size) to be achieved over the time frame ($\delta \neq 0$)

P_1 is the estimated baseline prevalence value

P_2 is the expected ending prevalence value, equal to

$$P_1 \pm \delta$$

$$\bar{P} = \frac{P_1 + P_2}{2}$$

$z_{1-\alpha}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−α. For 1−α = 0.95, the corresponding value is $z_{0.95} = 1.64$

$z_{1-\beta}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−β which is also known as analytical power. Here in our case (1−β = 0.80, with the corresponding value of $z_{0.80} = 0.84$.

Deff = 1+(n-1) *ICC - is the estimated design effect of the survey. Note that the sample size calculations are quite sensitive to this parameter since it depends on ICC and number of households per cluster (EAs in this case).

Table 15: Required sample sizes for different values of expected change overtime (Scenario 1 with matching)

Level of change		delta				Deff	1st comp	2nd comp	1stX2nd comp	Sample per arm	Total sample size
Pipe water in dwelling (%)											
10%	15.80%	1.58%	17.38%	1.64	0.84	1.2	6.1504	0.2766	1.7014	8178	16357
15%	15.80%	2.37%	18.17%	1.64	0.84	1.2	6.1504	0.2817	1.7327	3702	7404
20%	15.80%	3.16%	18.96%	1.64	0.84	1.2	6.1504	0.2867	1.7632	2119	4238
25%	15.80%	3.92%	19.72%	1.64	0.84	1.2	6.1504	0.2913	1.7918	1400	2801
30%	15.80%	4.68%	20.48%	1.64	0.84	1.2	6.1504	0.2959	1.8197	998	1997
40%	15.80%	6.32%	22.12%	1.64	0.84	1.2	6.1504	0.3053	1.8778	564	1128
Pipe water in yard/premise (%)											
10%	40.0%	4.00%	44.00%	1.64	0.84	1.2	6.1504	0.4864	2.9916	2244	4487
15%	40.0%	6.00%	46.00%	1.64	0.84	1.2	6.1504	0.4884	3.0039	1001	2003
20%	40.0%	8.00%	48.00%	1.64	0.84	1.2	6.1504	0.4896	3.0112	565	1129
25%	40.0%	9.92%	49.92%	1.64	0.84	1.2	6.1504	0.4900	3.0137	367	735
30%	40.0%	11.84%	51.84%	1.64	0.84	1.2	6.1504	0.4897	3.0116	258	516
Tap water in yard (%)											
10%	14.8%	1.48%	16.28%	1.64	0.84	1.2	6.1504	0.2624	1.6138	8841	17682
15%	14.8%	2.22%	17.02%	1.64	0.84	1.2	6.1504	0.2673	1.6442	4003	8007
20%	14.8%	2.96%	17.76%	1.64	0.84	1.2	6.1504	0.2722	1.6739	2293	4585
25%	14.8%	3.67%	18.47%	1.64	0.84	1.2	6.1504	0.2767	1.7017	1516	3032
30%	14.8%	4.38%	19.18%	1.64	0.84	1.2	6.1504	0.2811	1.7290	1081	2162
40%	14.8%	5.92%	20.72%	1.64	0.84	1.2	6.1504	0.2904	1.7859	611	1223
Tap water outside yard (%)											
10%	30.19%	3.02%	33.21%	1.64	0.84	1.2	6.1504	0.4326	2.6604	3503	7005
15%	30.19%	4.53%	34.72%	1.64	0.84	1.2	6.1504	0.4374	2.6902	1574	3148
20%	30.19%	6.04%	36.23%	1.64	0.84	1.2	6.1504	0.4418	2.7172	894	1789
25%	30.19%	7.49%	37.68%	1.64	0.84	1.2	6.1504	0.4456	2.7404	587	1173
30%	30.19%	8.94%	39.13%	1.64	0.84	1.2	6.1504	0.4489	2.7611	415	830
40%	30.19%	12.08%	42.27%	1.64	0.84	1.2	6.1504	0.4548	2.7970	230	460
Pipe water in neighbor's yard/premise (%)											
10%	10.8%	1.08%	11.85%	1.64	0.84	1.2	6.1504	0.2005	1.2334	12760	25520
10%	10.8%	1.62%	12.39%	1.64	0.84	1.2	6.1504	0.2046	1.2585	5786	11573
20%	10.8%	2.15%	12.92%	1.64	0.84	1.2	6.1504	0.2086	1.2832	3319	6638
25%	10.8%	2.67%	13.44%	1.64	0.84	1.2	6.1504	0.2124	1.3066	2198	4396
30%	10.8%	3.19%	13.96%	1.64	0.84	1.2	6.1504	0.2162	1.3297	1570	3140
35%	10.8%	3.77%	14.54%	1.64	0.84	1.2	6.1504	0.2204	1.3553	1145	2289
40%	10.8%	4.31%	15.08%	1.64	0.84	1.2	6.1504	0.2241	1.3786	891	1783
50%	10.8%	5.39%	16.16%	1.64	0.84	1.2	6.1504	0.2316	1.4241	589	1179
Packet water (%)											
10%	64.0%	6.40%	70.44%	1.64	0.84	1.2	6.1504	0.4385	2.6969	789	1578
15%	64.0%	9.61%	73.65%	1.64	0.84	1.2	6.1504	0.4244	2.6101	339	679
20%	64.0%	12.81%	76.85%	1.64	0.84	1.2	6.1504	0.4082	2.5106	184	367
25%	64.0%	15.88%	79.92%	1.64	0.84	1.2	6.1504	0.3908	2.4033	114	229
30%	64.0%	18.96%	83.00%	1.64	0.84	1.2	6.1504	0.3714	2.2844	76	153
40%	64.0%	25.62%	89.66%	1.64	0.84	1.2	6.1504	0.3230	1.9868	36	73
Note: Methodology used from											

Table 16: Required sample sizes for different levels of expected change overtime (Scenario 2 longitudinal study)

Level of change	$P_{1,est}$	delta	$P_{2,est}$	$\bar{P}$	$z_{1-\alpha}$	$z_{1-\beta}$	Deff	1st comp	2nd comp	1st+2nd comp	$n_{initial}$	Non response rate (NRR)	NRR adjusted
Pipe water in dwelling (%)													
10%	15.80%	1.58%	17.38%	16.59%	1.64	0.84	1.18	0.8628	0.4418	1.3046	8045	2%	8209
15%	15.80%	2.37%	18.17%	16.99%	1.64	0.84	1.18	0.8709	0.4459	1.3168	3642	2%	3717
20%	15.80%	3.16%	18.96%	17.38%	1.64	0.84	1.18	0.8789	0.4498	1.3286	2086	2%	2129
25%	15.80%	3.92%	19.72%	17.76%	1.64	0.84	1.18	0.8864	0.4534	1.3398	1379	2%	1408
30%	15.80%	4.68%	20.48%	18.14%	1.64	0.84	1.18	0.8937	0.4569	1.3506	984	2%	1004
Pipe water in yard/premise (%)													
10%	40.0%	4.00%	44.00%	42.00%	1.64	0.84	1.18	1.1447	0.5858	1.7306	2209	2%	2254
15%	40.0%	6.00%	46.00%	43.00%	1.64	0.84	1.18	1.1482	0.5870	1.7353	987	2%	1007
20%	40.0%	8.00%	48.00%	44.00%	1.64	0.84	1.18	1.1513	0.5878	1.7390	558	2%	569
25%	40.0%	9.92%	49.92%	44.96%	1.64	0.84	1.18	1.1537	0.59	1.74	364	2%	371
30%	40.0%	11.84%	51.84%	45.92%	1.64	0.84	1.18	1.1558	0.5878	1.7436	256	2%	261
Tap water in yard (%)													
10%	14.8%	1.48%	16.28%	15.54%	1.64	0.84	1.18	0.8403	0.4303	1.2705	8696	2%	8874
15%	14.8%	2.22%	17.02%	15.91%	1.64	0.84	1.18	0.8483	0.4343	1.2826	3939	2%	4019
20%	14.8%	2.96%	17.76%	16.28%	1.64	0.84	1.18	0.8563	0.4382	1.2945	2257	2%	2303
25%	14.8%	3.67%	18.47%	16.64%	1.64	0.84	1.18	0.8637	0.4418	1.3055	1493	2%	1523
30%	14.8%	4.38%	19.18%	16.99%	1.64	0.84	1.18	0.8710	0.4454	1.3164	1065	2%	1087
Tap water outside yard (%)													
10%	30.19%	3.02%	33.21%	31.70%	1.64	0.84	1.18	1.0792	0.5525	1.6317	3447	2%	3517
15%	30.19%	4.53%	34.72%	32.45%	1.64	0.84	1.18	1.0859	0.5555	1.6415	1550	2%	1582
20%	30.19%	6.04%	36.23%	33.21%	1.64	0.84	1.18	1.0923	0.5583	1.6506	882	2%	900
25%	30.19%	7.49%	37.68%	33.93%	1.64	0.84	1.18	1.0982	0.5607	1.6589	579	2%	591
30%	30.19%	8.94%	39.13%	34.66%	1.64	0.84	1.18	1.1037	0.5628	1.6665	410	2%	419
Pipe water in neighbor's yard/premise (%)													
10%	10.8%	1.08%	11.85%	11.31%	1.64	0.84	1.18	0.7345	0.3762	1.1107	12550	2%	12806
10%	10.8%	1.62%	12.39%	11.58%	1.64	0.84	1.18	0.7421	0.3800	1.1221	5692	2%	5809
20%	10.8%	2.15%	12.92%	11.85%	1.64	0.84	1.18	0.7495	0.3837	1.1332	3266	2%	3333
25%	10.8%	2.67%	13.44%	12.11%	1.64	0.84	1.18	0.7565	0.39	1.14	2164	2%	2208
30%	10.8%	3.19%	13.96%	12.36%	1.64	0.84	1.18	0.7634	0.3906	1.1540	1546	2%	1578
35%	10.8%	3.77%	14.54%	12.65%	1.64	0.84	1.18	0.7711	0.3943	1.1654	1128	2%	1151
40%	10.8%	4.31%	15.08%	12.92%	1.64	0.84	1.18	0.7780	0.3977	1.1757	879	2%	897
Packet water (%)													
10%	64.0%	6.40%	70.44%	67.24%	1.64	0.84	1.18	1.0885	0.5562	1.6448	778	2%	794
15%	64.0%	9.61%	73.65%	68.84%	1.64	0.84	1.18	1.0742	0.5472	1.6214	336	2%	343
20%	64.0%	12.81%	76.85%	70.44%	1.64	0.84	1.18	1.0583	0.5367	1.5950	183	2%	187
25%	64.0%	15.88%	79.92%	71.98%	1.64	0.84	1.18	1.0416	0.5251	1.5667	115	2%	117
30%	64.0%	18.96%	83.00%	73.52%	1.64	0.84	1.18	1.0234	0.5119	1.5353	77	2%	79

6.7 Annex 7: Baseline Memo

SOCIAL IMPACT
ADVANCING DEVELOPMENT EFFECTIVENESS

BASELINE MEMO
Millennium Challenge Corporation

December 2018

This memo was prepared independently by Basab Dasgupta, Julia Higgins and Carly Farver from Social Impact, Inc. at the request of MCC.

Evaluation of Sierra Leone Threshold Program

Baseline Memo

Background

Purpose

Analytical Approach

Observations and Recommendations

Annex

Background

Importance of DMA Activity Component for the evaluation of the MCC Threshold Program

The MCA (Millennium Challenge Account) Threshold Program Grant Agreement is being conducted to assist Sierra Leone in the implementation of critical institutional and policy reforms in the water and electricity sectors, through: (1) the Regulatory Strengthening Project (RSP); (2) the Water Sector Reform Project (WSRP); and (3) the Electricity Sector Reform Project (ESRP).

This document considers a component of the WSRP program only: The District Metering Area and Standpipe Demonstration Activity (DMA/Kiosk). The WSRP, in general, addresses the issue of inadequate access to reliable clean water within Freetown by focusing on activities designed to provide sector coordination, technical assistance and capacity building, and the DMA/Kiosk pilot study to test new management approaches and reduce non-revenue water (NRW). A majority of the WSRP is directed towards the Guma Valley Water Company (GVWC), the main provider of water services in the Freetown area. The DMA/Kiosk activity leverages technical assistance provided to GVWC under the larger WSRP Institutional Strengthening Activity "by identifying and establishing a controlled district metering area,[84] within which [GVWC] can implement and test its improved business practices and operationalize a non-revenue water strategy."[85] Improved business practices include stronger monitoring of water flows in and out of the selected DMAs, assessing the scope of potential losses, and implementing NRW reduction strategies including:

A community behavioral change campaign to raise awareness around illegal water connections and broken pipes

Reconnecting/converting illegal connections to legal connections

"Piloting of a public private partnership (PPP) model for the operation and management of public standpipes. Public standpipes provide water for a substantial portion of GVWC's customer base, but little revenue is collected by the GVWC from them and they provide unreliable service and often contaminated water to low-income users."[86]

[84] At the time of this memo, two DMAs within Freetown, Kingtom and Aberdeen had been selected for the pilot. These DMAs were selected in consultation with stakeholders, though a DMA selection review process by Adam Smith International (ASI).

[85] M&E Plan, Feb 2018, pg2

[86] DPE/M&E BPA: Evaluation of Sierra Leone Threshold Program, pg5

The DMA/Kiosk activity is a direct opportunity to test innovative strategies for NRW reduction and GVWC's new business and operational culture, including customer engagement. Given the nature of the pilot, the isolation of specific areas within the GVWC customer base, it also presents an opportunity to employ an "impact evaluation of the DMA approach… to determine whether the design and implementation of non-revenue water reduction strategies with GVWC within the DMA, including an outreach and communication plan for illegal connections, will lead to a commercially viable entity." [87]

Findings from the DMA/Kiosk activity may inform future GVWC operations as well as provide insight for the sector as a whole. This evaluation is also expected to make a significant contribution to addressing limited availability and accessibility of data on water services in Freetown.

Objectives of the IE and the role of a baseline

The DMA/Kiosk IE seeks to answer the following questions:[88]

Were the activities in the DMA/Kiosk activity effective at reducing NRW, and if so, which activities were the most effective?

How did the activities piloted impact service reliability, cost of water, water quality, water collection times, consumer citizenship attitudes and behaviors and satisfaction with water service in the targeted DMA? What evidence is there that the technical assistance provided to GVWC resulted in increased capacity to affect outcomes related to gender and social equity in the DMA?

What is the impact of the DMA Activity on standpipe management, the levels of water service provided by the developed/rehabilitated standpipes under the THP and post-THP (e.g. hours of water, water quality, etc.)? How does the private sector approach to standpipe management (MCC-sponsored kiosks) compare to other standpipe management approaches (e.g., community managed kiosks or networks) in terms of service reliability, maintenance, cost to households and revenues for GVWC?

Are there systems and results in the DMA around NRW and the kiosk pilot that can be sustainably replicated to other GVWC service areas? What factors/indicators are critical to ensure successful replication?

Social Impact's (SI) evaluation team's (ET) proposed method for answering these questions is through a propensity score matching approach, matching either enumeration areas (EAs) or clusters of contiguous EAs outside the treatment DMAs with EAs or clusters of EAs inside the treatment DMAs. Doing so requires the identification of a valid comparison area. Prior to SI's ET involvement a comparison DMA, Cockle Bay, was chosen from previously winnowed list of 4-5 potential treatment

[87] M&E Plan, Feb 2018, pg9

[88] M&E Plan, Feb 2018, pg11

DMAs. In theory, Cockle Bay has many of the same characteristics as the treatment DMAs though some differences are expected given its final exclusion from the treatment.

Baseline data and a review of the DMA selection process will help the ET confirm Cockle Bay as a valid comparison DMA. In addition, strong baseline data are required for the proposed evaluation which includes the matching of EAs between treatment and comparison DMAs on baseline characteristics.

Baseline conducted by third-party

Baseline data was collected by a third-party consultant, ASI, in May and June 2018, prior to the contracting of the ET. Data was collected in all 3 DMAs (2 Treatment and 1 Comparison) and targeted at the household level. The *District Metering Area – Baseline Data Collection Guidance* from MCCU give an overview of the survey design, proposed indicators and planned execution. *Customer Study Design and Qualitative Administration (Task 4.5.3)* and the *Knowledge Attitude and Practice Survey Report (Task 4.5.2)* from ASI summarize baseline findings.

In general, the ET considers two surveys to be the primary source of baseline data for the IE and these will be the focus of this memo:[89]

Knowledge, Attitude and Practice (KAP) Survey – aimed at providing an overview of the KAP of current and potential GVWC customers (n = 670)

Develop a background understanding of residential customer service status

Provide a baseline in relation to water use and service levels, behavioral practices and the social impacts of water

Customer Mapping Survey (CMS) – aimed at providing the location and inventory of most current and potential GVWC customers (n = 4472)[90]

Provide information to develop a functional interactive customer map and customer information system (CIS) including connection types, customer satisfaction

Provide information for customer tracking and billing within the DMA including information on billing and payment

[89] Within the scope of the evaluation additional documentation and data, including administrative data and SMEC mapping data may be incorporated but they are not the focus here.

[90] KAP sampling data estimates 6700 households within the three DMAs, it is unclear to the ET which HHs have been excluded from the CMS

MCC request for validation of the baseline done by ASI as an optional task

This memo is being exercised in the execution of Optional Task 12 as part of SI's evaluation contract with MCC. Option Task 12: Advise on Baseline Data Collection notes "*The Evaluator shall develop a summary report of the findings of its review of the survey materials processes and provide recommendations to the Consultant and any subcontracted data collection firm.*"[91] This baseline memo serves as a summary report to assess ASI's baseline data approach and data collected to ascertain whether or not it fulfills the evaluation needs.

Purpose

This memo will seek to assess ASI's baseline approach or suggest an alternative data collection activity or evaluation design by responding to the following questions:

Are the sampling approach and sample size appropriate for comparative analysis?

An evaluation's sample size can significantly impact the quality of the research findings. A sample that is too small (insufficiently powered) can fail to find results even when a true change has been realized in the population due to a lack of precision. Alternatively, it may also find results where none exist due to capturing a unique non-representative sample of the population. When evaluating the sampling approach, the ET will look for (1) proper representation of the population being studied in order to answer evaluation questions (EQs) and (2) the approach is appropriate for the intended evaluation method (propensity score matching). Assuming the approach is valid and the sample size calculated is correct the ET can move on to examine other components of data collection such as data quality.

If the ET determines the sampling approach was incorrect it will determine and narrate a more appropriate approach, providing new power calculations and estimates of the appropriate sample size for this study. From the new calculations, the team will (1) determine if baseline data collected by ASI meets the sample size requirements and, if not, (2) determine whether the study area population is large enough to capture the appropriate sample through additional data collection.

Since the KAP survey incorporates necessary indicators on water use and service levels, behavioral practices and the social impacts of water that are not captured in the customer mapping survey it serves as the primary survey of focus to ensure the sample size requirements are met.

[91] DPE/M&E BPA: Evaluation of Sierra Leone Threshold Program, pg20

Are indicators identified/selected and questions asked on those indicators in the baseline questionnaires sufficient to answer all DMA/Kiosk EQs?

While customer survey data will likely support analysis across all DMA/Kiosk EQs touching on attitudes towards NRW, including illegal connection and pipe cutting (EQ1), accessibility and affordability of kiosks (EQ3) and unique community characteristics (EQ4) most customer household survey data will respond to EQ2: *How did the activities piloted impact service reliability, cost of water, water quality, water collection times, consumer citizenship attitudes and behaviors and satisfaction with water service in the targeted DMA? What evidence is there that the technical assistance provided to GVWC resulted in increased capacity to affect outcomes related to gender and social equity in the DMA?*

This is reflected in the key M&E indicators (Feb 2018) which note the following will be measured through the DMA survey:

Access to improved water supply: Percentage of households whose main source of drinking water is a private piped connection, public tap/standpipe, tube-well, protected dug well, protected spring or rainwater.

Percentage of population using safely managed drinking water services: Population using an improved water source which is located on the premises, available when needed and free of fecal (and priority chemical) contamination.

Percentage of population using an improved water source: Population using an improved water source with a total collection time of no more than 30 minutes roundtrip including queuing.

Residential water consumption: The average water consumption in liters per person per day.

Time spent fetching water from home in last week: Average time spent gathering water, based on household water gathering activities.

Consumer satisfaction with water reliability: Percent of households that are satisfied with water reliability (continuity of service)

Consumer satisfaction with water quality: Percent of households that are satisfied with water quality

While key indicators are reviewed and revised throughout the program the ET will consider these as the primary requirements for a valid baseline study though section 3.2 will include additional recommendations by the ET.[92]

[92] Additional indicators may want to consider disaggregation by gender/age for time collecting water, and include experience of violence/bullying at water collection points depending on the level of the study's interest in gender-based outcomes

Is data quality of baseline surveys reliable? Can results be replicated?

Assuming ASI baseline data collection instruments include key indicators it is important to confirm that the collected data are reliable and complete. While extensive time could be given to a detailed examination and cleaning of the data provided, for this memo the ET has focused on whether the ET can replicate key findings and tables from the ASI reports. If tables cannot be replicated due challenges in reconstructing indicators, missing, incomplete or improperly coded data the ET will need access to any cleaning files or data analysis documents used by ASI.[93]

Is Cockle Bay, the comparison DMA identified by ASI, an appropriate comparison DMA?

One of the key requirements of an impact evaluation is a valid counterfactual (comparison) group. As noted at the proposal stage "*EQs for this activity inquire about changes that can be causally attributed to the program, which involves estimating a counterfactual. This is often challenging in infrastructure projects, due to limitations in the ability to randomize interventions, selective targeting, and other factors… Presumably, the DMAs were selected purposefully due to its suitability for treatment from an engineering or socioeconomic perspective. For example, the DMAs might have been selected because they contained existing public standpipes in neighborhoods or wards with significant populations of households that have demonstrated demand for an improved water source but did not have the willingness and ability to pay for a private connection. Such unique characteristics mean that a comparison community needs to be carefully matched and selected."*[94]

Since both treatment and comparison DMAs were identified prior to the involvement of the ET this section analyzes the selection process to determine whether the identified DMA, Cockle Bay, can serve as a valid comparison. It does this through a two-step process:

Review of ASI's Selection Methodology which narrowed potential treatment areas down to five DMAs and eventually selected two for treatment and one for comparison. Here the ET examines how the final treatment DMAs were chosen and notes the difference between them and the comparison DMA. The ET also considers whether these differences are sufficient to warrant seeking a revised comparison DMA (treatment DMAs are understood to be fixed), the feasibility of finding a better alternative, or the potential that no valid comparison DMA exists and thus the requirement to revise the evaluation method.

[93] The ET had a call with ASI on Dec 12, 2018 and understand that no cleaning or analysis documentation outside of provided reports exists. The ET is, therefore, working on the assumption that all findings can be replicated from the provided, raw data.

[94] SI Technical Proposal

While slight differences are expected between the treatment and comparison DMAs based on the selection methodology, if these differences are considered surmountable given the value added of keeping the comparison, the team will also examine household level indicators (not captured in the selection methodology) as a further step in ensuring comparability. Ideally the communities themselves would match on community level indicators such as income levels, residency, average level of education, and other demographic factors. This would assist in serving as a proxy for other community level services (access to schools, sanitation, trash removal, electricity) which may play a role water quality outcomes and access. However, while matching at the community level is important, it is not a prerequisite for the evaluation design as a long as enough households can be matched across DMAs. Therefore, our analysis focuses primarily on household level matching.

Analytical Approach

The Analytical Approach section responds to the questions outlined in the previous section. For example, Section 3.1 responds to question 2.1 above on whether the sampling approach and size is appropriate for the comparative analysis. Each question is incorporated again below for reference.

Sampling Approach and sample size

Are the sampling approach and sample size appropriate for comparative analysis?

Since the KAP survey incorporates the most comprehensive household level data, and therefore is the most useful baseline data source, the ET gives it primary focus in responding to this question.

From the KAP Survey Report, the ET understands the KAP sample was selected using a "cross-sectional multi-stage clustering" sample design. The first stage included the clustering of households within Enumeration Areas (EAs) based off the Sierra Leone 2015 census; 67 EAs were identified among the 3 DMAs. The KAP survey then chose to sample 10 percent of households randomly from each EA. The authors of the report estimated that a sample size of 551 households was required to achieve a 95 percent confidence interval and 4 percent margin of error within the areas sampled.[95] Given that there are 67 EAs the study identifies a total sample size of 670 households, well above the estimated required sample of 551.[96] Table 1 shows the distribution of households across the EAs and in the total sample as reported in KAP study.

[95] Task 4.5.2: Knowledge Attitude and Practice (KAP) survey report Pg14

[96] The EAs were evenly distributed by Statistic Sierra Leone (SSL) in 2015 with an equal number of households of 100 for each EA. The report informs that it uses the same EAs that SSL has constructed for

Table 1: Sample size based on the EAs[97]

DMA	Number of EAs	Number of	Sample size (10
Aberdeen	20	2,000	200
Cockle Bay	15	1,500	150
King Tom	32	3,200	320
TOTAL	67	6,700	670

To validate the sample size, the ET first discusses the appropriateness of the sampling approach for the proposed evaluation and then assesses the sufficiency of sample size through the lens of available literature and an estimation of sample size required to assure at least 80 percent analytical power.

Validation of sampling approach

The ET determined that this sampling approach, while robust for a one-period *descriptive analytical study*, is not appropriate for a *comparative analytical study* such as impact evaluation or longitudinal study.

The descriptive survey is generally used to provide a snapshot of the situation at a single point in time. Therefore, the proposed sample size is aimed at achieving a reasonable level of precision (i.e., a small standard error) by specifying a "margin of error" (MOE) for indicator estimates. The comparative analytical survey, on the other hand, is used when the objective of the study is to conduct statistical tests of differences between indicators from different groups or at different time points. This survey requires that the sample size is large enough (i) to control for the levels of inferential errors associated with the statistical tests of differences; (ii) to maintain sufficient statistical/analytical power (at least 80 percent) and (iii) to assure maximum comparability (match) across groups (comparison and treatment) particularly when an IE methodology is used. Because of their differences in objectives, the sampling approach for descriptive surveys are simpler and tend to result in smaller sample sizes than those for comparative analytical surveys, although this is not always the case.

Given the differences in the sampling approach, the ET looked to address the question of whether the available survey data can still serve as a valid baseline under a different approach. That is, does the KAP survey sample have sufficient observations to control for level of inferential errors, maintain sufficient analytical power and assure maximum comparability.

Sample size requirements for comparative study

In this section, the ET discusses the different scenarios, comparison group matching and longitudinal study design, which warrant a sampling approach different from that used for the KAP Survey. Analyzing these different approaches allows the ET to determine the optimum sample sizes which can determine whether the available sample size from the KAP survey is optimum (even though the initial approach may not be robust).

the 2015 Census. The three DMAs have altogether 67 EAs. As a result, a random selection of 10 percent of households from each EA leads to 670 households total from the three DMA areas for the KAP survey.

[97] Source: KAP survey Report, 2018

The main difference between the sampling approach of a *descriptive survey* and a *comparative survey* is that unlike the descriptive survey, the comparative survey starts with the question of the minimum detectable effect size that it is expecting to be achieved over a period of time. Although it can be a hypothetical number (average or percent change), the estimation of optimal sample size significantly hinges on this number to assure sufficient statistical/analytical power after controlling for specific level of inferential errors. As a result, based on baseline value of each outcome of interest (from KAP survey data) and corresponding minimum detectable effect sizes over time, the required sample size will also change. The largest required sample size out of all of potential outcomes of interest is then selected so that the sample size is sufficient to capture changes across indicators.

Scenario 1: IE with a comparison group (matching) Due to high interest in an impact evaluation among stakeholders, the ET first analyzes a scenario where highly similar households across treatment and comparison DMAs will be identified and matched.

In the case where a randomized control trial design is not feasible because treatment group is already identified,[98] a quasi-experimental analysis using matching becomes the next best option for an impact evaluation. A statistical matching approach generally attempts to establish a valid comparison group by matching observable baseline characteristics of treatment and comparison group such that the differences between treatment and comparison groups are minimized and approximate the results of a randomized selection. Once matching is established at baseline with similar baseline characteristics, a simple comparison of difference or other statistical approach such as difference in difference method can provide information about the changes over time which can be attributable to the treatment.

Based on this concept of matching on baseline characteristics the ET first assesses whether it can confirm a valid comparison group from Cockle Bay DMA for the two treatment DMAs (this is expanded upon further in section 3.4). Since the outcome measures will capture household level changes, the ET uses household level characteristics to match households from two treatment DMAs with households from comparison DMA, Cockle Bay. Instead of creating two different arms of treatment for two different treatment DMAs, ET combines them together and consider them as households from same treatment group. The assumption is reasonable because the review of documents and the ET's conversation with the stakeholders suggest that characteristically these two treatment DMAs are similar. Moreover, they will receive the same treatment/intervention without any fundamental difference.

Once the matching of households between treatment and comparison DMAs is completed (step 1), the ET separately calculates the optimal sample size required for a valid IE using a statistical approach and compares the available matched sample size with the sufficient size required to conduct an IE (step 2). The methodological details for identifying optimal sample size are provided in the Annex 1.

[98] In the case of SL threshold DMA Activity, Aberdeen and Kingtom are already identified to receive the treatment.

Step 1: Matching households from KAP survey. We used sex of the respondent, sex of the household head, ownership status of the residence, level of education, and household's religion as baseline characteristics to match comparison households from Cockle Bay to beneficiary households from the treatment group comprised of Aberdeen and Kingtom DMAs. The matching method could identify 144 comparison households from Cockle Bay for 407 treatment households jointly from Aberdeen and Kingtom. Note that despite using a bare minimum number of basic household characteristics we matched upon here, we could not reach a total sample size of 551 households which is equal to the minimum number that the KAP survey reports to be sufficient.

After matching we validate the balance across groups by comparing several outcome indicators. Table 2 provides an example of this validation technique by comparing matched treatment and control households on the indicator of *household water sources during the dry season.* The table, and other indicators evaluated, shows that matched households are well balanced and lessons the concern that Cockle Bay cannot provide a valid comparison.

Table 2: Access to sources of water in Dry season for the matched sample (percent)

	Number of households		Sample mean			
	Comparison	Treatment	Comparison	Treatment	Diff [T-C]	P value
Households with access to piped water in dwelling (percent)	144	407	13.2 percent	17.7 percent	4.5 percent	0.19
Households with access to piped water in yard (percent)	144	407	37.5 percent	38.1 percent	0.6 percent	0.90
Households with access to tap in own yard (percent)	144	407	18.1 percent	14.7 percent	-3.3 percent	0.37
Households with access to tap not in own yard (percent)	144	407	27.8 percent	30.5 percent	2.7 percent	0.54
Households with access to piped water in neighbor's dwelling (percent)	144	407	13.2 percent	10.6 percent	-2.6 percent	0.41
Households' use of packet water in dwelling (percent)	144	407	69.4 percent	63.4 percent	-6.1 percent	0.18

Note: Based on SI's calculation using unpaired t-test across groups

Step 2: While the above analysis in Step 1 suggests that the matching method is able to identify a valid comparison groups of households from Cockle Bay, it does not assure whether the sample size is optimum to guarantee an analytical power of at least 80 percent to conduct a meaningful IE with internal validity. In this step the ET uses a statistical approach to estimate the optimal sample size for

an IE. The detail methodology is provided in the Annex. The ET used a cluster randomization method with household level outcomes.

Table A1 in the Annex computes different sample sizes based on various levels of change expected to be realized by the project.[99] It considers multiple outcome indicators and corresponding intra-class correlation (ICC) to incorporate cluster level variation, noting the required sample size per arm. The final sample size refers to the required sample size with 2 percent non-response rate. Table A1 in the Annex suggests that around 1,000 households per arm (or a total of 2000) is needed to assure a minimum detectable change of (i) 30 percent (or, 4.7 percentage points) increase in access to piped water,[100] (ii) 15 percent increase in access to piped water in own premise; (iii) 30 percent increase in access to tap water in yard; (iv) 20 percent increase in access to tap water outside yard and (v) 35 percent increase in access to water in neighbors' yard with 80 percent analytical power. The ET focused on primary water source as a key outcome indicator of the evaluation.

The ET's sample size calculations rule out the possibility of an IE using KAP survey data unless the expected minimum detectable effect size is substantially increased to levels unlikely to be achieved in the course of the project.

Customer mapping data: Since the sample size from the KAP data is considered inadequate according to the analysis above, the ET considered the potential uses of the customer mapping data as (1) the primary data source and (2) its inclusion in a sampling frame for a new baseline survey. In this section the ET examines whether the sample size of the customer mapping data meets the requirements described above for a comparative study. The appropriateness of the customer mapping data as the primary baseline data source due to available indicators is explored further in section 3.3.

The customer mapping dataset suggests there are sufficient number of households across DMAs for a matching IE design. The ET utilized a similar matching approach as the KAP survey, but substituted some demographic data (unavailable in the customer mapping survey) with outcome indicators relevant for this study, such as (i) *service connection type*: whether the household has GVWC connection, possess metered connection, type of service connection, and whether they are willing to connect to GVWC service if they are not connected yet and (ii) *time to collect water*: whether household

[99] We include level of change up to 30 percent. Since required sample decreases as expected level of changes increase, a higher level change may reduce sample size further which SI can discuss with MCC what settle on the expected level of change they are envisioning.

[100] Table A1 in the appendix provides greater detail regarding sample size selection and includes alternative detectable change sizes and their associated required sample size. This value should be read as a 30 percent increase over the baseline household value. Assuming appropriate matching there should be no detectable difference between the control and treatment households at baseline. Therefore, the average access level across all three DMAs, at baseline, of 15.8 percent to piped water is used. At 30 percent increase translates into a 4.74 percentage point increase during the time of the evaluation.

members need to travel to fetch water, waiting time to collect water. The matching process identifies 1,334 comparison households from Cockle Bay that match 1942 treatment households from Aberdeen and Kingtom (Table 3).

Table 3: Number of households across DMAs

DMAs	Unmatched sample		Matched sample	
	Freq.	Percent	Freq	Percent
Cockle Bay	1,612	36.87	1,334	40.72
Aberdeen	1,155	26.42	804	24.54
Kingtom	1,605	36.71	1,138	34.74
Total	4,372	100	3,276	100

Source: Customer mapping survey data.

Scenario 2: Longitudinal study without a comparison group. While an IE is feasible given the available study population as noted in the preceding section, it may not be advisable or cost effective in terms of time and money. In that case the ET also considered the potential of conducting a longitudinal analysis involving the same households over time from the treatment DMAs. For this design, comparison is made overtime for the same households before and after the treatment. Similar to scenario 1, the ET proposes to measure percentage changes in binary outcomes (say, access to different components of improved water facilities (Y/N)). In that case Table A2 in Annex demonstrates the varying sample sizes on each of the proposed indicators depending on size of expected desired change that could be confidently measured (the effect size) with sufficient analytical power.

The detail methodology is given in the Annex 1. Similar to scenario 1, it uses information on desired level of change, and ICC across clusters to have a cluster (EA) level analysis based on household level outcomes. The only difference between scenario 1 and 2 in terms of design is that we track households over time from the treatment DMAs. No matching between treatment and comparison DMAs is necessary since same households, in the treatment DMAs only, will be visited in baseline, midline and endline. The advantage of this approach is that it requires a smaller sample size compared to a rigorous IE. However, it is impossible to attributed causal effects to the program using this design.

The Table A2 in the Annex suggests that around 1,000 households overall is needed to assure a minimum detectable change of (i) 30 percent in access to piped water in premise, (ii) 15 percent in access to piped water in own premise; (iii) 30 percent in access to tap water in yard; (iv) 20 percent in access to tap water outside yard and (v) 35 percent in access to water in neighbors' yard with 80 percent analytical power. Note that, the required sample size is almost half as compared to that needed for an IE as described in scenario 1 but still greater than what is included in the KAP survey.

Review of documentation including M&E Plan, Evaluation Questions and Baseline questionnaires (2.2)

Are indicators identified/selected and questions asked on those indicators in the baseline questionnaires sufficient to answer all DMA/Kiosk EQs?

The EQs for the DMA/Kiosk pilot activity have been noted in section 1.2 above with EQ2 to be the main focus of the household survey. Primary indicators of interest have drawn from the M&E Plan (Feb 2018) as noted in section 2.2.

Table 4 includes both key indicators as noted in the M&E plan as well as additional demographic data necessary for a successful evaluation and suggested additional outcomes to be considered. It also indicates whether these indicators can be addressed through the baseline data provided to the ET. This breakdown is important to understand what information is missing from the current baseline data, as well as what information may be lost in dropping the KAP survey due to insufficient sample size.

Table 4: Inclusion of key measurements in baseline surveys

Measurement	**Indicator**	**KAP** Survey Variable (*SI variable name*)	**Customer Mapping** Survey Variable (SI variable name)
☐ Measurement captured	Measurement or proxy included but not ideal	Measurement excluded	
Demographic Indicators			
Demographic data (used for matching)	Household Monthly Income	**Not Asked** - Indirect Inicators of wealth include: Cost of renting A4 (*A4_dwellpay*) Vehicle ownership A5A (*A5a_4wheel*) Toilet Type A8A (*A8a_toilet)* Source of Income A10 (*A10_income)*	
	Sex of HH Head	AD (*res_HHH_f*)	18 (*hhh_sex*)
	HH Head Highest Level of Education	A9 (*A9_educ)* - Question asked to respondent, not necessarily HH Head	
	Household Religion	A14 (*A14_religion)*	

	Household Size	A12 (*A12A-A12E)*	17 (*hh_size*) - Not disaggregated by age or gender
	Number of Children in HH	A12 (*A12A-A12E)*	
	HH Property Type		8 (*residential*)
	HH Ownership Status	A3 (*A3_owner)* - Question is asked to respondent, not necessarily HH head	9 (*own*) - Question is asked to respondent, not necessarily HH head or property owner (buisnesses)
M&E Plan Indicators			
Access to Improved Water Supply	All Sources of Water (1)	B2 (*B2_dry_all)*	22 (*watersource*) - not disaggregated by season
	Main Source of Water	B3 (*B3a_maindry*)	
% of population using safely managed drinking water services	Measurement of fecal contaimination in drinking water (stored or not)		
	Measurement of fecal contaimination at main water source		
% of population using an improved water source	Main Source of Water	B3 (*B3a_maindry*)	
	Total Collection Time for water from main source (disaggregated by gender, age)	B22 (*B22_dry_min)* - Includes primary water collector but not disaggreagetd by time for each individual	75 (*wait_time*) - does not identify primary water collector nor disaggregate by individual
Residential water consumption	Average water consumption per HH member per day		
Time spent fetching water in the past week	Total collection time from all water sources	Only included for main source	75 (*wait_time*) - noted for those who have to "fetch" water but may not be inclusive of all water collection activities

Consumer satisfaction with water reliability	Satisfaction with reliability		84 (*cs_reli*) - Any provider 88 (*cs_hours*) - GVWC number of hr supply
Consumer satisfaction with water quality	Satisfaction with quality	B6 (*B6_maindry_quality)*	53 (*ct_tankqual*) - Refers only to quality of GVWC water stored in tank 85-87 *(cs_clean - cs_badtaste)*
Additional Important Indicators			
Cost of water	Average cost over X period per household (2)	B5a-B5e (*B5a_maindry_pay_yn*)	50 (*avg_bill_amt*) - Asked of GVWC customers only for GVWC directly supplied water 78 (*contain costs)* - Cost per container
Water Availability	Hours water is available from a main source	B7-B9 (*B7_maindry_days)*	63-64 (*connect_days-connect_hrs)* - GVWC connections only
NRW	% of HHs that believe one should pay for water	C1A (*C1a_shouldpay)*	
	Acknowledgement of pipe cutting and attitudes	D1A (*D1a_cut*) D1D (*D1d_auth)*- Would report to authority	
	% of HH aware of connections without billing	D4A (*D4c_illigal)*	
Precieved Safety on Water Collection	% of HH that believe water collection source point is safe	E1A (*E1a_safe*) - E7 (*E7_reportvio)*	
(1) Any KAP indicators asked for Dry Season were also repeated, with data available, for rainy season - water sources NOT disaggregated for CM			

(2) Costs are calculated in the KAP for the top 3 water sources in each season			

Measurement	Indicator	KAP Survey Variable (*SI variable name*)	Customer Mapping Survey Variable (SI variable name)
☐ Measurement captured	Measurement or proxy included but not ideal	Measurement excluded	
Demographic Indicators			
Demographic data (used for matching)	Household Monthly Income	**Not Asked** - Indirect Indicators of wealth include: Cost of renting A4 (*A4_dwellpay*) Vehicle ownership A5A (*A5a_4wheel*) Toilet Type A8A (*A8a_toilet)* Source of Income A10 (*A10_income)*	
	Sex of HH Head	AD (*res_HHH_f*)	18 (*hhh_sex*)
	HH Head Highest Level of Education	A9 (*A9_educ)* - Question asked to respondent, not necessarily HH Head	
	Household Religion	A14 (*A14_religion)*	
	Household Size	A12 (*A12A-A12E)*	17 (*hh_size*) - Not disaggregated by age or gender
	Number of Children in HH	A12 (*A12A-A12E)*	
	HH Property Type		8 (*residential*)
	HH Ownership Status	A3 (*A3_owner)* - Question is asked to respondent, not necessarily HH head	9 (*own*) - Question is asked to respondent, not necessarily HH head or property owner (buisnesses)
M&E Plan Indicators			
Access to Improved Water Supply	All Sources of Water (1)	B2 (*B2_dry_all)*	22 (*watersource*) - not disaggregated by season

	Main Source of Water	B3 (*B3a_maindry*)	
% of population using safely managed drinking water services	Measurement of fecal contaimination in drinking water (stored or not)		
	Measurement of fecal contamination at main water source		
% of population using an improved water source	Main Source of Water	B3 (*B3a_maindry*)	
	Total Collection Time for water from main source (disaggregated by gender, age)	B22 (*B22_dry_min)* - Includes primary water collector but not disaggreagetd by time for each individual	75 (*wait_time*) - does not identify primary water collector nor disaggregate by individual
Residential water consumption	Average water consumption per HH member per day		
Time spent fetching water in the past week	Total collection time from all water sources	Only included for main source	75 (*wait_time*) - noted for those who have to "fetch" water but may not be inclusive of all water collection activities
Consumer satisfaction with water reliability	Satisfaction with reliability		84 (*cs_reli*) - Any provider 88 (*cs_hours*) - GVWC number of hr supply
Consumer satisfaction with water quality	Satisfaction with quality	B6 (*B6_maindry_quality)*	53 (*ct_tankqual*) - Refers only to quality of GVWC water stored in tank 85-87 *(cs_clean - cs_badtaste)*
Additional Important Indicators			

Cost of water	Average cost over X period per household (2)	B5a-B5e (*B5a_maindry_pay_yn*)	50 (*avg_bill_amt*) - Asked of GVWC customers only for GVWC directly supplied water 78 (*contain costs)* - Cost per container
Payment Method	Method of Payment		49 (*payment_method*)
Water Availability	Hours water is available from a main source	B7-B9 (*B7_maindry_days)*	63-64 (*connect_days-connect_hrs)* - GVWC connections only
NRW	% of HHs that believe one should pay for water	C1A (*C1a_shouldpay)*	
	Acknowledgement of pipe cutting and attitudes	D1A (*D1a_cut*) D1D (*D1d_auth)*- Would report to authority	
	% of HH aware of connections without billing	D4A (*D4c_illigal)*	
Perceived Safety on Water Collection	% of HH that believe water collection source point is safe	E1A (*E1a_safe*) - E7 (*E7_reportvio)*	
(1) Any KAP indicators asked for Dry Season were also repeated, with data available, for rainy season - water sources NOT disaggregated for CM			
(2) Costs are calculated in the KAP for the top 3 water sources in each season			

Measurement	**Indicator**	**KAP** **Survey Variable (*SI variable name*)**	**Customer Mapping** **Survey Variable (SI variable name)**

☐ Measurement captured	Measurement or proxy included but not ideal	Measurement excluded	
Demographic Indicators			
Demographic data (used for matching)	Household Monthly Income	**Not Asked** - Indirect Inicators of wealth include: Cost of renting A4 (*A4_dwellpay*) Vehicle ownership A5A (*A5a_4wheel*) Toilet Type A8A (*A8a_toilet)* Source of Income A10 (*A10_income)*	
	Sex of HH Head	AD (*res_HHH_f*)	18 (*hhh_sex*)
	HH Head Highest Level of Education	A9 (*A9_educ)* - Question asked to respondent, not necessarily HH Head	
	Household Religion	A14 (*A14_religion)*	
	Household Size	A12 (*A12A-A12E)*	17 (*hh_size*) - Not disaggregated by age or gender
	Number of Children in HH	A12 (*A12A-A12E)*	
	HH Property Type		8 (*residential*)

	HH Ownership Status	A3 (*A3_owner)* - Question is asked to respondent, not necessarily HH head	9 (*own*) - Question is asked to respondent, not necessarily HH head or property owner (buisnesses)
	M&E Plan Indicators		
Access to Improved Water Supply	All Sources of Water (1)	B2 (*B2_dry_all)*	22 (*watersource*) - not disaggregated by season
	Main Source of Water	B3 (*B3a_maindry*)	
% of population using safely managed drinking water services	Measurement of fecal contaimination in drinking water (stored or not)		
	Measurement of fecal contaimination at main water source		
% of population using an improved water source	Main Source of Water	B3 (*B3a_maindry*)	
	Total Collection Time for water from main source (disaggregated by gender, age)	B22 (*B22_dry_min)* - Includes primary water collector but not disaggreagetd by time for each individual	75 (*wait_time*) - does not identify primary water collector nor disaggregate by individual
Residential water consumption	Average water consumption per HH member per day		
Time spent fetching water in the past week	Total collection time from all water sources	Only included for main source	75 (*wait_time*) - noted for those who have to "fetch" water but may not be

			inclusive of all water collection activities
Consumer satisfaction with water reliability	Satisfaction with reliability		84 (*cs_reli*) - Any provider 88 (*cs_hours*) - GVWC number of hr supply
Consumer satisfaction with water quality	Satisfaction with quality	B6 (*B6_maindry_quality)*	53 (*ct_tankqual*) - Refers only to quality of GVWC water stored in tank 85-87 *(cs_clean - cs_badtaste)*
Additional Important Indicators			
Cost of water	Average cost over X period per household (2)	B5a-B5e (*B5a_maindry_pay_yn*)	50 (*avg_bill_amt*) - Asked of GVWC customers only for GVWC directly supplied water 78 (*contain costs)* - Cost per container
Payment Method	Method of Payment		49 (*payment_method*)
Water Availability	Hours water is available from a main source	B7-B9 (*B7_maindry_days)*	63-64 (*connect_days -connect_hrs)* - GVWC connections only
NRW	% of HHs that believe one should pay for water	C1A (*C1a_shouldpay)*	

	Acknowledgement of pipe cutting and attitudes	D1A (*D1a_cut*) D1D (*D1d_auth)*- Would report to authority	
	% of HH aware of connections without billing	D4A (*D4c_illigal)*	
Precieved Safety on Water Collection	% of HH that believe water collection source point is safe	E1A (*E1a_safe*) - E7 (*E7_reportvio)*	
(1) Any KAP indicators asked for Dry Season were also repeated, with data available, for rainy season - water sources NOT disaggregated for CM			
(2) Costs are calculated in the KAP for the top 3 water sources in each season			

It should be noted that inclusion of an indicator in the table above is not a validation of the quality of data collected, only that it was included as a survey question. This is explored further in section 3.3 as the ET attempted the replication of reported data. However, it is valuable to note in this section the limitations in data quality, especially in the Customer Mapping survey since this survey would serve as a potential substitution for KAP.

To highlight this point the ET looked at one variable included in both surveys, *Days of water availability,* and examined the data quality provided from each survey. In the KAP survey this question is asked of all respondents regardless of the main source of water for both dry and rainy seasons. The data shows 100 percent response rates and all values are reasonable (for a 7-day week):

. tab B7_maindry_days, m

B7. On average, how many days is water available in your main source of water i	Freq.	Percent	Cum.
1	8	1.19	1.19
2	24	3.58	4.78
3	339	50.60	55.37
4	133	19.85	75.22
5	43	6.42	81.64
6	11	1.64	83.28
7	112	16.72	100.00
Total	670	100.00	

In the CMS this same question is asked but only for GVWC water connections only (presumably household connections). This limits the sample size to GVWC household and businesses only, dropping the sample from 4,472 to 1,452 across all three DMAs. Furthermore, the data provided to the ET includes a number of coding errors with responses outside the acceptable range (0-7):

. tab connect_days

63. On average, how many days in the week do you get water from your GVWC connec	Freq.	Percent	Cum.
0	46	3.17	3.17
1	22	1.52	4.68
10	1	0.07	4.75
12	2	0.14	4.89
1460	1	0.07	4.96
15	1	0.07	5.03
2	106	7.30	12.33
24	9	0.62	12.95
3	929	63.98	76.93
35	3	0.21	77.13
4	175	12.05	89.19
5	57	3.93	93.11
6	20	1.38	94.49
60	1	0.07	94.56
65000	3	0.21	94.77
7	74	5.10	99.86
8	1	0.07	99.93
80	1	0.07	100.00
Total	1,452	100.00	

This illustration highlights the challenge of working with a dataset collected by a third party and should be explored further if the preferred route is to use customer mapping data as the project baseline.

Reconstructing summary statistics to validate data quality (2.3)

Is data quality of baseline surveys reliable? Can results be replicated?

Replication, the ability to replicate results produced in a report from raw data, is a key indicator as to the usability of the data for future analysis and a proxy for its trustworthiness. While the KAP Survey collected data relevant to key M&E indicators, ASI's analysis was not easily replicable in the majority of contexts. Using the KAP report as a reference for indicators that were measured, SI compared figures included in ASI's report to outcomes reflected in the raw dataset. In the instances where a high degree of uncertainty was involved in reconstructing ASI analysis, SI constructed indicator outcomes using key variables in the raw dataset; the process used for each indicator is summarized in Table 5 below.

While numbers were not easily replicable across indicators for the KAP survey, this is likely attributed to lack of insight into how analysis was constructed rather than an explicit instance of low data quality. For example, in many cases, it is unclear which type of water source (primary, secondary, etc.) is used for calculation of key outcomes. Despite this, data quality concerns, such as lack of standardization across variables and inconsistent coding and categorization of responses are present and pose a challenge to SI's work down the line. Furthermore, the KAP survey omits data for two ley indicators: residential water consumption and percentage of the population using safely managed drinking water services. The tables in Annex 3 reflect the summary statistics that were generated by SI across key M&E indicators.

Table 5: Summary of KAP Indicator Analysis Reconstruction

Indicator	Definition	Status in KAP report
Access to improved water supply	Percentage of households whose main source of drinking water is a private piped connection, public tap/standpipe, tube-well, protected dug well, protected spring or rainwater.	Though variables were constructed to designate improved and unimproved water sources using the definition listed in Column 2, a discrepancy is present between outcomes observed in the ASI report and raw dataset.
Percentage of population using safely managed drinking water services	Population using an improved water source which is located on the premises, available when needed and free of fecal (and priority chemical) contamination.	ASI did not measure fecal contamination, so no information on this indicator was included in SI analysis.

Percentage of population using an improved water source	Population using an improved water source with a total collection time of no more than 30 minutes roundtrip including queuing.	SI calculated the percentage of households that use an improved water source and do not spend more than 30 minutes collecting water.
Residential water consumption	The average water consumption in liters per person per day.	ASI did not measure residential water consumption, so no information on this indicator was included in SI analysis.
Time spent fetching water from home in last week	Average time spent gathering water, based on household water gathering activities.	SI reported the average time spent gathering water, by DMA. Though ASI included this figure, it is unclear how the means were weighted.
Consumer satisfaction with water reliability	Percent of households that are satisfied with water reliability (continuity of service)	ASI did not explicitly define how satisfaction was measured or calculated, therefore, SI created new variables to indicate a reliable water connection in the rainy and dry seasons, using 20-24 hours of continuous service per day as the standard for a 'reliable' connection.
Consumer satisfaction with water quality	Percent of households that are satisfied with water quality	ASI did not explicitly define how satisfaction was measured or calculated, therefore, SI totaled the percentage of households that described the quality of their water as 'good' in the KAP survey. Households that indicated 'good' alongside a negative characteristic such as 'bad taste' or 'bad odor' were omitted from summary statistics reporting.

Customer Mapping Survey (CMS) analysis was more easily replicable than KAP analysis. Table 6 below maps variables that are closely tied with satisfaction and quality indicators to their respective locations in the CMS report and notes their consistency. Overall, the CMS figures in the raw data

clearly map on to figures included in the report. Slight discrepancies between the numbers reported in the data and figures and text included in the report are likely due to cleaning or recoding that SI is unable to replicate. While more indicators can be addressed through KAP data, CMS analysis was more easily replicable. Tables showing a detailed breakdown of each variable, disaggregated by DMA, can be found in Annex 3.

Table 6: Summary of CMS Indicator Analysis Reconstruction

Indicator type	Variable/Survey Question	Status in CMS Report
Satisfaction	81. How satisfied are you with the level of customer service from your current water provider?	Disaggregations are validated in Figure 42 of the CMS report. Totals are validated in the report's text on page 2.
	83. Are you satisfied with the cost of water?	Disaggregations are validated in Figure 44 of the CMS report. Totals are validated in the report's text on page 2.
	84. Are you satisfied with the reliability of water service from your service provider?	Disaggregations are validated in Figure 45 of the CMS report. Totals are validated in the report's text on page 2.
	88. Are you satisfied with the number of hours of water supply?	Disaggregations are validated in Figure 49 of the CMS report.
	89. Are you satisfied with the process of receiving and paying your bills?	Disaggregations are validated in Figure 50 of the CMS report. Different totals are reflected in the report's text on page 2.
Quality	87. Does the water you get from your water service provider taste bad?	Disaggregations are validated in Figure 48 of the CMS report. Totals are validated in the report's text on page 2.
	86. Does the water you get from your water service provider have a bad odor?	Disaggregations are validated in Figure 47 of the CMS report. Totals are validated in the report's text on page 2.
	85. Is the water you get from your water service provider clean?	Disaggregations are validated in Figure 46 of the CMS report. Totals are validated in the report's text on page 2.

Quality	82. Does your water service provider respond promptly to your complaints about service?	Disaggregations are validated in Figure 43 of the CMS report. Totals are validated in the report's text on page 2.
	80. Good customer service from the people selling you water is important to you?	Disggregations are validated in Figure 41 of the CMS report. Totals are validated in the report's text on page 2.

Document review and DMA comparison of high-level indicators on wealth and water connectivity using Baseline Summary statistics (2.4)

Is Cockle Bay, the comparison DMA identified by ASI, an appropriate comparison DMA?

The DMA Selection Methodology excel file provides the most comprehensive insight into the DMA selection process. Though a series of iterations, including consultations with GVWC, ASI narrowed the DMA selection down to five potential areas for the DMA/kiosk pilot: Aberdeen, Cockle Bay (formerly called Murry Town), Kingtom, Kissy, and Lumley. Kissey and Lumely were eventually discarded due to concerns about resettlement (Kissy) and residential verse commercial composition (Lumley). This left Aberdeen, Kingtom and Cockle Bay. The former two we selected as Treatment DMAs while Cockle Bay became the comparison. Since this final selection was not random but rather part of the consultation process we first examine the selection process to understand what difference, if any, arise between the treatment and comparison DMAs. We then supplement this through comparison of DMA demographics using the baseline data provided to understand potential socio-economic differences.

To start, according to the DMA Selection Methodology, the selection process included the following iterations:

Selection based on DFID works. Starting in 2016 DFID agreed to provide up to £38 million to rehabilitate Freetown's water supply system through the design and construction of key system elements identified by engineers from GVWC. The goal of the program, implemented alongside MCCU's institutional strengthening, was to increase sustainable access to safe water in Freetown. To avoid treatment contamination one criteria of the selected DMAs was that they were not part of a DFID Intervention/Rehabilitation Works area. All three DMAs met these criteria.

Selection based on water supply: Areas selected for the DMA/kiosk pilot had to have sufficient water supply prior to program intervention in order to test and discern effects in supply and NRW reduction. All three areas met these criteria with varying success. It was noted that water is not rationed (e.g. the

valves are never closed) in Aberdeen and some areas get 24 hr. supply depending on water pressure. Likewise, 24 hr. supply was also noted to exist in some areas in Kingtom. This was not the case for the comparison DMA where water is, in theory, rationed 4 days per week with an average of 12 hr. supply per day.

Selection based on technical criteria: Technical criteria included (1) known level of leaks/illegal water connections, (2) length of distribution pipes, (3) complexity to isolate the area, (4) cost and method of isolation, (5) GVWC's qualitative perspective with a final determination of whether or not the DMA was viably technical. All three were determined viably technical but GVWC's feedback was largely opposed to Cockle Bay noting its similarities to Aberdeen but with fewer customers and that it had few values and potential control points. This was likely one of the key deciding inputs in not selecting Cockle Bay for treatment.

Selection based on costs: Isolation cost estimates ranged for the 3 DMAs from $15,500 (Kingtom) to $50,000 (Aberdeen). All three DMAs were considered eligible due to costs.

Selection based on encroachment: Encroachment raises the potential need for resettlement due to infrastructure projects. In treatment DMAs this was not anticipated to be an issue while an evaluation of Cockle Bay was pending at the time of the decision.

Selection based on representative mix: Representative mix included a variety of factors such as current GVWC customers, households, male and female populations, existing standpipes, commercial entities, informal businesses, presence of institutions (schools, religious) and presence of government (police, fire, army). While the individual components vary across the three DMAs, all three were considered to have a sufficient, representative mix.

While the selection methodology highlights some potential concerns as to the differences between the treatment and comparison DMA the ET, at this time, does not think it is sufficient to validate dropping the comparison DMA in favor of another. The feasibility of conducting another assessment with the level of detail and consideration, including technical expertise allocated in the ASI exercise, is low and likely redundant. Reviewing the available options from the original list does not immediately identify a better outcome especially since the treatment DMAs are now fixed.

However, as part the evaluation process the team will want to gather additional information as to potential differences, especially in GVWC's selection choice, as this may influence, or be influenced, by other factors not listed here (i.e. customer relationships in the three DMAs, payment practices to date, etc.) that may impact the expected outcomes or cause the realized outcomes to be inappropriately attributed to the DMA/Kiosk pilot over other preexisting differences. In addition, one of the largest lessons learned from a review of the DMA Selection Methodology is that there are large, and potentially confounding, differences between DMAs in the Freetown area. While the IE activity is

aimed at generating lessons learned that can then be extrapolated to future works both within and outside of Freetown the ET and MCC will need to be careful when considering the validity of these "all-star" DMAs compared to the larger Freetown area of unselected DMAs. It is highly likely that should the IE realize the intended results any future works will need to proceed with caution, future piloting, and likely further evaluation (this will be explored further in responding to EQ4 regarding replication).

While it is unlikely, given the cost and technical considerations in DMA selection, to identify an alternative comparison it is still necessary to validate the available comparison. If the best-case comparison DMA is still largely different than the proposed treatment areas it may be worth considering an alternative design rather than maintaining a weak counterfactual. To confirm validity the ET considered baseline household characteristics as well as the DMA Selection Methodology. Upon matching conducted in section 3.1 the ET was able to run some comparison statistics to confirm matched households were not statistically different on key outcomes at baseline (see Table 2).

Observations and Recommendations

Summary of Observations

The following summarizes key observations from SI's review of the available baseline data.

Observation 1: Cockle Bay can be used as comparison DMA. Based on the ET's review of the reports and data made available, the ET thinks that Cockle Bay can be used as a comparison DMA for the evaluation. While there are some differences in the selection methodology and DMA-level outcomes, the proposed IE is dependent at matching at the household level. When DMAs are compared on household level characteristics there are sufficient similarities to justify Cockle Bay as a comparison DMA.

Observation 2: The KAP survey is more exhaustive in terms of number of relevant questions required to successfully complete the IE. Based on review of documentation including the M&E Plan, evaluation questions, baseline questionnaires, and baseline data the ET compared the KAP and customer mapping surveys in section 3.2-3.3. The KAP survey is successful at addressing a majority of key indicators at various levels of disaggregation (by gender or season). The CMS is significantly less successful, which is expected given its limited scope and timeframe for execution at the household. While the ET was more successful at replicating CMS outcomes than KAP this is likely due to the straightforward, limited options, of the CMS data in constructing indicators, more than an indication of

better reliability in the CMS data. In fact, the ET found the KAP dataset to be more consistent with valid data coding as noted in section 3.2.

Observation 3: ET proposes that the sample size provided in the KAP data is not sufficient for a rigorous impact evaluation. The sampling approach used a methodology that is appropriate for a descriptive survey which is mostly used for a one-time snapshot study. For a comparative study, the sample size needs to be large enough not only to control for the levels of inferential errors associated with the statistical tests of differences, but also to maintain sufficient statistical/analytical power (at least 80 percent) and to assure maximum comparability across groups (comparison and treatment)—particularly for an IE. Our matching exercise suggests that with a bare minimum number of indicators/characteristics for matching, the ET could identify 551 matched households. A slight change in indicators to be more inclusive leads to a drastic fall in the number of matched households.

Observation 4: Matching households using the CMS is more stable and yields a sufficient number of matched households required for a robust IE. However, CMS data lacks the desired number of baseline measurements. While customer mapping data sample size is sufficient, the CMS data are not as comprehensive as the KAP survey and does not have all the indicators that the evaluation needs. Moreover, some key indicators have a significantly large number of missing values which may lead to biased estimates of baseline differences across treatment and comparison groups.

Recommendations

In the execution of this memo the ET considered several potential outcomes to be viable depending on the findings. These included:

Option 1a: Comparison DMA is sufficient for an IE and existing data are appropriate to use as a baseline without the need for an additional data collection activity.

Option 1b: Comparison DMA is sufficient for an IE but SI recommends an additional data collection activity to collect new or additional data from households to maintain sufficient analytical power.

Option 2: Comparison DMA is not adequate, and SI identifies another comparison DMA with corresponding baseline data collection for IE.

Option 3: Comparison DMA is not adequate for the IE design or existing data meets some but not all criteria prompting the consideration of alternative approaches for the evaluation. In that case SI will propose the best method feasible.

Based on our observations SI does not recommend:

Option 1a: Comparison DMA is sufficient for an IE and existing data are appropriate to use as a baseline without the need for an additional data collection activity.

The ET does not believe existing data are appropriate for use as an IE baseline. If this option is pursued, the customer mapping data will be the primary baseline data in order to reach a sufficient sample size. The ET and MCC must recognize the limitations to the available dataset including a lack of key variables and concerns about data quality and cleaning.

Option 2: Comparison DMA is not adequate, and SI identifies another comparison DMA with corresponding baseline data collection for IE.

The ET believes that the comparison DMA is adequate and therefore excludes this conclusion.

Based on our observations SI recommends pursuing one of the following options:

Option 1b: Comparison DMA is sufficient for an IE but SI recommends an additional data collection activity to collect new or additional data from households to maintain sufficient analytical power.

In this option, SI would conduct additional baseline data collection, likely using the CMS as a listing dataset. If CMS reliability can be validated any additional data collection would aim to match households to CMS data in order to utilize information already collected. Additional data collection would also allow the team to consider capturing indicators not included in either KAP or CMS data.

Option 2: Comparison DMA is not adequate for the IE design or existing data meets some but not all criteria prompting the consideration of alternative approaches for the evaluation. In that case SI will propose the best method feasible.

In this option, the ET can discuss with MCC alternative evaluation designs including a rigorous PE. This option would exclude an IE evaluation design but would avoid additional baseline data collection.

All of these options, including those SI does not recommend, can be explored in further conversations with MCC. The ET would like to note, however, that should additional baseline data collection be determined as the best route forward, this exercise would need to be executed quickly, prior to the start of treatment in early 2019.

Annex 1. Formulas used to calculate sample size for different scenarios

Comparative study to capture changes over time

The following sampling strategies are used for Scenario 1 and 2 as described in the text.

Scenario 1: Sample Size Calculation with Matching at base line

For binary outcomes, the number of households per arm, assuming a cluster-level analysis, is calculated as[101]

$$n_{initial} = Deff * \left[\frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 + [P_1(1-P_1) + P_2(1-P_2)]}{\delta^2}\right]$$

where:

δ represents the meaningful change (minimum detectable effect size) to be achieved over the time frame ($\delta \neq 0$)

P_1 is the estimated baseline prevalence value

P_2 is the expected ending prevalence value, equal to

$P_1 \pm \delta$

$z_{1-\alpha}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1-α. For 1−α= 0.95, the corresponding value is $z_{0.95} = 1.64$

$z_{1-\beta}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−β which is also known as analytical power? Here in our case (1−β = 0.80, with the corresponding value of $z_{0.80} = 0.84$.

Deff= 1+(n-1) *ICC -is the estimated design effect of the survey. Note that the sample size calculations are quite sensitive to this parameter since it depends of inter cluster correlation (ICC) and number of households per cluster (EAs in this case).

Scenario 2: Sample size calculation for longitudinal study (Without matching)

[101] Rutterford, C; A. Copus and S. Eldridge (2015): Methods for sample size determination in cluster randomized trials, International Journal of Epidemiology, 2015, 1–17, Advance Access Publish in July,13 2015.

Because these indicators are assumed to be proportions (prevalence), the appropriate formula to calculate the required sample size is as follows:

$$n_{initial} = Deff * \left[\frac{z_{1-\alpha}\sqrt{2\bar{P}(1-\bar{P})} + z_{1-\beta}\sqrt{P_1(1-P_{1)}) + P_2(1-P_{2)})}}{[\delta]}\right]^2$$

where:

δ represents the meaningful change (minimum detectable effect size) to be achieved over the time frame ($\delta \neq 0$)

P_1 is the estimated baseline prevalence value

P_2 is the expected ending prevalence value, equal to

$P_1 \pm \delta$

$$\bar{P} = \frac{P_1 + P_2}{2}$$

$z_{1-\alpha}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−α. For 1−α = 0.95, the corresponding value is $z_{0.95} = 1.64$

$z_{1-\beta}$ is the value from the Normal Probability Distribution corresponding to a confidence level 1−β which is also known as analytical power? Here in our case (1−β) = 0.80, with the corresponding value of $z_{0.80} = 0.84$.

Deff = 1+(n-1) *ICC - is the estimated design effect of the survey. Note that the sample size calculations are quite sensitive to this parameter since it depends of inter cluster correlation (ICC) and number of households per cluster (EAs in this case).

Table A1: Required sample sizes for different values of expected change overtime (Scenario 1 with matching)

Level of change	[illegible]	delta	[illegible]	[illegible]	[illegible]	Deff	1st comp	2nd comp	1stX2nd comp	Sample per arm	Total sample size
Pipe water in dwelling (%)											
10%	15.80%	1.58%	17.38%	1.64	0.84	1.2	6.1504	0.2766	1.7014	8178	16357
15%	15.80%	2.37%	18.17%	1.64	0.84	1.2	6.1504	0.2817	1.7327	3702	7404
20%	15.80%	3.16%	18.96%	1.64	0.84	1.2	6.1504	0.2867	1.7632	2119	4238
25%	15.80%	3.92%	19.72%	1.64	0.84	1.2	6.1504	0.2913	1.7918	1400	2801
30%	15.80%	4.68%	20.48%	1.64	0.84	1.2	6.1504	0.2959	1.8197	998	1997
40%	15.80%	6.32%	22.12%	1.64	0.84	1.2	6.1504	0.3053	1.8778	564	1128
Pipe water in yard/premise (%)											
10%	40.0%	4.00%	44.00%	1.64	0.84	1.2	6.1504	0.4864	2.9916	2244	4487
15%	40.0%	6.00%	46.00%	1.64	0.84	1.2	6.1504	0.4884	3.0039	1001	2003
20%	40.0%	8.00%	48.00%	1.64	0.84	1.2	6.1504	0.4896	3.0112	565	1129
25%	40.0%	9.92%	49.92%	1.64	0.84	1.2	6.1504	0.4900	3.0137	367	735
30%	40.0%	11.84%	51.84%	1.64	0.84	1.2	6.1504	0.4897	3.0116	258	516
Tap water in yard (%)											
10%	14.8%	1.48%	16.28%	1.64	0.84	1.2	6.1504	0.2624	1.6138	8841	17682
15%	14.8%	2.22%	17.02%	1.64	0.84	1.2	6.1504	0.2673	1.6442	4003	8007
20%	14.8%	2.96%	17.76%	1.64	0.84	1.2	6.1504	0.2722	1.6739	2293	4585
25%	14.8%	3.67%	18.47%	1.64	0.84	1.2	6.1504	0.2767	1.7017	1516	3032
30%	14.8%	4.38%	19.18%	1.64	0.84	1.2	6.1504	0.2811	1.7290	1081	2162
40%	14.8%	5.92%	20.72%	1.64	0.84	1.2	6.1504	0.2904	1.7859	611	1223
Tap water outside yard (%)											
10%	30.19%	3.02%	33.21%	1.64	0.84	1.2	6.1504	0.4326	2.6604	3503	7005
15%	30.19%	4.53%	34.72%	1.64	0.84	1.2	6.1504	0.4374	2.6902	1574	3148
20%	30.19%	6.04%	36.23%	1.64	0.84	1.2	6.1504	0.4418	2.7172	894	1789
25%	30.19%	7.49%	37.68%	1.64	0.84	1.2	6.1504	0.4456	2.7404	587	1173
30%	30.19%	8.94%	39.13%	1.64	0.84	1.2	6.1504	0.4489	2.7611	415	830
40%	30.19%	12.08%	42.27%	1.64	0.84	1.2	6.1504	0.4548	2.7970	230	460
Pipe water in neighbor's yard/premise (%)											
10%	10.8%	1.08%	11.85%	1.64	0.84	1.2	6.1504	0.2005	1.2334	12760	25520
10%	10.8%	1.62%	12.39%	1.64	0.84	1.2	6.1504	0.2046	1.2585	5786	11573
20%	10.8%	2.15%	12.92%	1.64	0.84	1.2	6.1504	0.2086	1.2832	3319	6638
25%	10.8%	2.67%	13.44%	1.64	0.84	1.2	6.1504	0.2124	1.3066	2198	4396
30%	10.8%	3.19%	13.96%	1.64	0.84	1.2	6.1504	0.2162	1.3297	1570	3140
35%	10.8%	3.77%	14.54%	1.64	0.84	1.2	6.1504	0.2204	1.3553	1145	2289
40%	10.8%	4.31%	15.08%	1.64	0.84	1.2	6.1504	0.2241	1.3786	891	1783
50%	10.8%	5.39%	16.16%	1.64	0.84	1.2	6.1504	0.2316	1.4241	589	1179
Packet water (%)											
10%	64.0%	6.40%	70.44%	1.64	0.84	1.2	6.1504	0.4385	2.6969	789	1578
15%	64.0%	9.61%	73.65%	1.64	0.84	1.2	6.1504	0.4244	2.6101	339	679
20%	64.0%	12.81%	76.85%	1.64	0.84	1.2	6.1504	0.4082	2.5106	184	367
25%	64.0%	15.88%	79.92%	1.64	0.84	1.2	6.1504	0.3908	2.4033	114	229
30%	64.0%	18.96%	83.00%	1.64	0.84	1.2	6.1504	0.3714	2.2844	76	153
40%	64.0%	25.62%	89.66%	1.64	0.84	1.2	6.1504	0.3230	1.9868	36	73
Note: Methodology used from											

Table A2: Required sample sizes for different levels of expected change overtime (Scenario 2 longitudinal study)

Level of change	$P_{1,est}$	delta	$P_{2,est}$	$\bar{P}$	$z_{1-\alpha}$	$z_{1-\beta}$	Deff	1st comp	2nd comp	1st+2nd comp	$n_{initial}$	Non response rate (NRR)	NRR adjusted
Pipe water in dwelling (%)													
10%	15.80%	1.58%	17.38%	16.59%	1.64	0.84	1.18	0.8628	0.4418	1.3046	8045	2%	8209
15%	15.80%	2.37%	18.17%	16.99%	1.64	0.84	1.18	0.8709	0.4459	1.3168	3642	2%	3717
20%	15.80%	3.16%	18.96%	17.38%	1.64	0.84	1.18	0.8789	0.4498	1.3286	2086	2%	2129
25%	15.80%	3.92%	19.72%	17.76%	1.64	0.84	1.18	0.8864	0.4534	1.3398	1379	2%	1408
30%	15.80%	4.68%	20.48%	18.14%	1.64	0.84	1.18	0.8937	0.4569	1.3506	984	2%	1004
Pipe water in yard/premise (%)													
10%	40.0%	4.00%	44.00%	42.00%	1.64	0.84	1.18	1.1447	0.5858	1.7306	2209	2%	2254
15%	40.0%	6.00%	46.00%	43.00%	1.64	0.84	1.18	1.1482	0.5870	1.7353	987	2%	1007
20%	40.0%	8.00%	48.00%	44.00%	1.64	0.84	1.18	1.1513	0.5878	1.7390	558	2%	569
25%	40.0%	9.92%	49.92%	44.96%	1.64	0.84	1.18	1.1537	0.59	1.74	364	2%	371
30%	40.0%	11.84%	51.84%	45.92%	1.64	0.84	1.18	1.1558	0.5878	1.7436	256	2%	261
Tap water in yard (%)													
10%	14.8%	1.48%	16.28%	15.54%	1.64	0.84	1.18	0.8403	0.4303	1.2705	8696	2%	8874
15%	14.8%	2.22%	17.02%	15.91%	1.64	0.84	1.18	0.8483	0.4343	1.2826	3939	2%	4019
20%	14.8%	2.96%	17.76%	16.28%	1.64	0.84	1.18	0.8563	0.4382	1.2945	2257	2%	2303
25%	14.8%	3.67%	18.47%	16.64%	1.64	0.84	1.18	0.8637	0.4418	1.3055	1493	2%	1523
30%	14.8%	4.38%	19.18%	16.99%	1.64	0.84	1.18	0.8710	0.4454	1.3164	1065	2%	1087
Tap water outside yard (%)													
10%	30.19%	3.02%	33.21%	31.70%	1.64	0.84	1.18	1.0792	0.5525	1.6317	3447	2%	3517
15%	30.19%	4.53%	34.72%	32.45%	1.64	0.84	1.18	1.0859	0.5555	1.6415	1550	2%	1582
20%	30.19%	6.04%	36.23%	33.21%	1.64	0.84	1.18	1.0923	0.5583	1.6506	882	2%	900
25%	30.19%	7.49%	37.68%	33.93%	1.64	0.84	1.18	1.0982	0.5607	1.6589	579	2%	591
30%	30.19%	8.94%	39.13%	34.66%	1.64	0.84	1.18	1.1037	0.5628	1.6665	410	2%	419
Pipe water in neighbor's yard/premise (%)													
10%	10.8%	1.08%	11.85%	11.31%	1.64	0.84	1.18	0.7345	0.3762	1.1107	12550	2%	12806
10%	10.8%	1.62%	12.39%	11.58%	1.64	0.84	1.18	0.7421	0.3800	1.1221	5692	2%	5809
20%	10.8%	2.15%	12.92%	11.85%	1.64	0.84	1.18	0.7495	0.3837	1.1332	3266	2%	3333
25%	10.8%	2.67%	13.44%	12.11%	1.64	0.84	1.18	0.7565	0.39	1.14	2164	2%	2208
30%	10.8%	3.19%	13.96%	12.36%	1.64	0.84	1.18	0.7634	0.3906	1.1540	1546	2%	1578
35%	10.8%	3.77%	14.54%	12.65%	1.64	0.84	1.18	0.7711	0.3943	1.1654	1128	2%	1151
40%	10.8%	4.31%	15.08%	12.92%	1.64	0.84	1.18	0.7780	0.3977	1.1757	879	2%	897
Packet water (%)													
10%	64.0%	6.40%	70.44%	67.24%	1.64	0.84	1.18	1.0885	0.5562	1.6448	778	2%	794
15%	64.0%	9.61%	73.65%	68.84%	1.64	0.84	1.18	1.0742	0.5472	1.6214	336	2%	343
20%	64.0%	12.81%	76.85%	70.44%	1.64	0.84	1.18	1.0583	0.5367	1.5950	183	2%	187
25%	64.0%	15.88%	79.92%	71.98%	1.64	0.84	1.18	1.0416	0.5251	1.5667	115	2%	117
30%	64.0%	18.96%	83.00%	73.52%	1.64	0.84	1.18	1.0234	0.5119	1.5353	77	2%	79

Annex 2. Data Collection Activities

Based on discussion with MCC, if a new baseline is decided to be conducted then the following activities will be undertaken and a timeline will be proposed:

Start-up activities

- Sample size determination and activity budgeting
- Identification and Procurement of Survey Firm
- Preparation of Survey Tools, Protocols and Manuals
- IRB approval

Survey activities

- Pretesting and Pilot
- Survey Training
- Household Survey
- Access to Water and Water Quality[102]
- Utility/Administrative Data (from GVWC)[103]

Reporting

- Data Cleaning and Analysis
- Data Quality Report
- Draft Report
- Stakeholder Feedback (if required)
- Final Report
- Final De-Identified Dataset Submission to MCC

[102] This data may be incorporated into the household survey as part a subjective household measure or will be captured as part of a collaboration with Sweet Sense/UC Boulder. separately from the new baseline activity.

[103] Regardless of whether new baseline data are required SI will work to collect this information for overall project monitoring.

Annex 3. Summary Statistics and Analysis Reconstruction

KAP

Indicator	Variable(s) used	Figure number(s) in ASI report	DMA	Rainy season, ASI report	Rainy season, SI analysis	Dry Season, ASI report	Dry season, SI analysis
Access to improved water supply	B2_dry_all B28_wsource	Figures 15 & 16	Aberdeen	73 percent	96 percent	98 percent	94 percent
			Cockle Bay	78 percent	94 percent	99 percent	91 percent
			King Tom	69 percent	96 percent	98 percent	95 percent

Indicator	Variable(s) used	DMA	Rainy season, SI analysis (mean time spent, min)	Dry season, SI analysis (mean time spent, min)
Time spent fetching water from home in last week (percent in 30 min or less)	B22_dry_time_min B48_rain_time_min	Aberdeen	1.18	1.15
		Cockle Bay	1.18	1.14
		King Tom	1.13	1.17

Indicator	Variable(s) used	DMA	Rainy season, SI analysis	Dry season, SI analysis
Consumer satisfaction with water reliability	B8_maindry_hrs B43_secrain_hrs_num (could not locate main rain var)	Aberdeen	55 percent	31 percent
		Cockle Bay	44 percent	30 percent
		King Tom	65 percent	25 percent

Indicator	Variable(s) used	DMA	Rainy season, SI analysis	Dry season, SI analysis
Consumer satisfaction with water quality	B6_maindry_good B33_mainrain_good	Aberdeen	80 percent	72 percent
		Cockle Bay	77 percent	66 percent
		King Tom	83 percent	73 percent

Indicator	Variable(s) used	DMA	Rainy season, SI analysis	Dry season, SI analysis
Population using an improved water source with a total collection time of no more than 30 minutes roundtrip including queuing.	B22_dry_time_min B48_rain_time_min Improved/unimproved binary B2 variables	Aberdeen	30 percent	80 percent
		Cockle Bay	28 percent	76 percent
		King Tom	28 percent	73 percent

CMS

81. How satisfied are you with the level of customer service from your current water provider?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
A little satisfied	836	51.9	541	47.1	735	45.9	2,112.00	48.4
Don't know	100	6.2	96	8.4	74	4.6	270	6.2

Not at all	431	26.7	351	30.5	635	39.6	1,417.00	32.5
Very satisfied	245	15.2	161	14	158	9.9	564	12.9
Total	1,612.00	100	1,149.00	100	1,602.00	100	4,363.00	100
83. Are you satisfied with the cost of water?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	575	37	335	30.7	474	29.8	1,384.00	32.7
No	415	26.7	437	40	705	44.3	1,557.00	36.7
Yes	563	36.3	320	29.3	413	25.9	1,296.00	30.6
Total	1,553.00	100	1,092.00	100	1,592.00	100	4,237.00	100
84. Are you satisfied with the reliability of water service from your service provider?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	193	12	127	11.1	213	13.3	533	12.2
No	912	56.6	786	68.4	1,086.00	67.9	2,784.00	63.9
Yes	507	31.5	236	20.5	300	18.8	1,043.00	23.9
Total	1,612.00	100	1,149.00	100	1,599.00	100	4,360.00	100
88. Are you satisfied with the number of hours of water supply?	Cockle Bay		Aberdeen		Kingtom		Total	

	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	205	12.8	116	10.1	100	6.3	421	9.7
No	929	57.9	782	67.9	1,189.00	74.5	2,900.00	66.6
Yes	471	29.3	254	22	306	19.2	1,031.00	23.7
Total	1,605.00	100	1,152.00	100	1,595.00	100	4,352.00	100
89. Are you satisfied with the process of receiving and paying your bills?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	1,083.00	71	558	50.7	1,078.00	68.6	2,719.00	64.8
No	176	11.5	225	20.4	300	19.1	701	16.7
Yes	267	17.5	318	28.9	194	12.3	779	18.6
Total	1,526.00	100	1,101.00	100	1,572.00	100	4,199.00	100

87. Does the water you get from your water service provider taste bad?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	117	7.3	100	8.7	65	4.1	282	6.5
No	1,290.00	80.1	829	71.8	1,067.00	66.7	3,186.00	73
Yes	204	12.7	225	19.5	468	29.3	897	20.5
Total	1,611.00	100	1,154.00	100	1,600.00	100	4,365.00	100

86. Does the water you get from your water service provider have a bad odor?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	90	5.6	81	7	24	1.5	195	4.5
No	1,308.00	81.2	861	74.6	1,118.00	69.7	3,287.00	75.3
Yes	212	13.2	212	18.4	461	28.8	885	20.3
Total	1,610.00	100	1,154.00	100	1,603.00	100	4,367.00	100
85. Is the water you get from your water service provider clean?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	84	5.2	77	6.7	17	1.1	178	4.1
No	384	23.8	393	34.1	731	45.6	1,508.00	34.5
Yes	1,143.00	70.9	683	59.2	854	53.3	2,680.00	61.4
Total	1,611.00	100	1,153.00	100	1,602.00	100	4,366.00	100
82. Does your water service provider respond promptly to your complaints about service?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Don't know	589	36.5	285	24.8	607	38.2	1,481.00	34
No	661	41	657	57.2	703	44.2	2,021.00	46.5
Yes	362	22.5	206	17.9	280	17.6	848	19.5

Total	1,612.00	100	1,148.00	100	1,590.00	100	4,350.00	100
80. Good customer service from the people selling you water is important to you?	Cockle Bay		Aberdeen		Kingtom		Total	
	No.	percent	No.	percent	No.	percent	No.	percent
Agree	1,102.00	68.4	734	63.8	738	47.2	2,574.00	59.5
Disagree	141	8.7	129	11.2	375	24	645	14.9
Don't know	81	5	99	8.6	63	4	243	5.6
Partly agree	288	17.9	188	16.3	388	24.8	864	20
Total	1,612.00	100	1,150.00	100	1,564.00	100	4,326.00	100

Illustration of the number of missing values in the CMS survey

CMS Variable (SI variable names)	Number of Missing Values
cs_billpay	173
cs_cost	135
Hhh_sex	337
Hh_size	349
Prop_use	4,042
Gvwc_prevent	2,107
Gvwc_easier	2,109
Storetype	277
Gvwc_account	2,880
Ct_dwell	2,642
Ct_tank	2,642
Ct_tapyard	2,642

Ct_tankqual	4,143
Ct_tankpump	4,144
Metered_yn	2,887

6.8 Annex 8: Approved Evaluation Activities

In consideration of the tradeoffs between cost and learning, not all data collection activities recommend in this report will be exercised. Specifically, data collection in the DMA comparison area beyond the household survey and available GVWC administrative data were not approved. The following table clarifies the final approval status of each activity.

Evaluation	Activity	Area (DMA)	Approval
Performance Evaluation	Desk Review (Data and Reports)		APPROVED
Performance Evaluation	Key Informant Interviews		APPROVED
Performance Evaluation	Process Mapping		APPROVED
Performance Evaluation	Direct Observation		APPROVED
DMA Evaluation	Household Surveys	TREATMENT	APPROVED
DMA Evaluation	ITSA GVWC Administrative Data	TREATMENT	APPROVED
DMA Evaluation	ITSA GVWC Sensor Data	TREATMENT	APPROVED
DMA Evaluation	ITSA GVWC Quality Testing	TREATMENT	APPROVED
DMA Evaluation	Focus Group Discussions	TREATMENT	APPROVED
DMA Evaluation	Key Informant Interviews	TREATMENT	APPROVED
DMA Evaluation	Direct Observation (Kiosks/Public Standpipes)	TREATMENT	APPROVED
DMA Evaluation	Process Mapping (Standpipe Management)	TREATMENT	APPROVED
DMA Evaluation	Household Surveys	COMPARISON	APPROVED
DMA Evaluation	ITSA GVWC Administrative Data	COMPARISON	APPROVED
DMA Evaluation	ITSA SI Sensor Data	COMPARISON	NOT APPROVED
DMA Evaluation	ITSA SI Water Quality Testing	COMPARISON	NOT APPROVED
DMA Evaluation	Focus Group Discussions	COMPARISON	NOT APPROVED
DMA Evaluation	Key Informant Interviews	COMPARISON	NOT APPROVED
DMA Evaluation	Direct Observation (Public Standpipes)	COMPARISON	NOT APPROVED

SOCIAL IMPACT, INC.

2300 CLARENDON BLVD SUITE 1000

ARLINGTON, VA 22201