IMPACT EVALUATION OF THE USAID IUWASH TANGGUH ACTIVITY – EVALUATION DESIGN REPORT

URBAN RESILIENCE BY BUILDING AND APPLYING NEW EVIDENCE IN WASH (URBAN WASH) PROJECT

November 2022

This report is made possible by the support of the American People through the United States Agency for International Development (USAID). The content of this report is the responsibility of Tetra Tech and does not necessarily reflect the views of USAID or the United States government.

Prepared for the United States Agency for International Development by the Urban Resilience by Building and Applying New evidence in WASH (URBAN WASH) project, under the GSA's One Acquisition Solicitation for Integrated Services (OASIS Unrestricted) Indefinite Delivery Indefinite Quantity Contract, contract number GS00Q14OADU138 and order number 7200AA21M00012.

Tetra Tech Contacts:

Miriam Otoo, Deputy Chief of Party
miriam.otoo@tetratech.com

Zach Borrenpohl, Project Manager
zach.borrenpohl@tetratech.com

Tetra Tech
1320 N. Courthouse Road, Suite 600, Arlington VA 22201
Tel: (703) 387-2100, Fax: (703) 414-5593
www.tetratech.com/intdev

Cover photo courtesy of Miguel Albornoz, NORC at the University of Chicago.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

ADS	Automated Directives System
ANCOVA	Analysis of Covariance
BAPPENAS	Ministry of National Development Planning
BPSPAMS	*Badan Pengelola Air Minum dan Sanitasi* (Community Water Supply System)
BC	Business Continuity
BRIN	*Badan Riset dan Inovasi Nasional* (National Research and Innovation Agency)
CAPI	Computer-Assisted Personal Interviewing
COR	Contracting Officer's Representative
DCOP	Deputy Chief of Party
DDL	Data Development Library
DID	Difference-in-Difference
DKI	*Daerah Khusus Ibukota* (Special Capital Region)
DQA	Data Quality Assurance
EDR	Evaluation Design Report
EQ	Evaluation Question
GOI	Government of Indonesia
GPS	Global Positioning System
HWISE	Household Water Insecurity Experiences
IDR	Indonesian Rupiah
IE	Impact Evaluation
IP	Implementing Partner
IRB	Institutional Review Board
IUWASH Tangguh	Indonesia Urban Resilient WASH
JMP	Joint Monitoring Programme
KII	Key Informant Interview
KK	*Kartu Keluarga* (Household Identification)
LG	Local Government
MDES	Minimum Detectable Effect Size
MEL	Monitoring, Evaluation, and Learning

MIS	Management Information System
MoEF	Ministry of Environment and Forestry
MOU	Memorandum of Understanding
NORC	NORC at the University of Chicago
NUWSP	National Urban Water Supply Project
PDAM	*Perusahaan Daerah Air Minum* (Local Water Utility)
PE	Performance Evaluation
PERPAMSI	*Persatuan Perusahaan Air Minum Seluruh Indonesia* (Association of Indonesian Water Supply Companies)
PODES	Village Potential Survey
PUPR	*Kementerian Pekerjaan Umum dan Perumahan Rakyat* (Ministry of Public Works and Housing)
RISPAM	*Rencana Induk Sistem Penyediaan Air Minum* (Water Supply System Master Plan)
RPAM	*Rencana Pengamanan Air Minum* (Drinking Water Safeguard Plan)
RPJMD	*Rencana Pembangunan Jangka Menengah Daerah* (Regional Medium-Term Development Plan)
RPJMN	*Rencana Pembangunan Jangka Menengah Nasional* (National Medium-Term Development Plan)
SDG	Sustainable Development Goal
SOW	Scope of Work
TA	Technical Assistance
UN	United Nations
UNICEF	United Nations Children's Fund
URBAN WASH	Urban Resilience by Building and Applying New Evidence in WASH
USAID	U.S. Agency for International Development
USAID/RFS	USAID's Bureau for Resilience and Food Security
WASH	Water, Sanitation, and Hygiene
WHO	World Health Organization
WRM	Water Resource Management

EXECUTIVE SUMMARY

The U.S. Agency for International Development's (USAID) Urban Resilience by Building Partnerships and Applying New Evidence in Water, Sanitation, and Hygiene (URBAN WASH) program will design and implement an impact evaluation (IE) of the USAID Indonesia Urban Resilient WASH (IUWASH) Tangguh Activity. IUWASH Tangguh ($44.1 million, 2022-2027) provides technical assistance (TA) at the national, provincial, and local level to strengthen water, sanitation, and hygiene (WASH) and water resource management (WRM) sector governance and financing, increase access to safely managed drinking water and sanitation services, improve WRM to support resilient drinking water services, and increase adoption of household behaviors that contribute to improved WASH and WRM services. IUWASH Tangguh will implement different packages of interventions in each of IUWASH Tangguh's 38 participating cities and districts. In each city and district, there will be hotspot[1] neighborhoods that receive particularly intense interventions to support improved WASH and WRM services. Hotspot neighborhoods focused on water supply interventions include households with lower baseline access to safely managed WASH and lower incomes on average than households elsewhere in treatment sites.

The IE will assess IUWASH Tangguh's causal effects on higher-order impacts related to household water security and resilient city-wide water services to generate recommendations for improving the Activity's effectiveness, identify lessons learned for similar future programming, and assess the feasibility of scaling the Activity up to additional geographic regions and populations. The guiding evaluation questions (EQ) are below.

EQ1: How has household water security in the targeted areas changed as a result of the interventions?

EQ2: How have urban water utilities (PDAMs) participating in IUWASH Tangguh and their local government counterparts changed WRM policies and practices in response to the activity? What implications, if any, does this have for the quantity and quality (i.e., availability) of their bulk water supply?[2]

EQ3: How has city-wide water service resilience changed as a result of the interventions?

URBAN WASH proposes a mixed-methods, quasi-experimental design for this IE. The proposed design includes two household-level quasi-experiments focused on household water security in response to EQ1 and one city/district-level quasi-experiment focused on resilient water service delivery for EQ3. The IE will sample household and institutional respondents from 31 pairs of statistically matched treatment and comparison cities and districts.[3] Within the scope of the broader IE, URBAN WASH proposes to use performance evaluation (PE) methods in response to EQ2, which will combine longitudinal analysis of administrative data on bulk water availability for PDAMs in treatment areas and thematic analysis of qualitative interviews with institutional personnel regarding changes in WRM practices perceived to result from the IUWASH Tangguh interventions.

The EQ1 quasi-experiments will assess changes in household water security—operationally defined as access to basic drinking water services, reliability of drinking water services, quantity of water consumed

---

1 "Hotspot" neighborhoods will subsequently be referred to as "community intervention locations" in all future impact evaluation reports and documents.

2 Note this EQ will employ performance evaluation (PE) methods alone, while the first and third EQ will employ IE methods.

3 This includes all treatment cities and districts from the Banten, West Java, Central Java, East Java, North Sumatra, and South Sulawesi provinces. Treatment sites from other provinces are excluded due to lack of comparison candidates.

across all sources, affordability of water consumed across all sources, and quality of drinking water (i.e., absence of fecal contamination). The two experiments are designed to quantify the separate contributions of IUWASH Tangguh's non-hotspot interventions and hotspot interventions on households that reside in hotspot neighborhoods. They accomplish this by statistically matching hotspot neighborhoods with similar neighborhoods elsewhere in treatment sites and comparison sites. These two experiments estimate the impact of IUWASH Tangguh on a specific type of neighborhood (one that resembles a hotspot neighborhood), and not the impact of the activity on the entire city/district population. The resulting estimates of Activity impact should thus be considered a ceiling on overall Activity impact—average treatment effects are likely to be smaller for the full set of households outside hotspot neighborhoods in treatment sites, which will have higher baseline household water security than households in hotspot neighborhoods. The construction of the EQ1 quasi-experimental groups may change over the course of the evaluation depending on the planned expansion of IUWASH Tangguh's hotspot intervention over the course of the Activity.

The EQ3 quasi-experiment will assess changes in risk identification, risk understanding, risk data use, planning for risk mitigation, financing for risk mitigation, and other indicators related to city-wide water service providers' ability to identify, mitigate against, and adapt to hazards to their water service provision. The quasi-experiment will compare these measures for treatment cities and districts with the values for a statistically matched set of comparison cities and districts. The proposed measures combine expert review of water safety planning documentation, survey results with PDAM and local government (LG) officials, and secondary PDAM performance data aggregated by the Ministry of Public Works and Housing. URBAN WASH will separately survey PDAM officials and LG officials for this quasi-experiment to cover institutional responsibilities for all types of water services in urban areas (PDAM, private, and community based).

For all three quasi-experiments, URBAN WASH will test and select the best performing of two possible analysis methods to quantify program impacts: difference-in-difference or analysis in covariance. In addition to measuring impacts on each specific indicator, URBAN WASH proposes to construct and estimate Activity impact on household water security and city-wide resilience indices. These analytical methods both accommodate baseline differences in outcomes of interest between treatment and comparison groups.

URBAN WASH's proposed design will also rely on complementary thematic analysis of endline qualitative interviews with institutional personnel to help explain and contextualize quantitative IE results from the quasi-experiments. The endline evaluation will include a cost-effectiveness analysis to identify the unit cost of activity impacts and assess what it might cost to replicate this activity and its outcomes—for example, if the Government of Indonesia desired to replicate IUWASH Tangguh's TA or if USAID intended to expand the interventions to additional geographies. URBAN WASH will seek opportunities to contextualize evaluation results from a gender perspective, where feasible.

URBAN WASH is collaborating with subcontractor NORC at the University of Chicago (NORC) to design and implement the evaluation and subcontractor Article 33 to collect household, PDAM, and LG survey data. URBAN WASH will collect baseline data from January to March 2023 and complete baseline reporting and dissemination by June 2023. The endline evaluation is planned between October 2025 and September 2026, allowing three years for Activity impacts to materialize. In the interim, URBAN WASH will coordinate with IUWASH Tangguh to monitor Activity implementation and plan any required modifications to the endline evaluation design.

1.0 INTRODUCTION AND BACKGROUND

1.1 EVALUATION BACKGROUND

On February 18, 2022, Tetra Tech's Urban Resilience by Building Partnerships and Applying New Evidence in Water, Sanitation, and Hygiene (URBAN WASH) program received a request from the United States Agency for International Development's (USAID) Indonesia Mission (USAID/Indonesia) to conduct an impact evaluation (IE) of the Indonesia Urban Resilient WASH Tangguh (USAID IUWASH Tangguh) Activity.

From May 17-19, URBAN WASH led a preliminary co-design workshop in Jakarta with representatives of USAID from Indonesia and Washington D.C. and personnel from DAI Global LLC (DAI)—USAID's IUWASH Tangguh implementing partner (IP). The workshop focused on identifying key parameters for the evaluation. In June, URBAN WASH issued a subcontract to NORC at the University of Chicago (NORC) to undertake the design and implementation of the IE.

From June to September 2022, NORC and URBAN WASH facilitated three additional remote co-design sessions to further refine potential designs and measurement approaches for the evaluation. NORC Evaluation Director Miguel Albornoz, Deputy Team Lead Trimo Pamudji, and Senior Urban Water Supply Advisor Risyana Sukarma also participated in a one-week scoping trip from August 22– 26, 2022, which included two in-person remote co-design sessions with the IUWASH Tangguh implementing team along with other scoping activities. Following this trip, URBAN WASH prepared an inception report outlining proposed and alternative evaluation designs, together with requests for additional information that would permit URBAN WASH to propose a final design.

This evaluation design report (EDR) outlines URBAN WASH's proposed design for the IE of IUWASH Tangguh. The proposed IE will span approximately three years, with baseline data collection from January-March 2023 and endline data collection from January-March 2026.

1.2 EVALUATION PURPOSE AND QUESTIONS

The purpose of this IE is to assess the impact of the Activity's interventions on desired higher order impacts including household water security, bulk water availability, and city-wide water service resilience. The guiding evaluation questions (EQ) are enumerated below. Note that the second EQ will rely entirely on performance evaluation (PE) methods within the scope of the broader IE.

> **EQ1:** How has household water security in the targeted areas changed as a result of the interventions?
>
> **EQ2:** How have urban water utilities (*Perusahaan Daerah Air Minum* [PDAM] in Bahasa Indonesia) participating in IUWASH Tangguh and their local government counterparts changed WRM policies and practices in response to the Activity? What implications, if any, does this have for the quantity and quality (i.e., availability) of their bulk water supply?
>
> **EQ3:** How has city-wide water service resilience changed as a result of the interventions?

1.3 INTENDED EVALUATION AUDIENCE AND USES

The primary audience of the IE is USAID/Indonesia. USAID/Indonesia expressed that it would like the evaluation to generate recommendations for improving the Activity's effectiveness, identify lessons learned for similar future programming, and assess the feasibility of scaling up the Activity to additional geographic regions and populations. USAID/Indonesia's Intermediate Result 3.2 posits that increased access to safely managed drinking water services and improved urban resilience to disaster will contribute to strengthening urban environmental management.[4] The IE will provide evidence to quantify IUWASH Tangguh's contribution to these intermediate results, and to support USAID/Indonesia's learning as it continues to partner with the Government of Indonesia (GOI) to pursue sustainable development goals (SDGs). Another primary audience for IE findings is IUWASH Tangguh, who can use IE findings to inform adaptive management and increased effectiveness.

Table 1: Intended Evaluation Audience(s) and Use(s)

AUDIENCE	ACTOR(S)	KEY USE(S)
Primary	USAID/Indonesia	Improve Activity effectiveness, identify lessons learned, and assess the feasibility of taking the Activity to a larger scale.
Primary	IUWASH Tangguh	Promote learning for adaptive management of Activity and any follow-ons.
Secondary	Government of Indonesia	Provide evidence regarding the current state of household water security and climate-resilient WASH and WRM throughout Indonesia, plus considerations for GOI adopting and scaling successful strategies to promote these.
Secondary	Other USAID operating units (e.g., USAID/RFS, USAID/Asia)	Generate evidence which might support USAID assistance in the water sector in other geographies.
Tertiary	Other WASH and WRM sector project implementers and researchers	Contribute to expanded sector knowledge base within Indonesia and internationally.

Secondary intended evaluation audiences include the GOI and other operating units within USAID. Within the GOI, the Ministry of Development Planning (BAPPENAS), the Ministry of Public Works and Housing (PUPR), the Ministry of Health, Indonesian Association of Water Supply Companies (PERPAMSI), and local governments (LGs) throughout Indonesia should find the evaluation particularly useful. USAID's Bureau for Resilience and Food Security (USAID/RFS) and Asia Bureau (USAID/Asia) each have an interest in learning from the IUWASH Tangguh approach and promoting cross-pollination with other geographies where similar USAID programming may be useful.

Finally, the IE will benefit Indonesian and international water, sanitation, and hygiene (WASH) and water resource management (WRM) project implementers and research organizations. Should USAID pursue any follow-on programs to IUWASH Tangguh, these could incorporate lessons learned from the IE in future theories of change and/or results frameworks. Meanwhile, evaluation findings could be shared with local and international academic or practitioner organizations to contribute to expanding the knowledge base of the sector as a whole, especially with the support of URBAN WASH's ongoing role in driving sector learning.

USAID. 2020. "Country Development Cooperation Strategy 2020 - 2025 | Indonesia | U.S. Agency for International Development," Accessed February 24, 2021, https://www.usaid.gov/indonesia/cdcs.

2.0 ACTIVITY DESCRIPTION

2.1 IUWASH TANGGUH ACTIVITY BACKGROUND

USAID IUWASH Tangguh seeks to advance Indonesia's development goals in increasing access to safely managed WASH in vulnerable urban areas and strengthening climate-resilient WASH services and WRM. The approximately $44.1 million Activity—implemented by DAI with a period of performance from April 4, 2022, to April 3, 2027—is a successor to the USAID IUWASH and IUWASH PLUS activities. These activities were also implemented by DAI from 2016-2022. The IUWASH Tangguh team will provide technical assistance (TA) to GOI, private sector, and civil society stakeholders to achieve four Objectives, shown in Figure 1.[5]

Figure 1: IUWASH Tangguh Objectives

IUWASH Tangguh interventions operate at the national, provincial, and local levels in 38 cities and districts in 10 USAID priority provinces.[6] The Activity will pursue improvements to governance and the enabling environment at each of the national, provincial, and local (i.e., city/district) levels. Simultaneously, in partnership with LGs and PDAMs, IUWASH Tangguh will use diagnostic tools to assess shortcomings in governance and utility performance relevant to WASH and WRM in each city and district. The Activity will tailor interventions in each city and district to the diagnosed shortcomings, as appropriate to relevant stakeholders' baseline capabilities and interests. Select cities and districts will receive "full support" for safely managed drinking water supply, safely managed sanitation, and climate resilient WRM interventions. Other cities and districts—depending on capacity, resources, or needs—may only receive a sub-set of this support. The specific TA provided will vary from site to site based on the site's baseline capabilities and needs.[7]

---

[5] DAI Global, LLC, 2016. "USAID/Indonesia Urban Resilient, Sanitation, and Hygiene (IUWASH Tangguh): Project Year 1 Work Plan" (USAID/Indonesia, July 2, 2022).

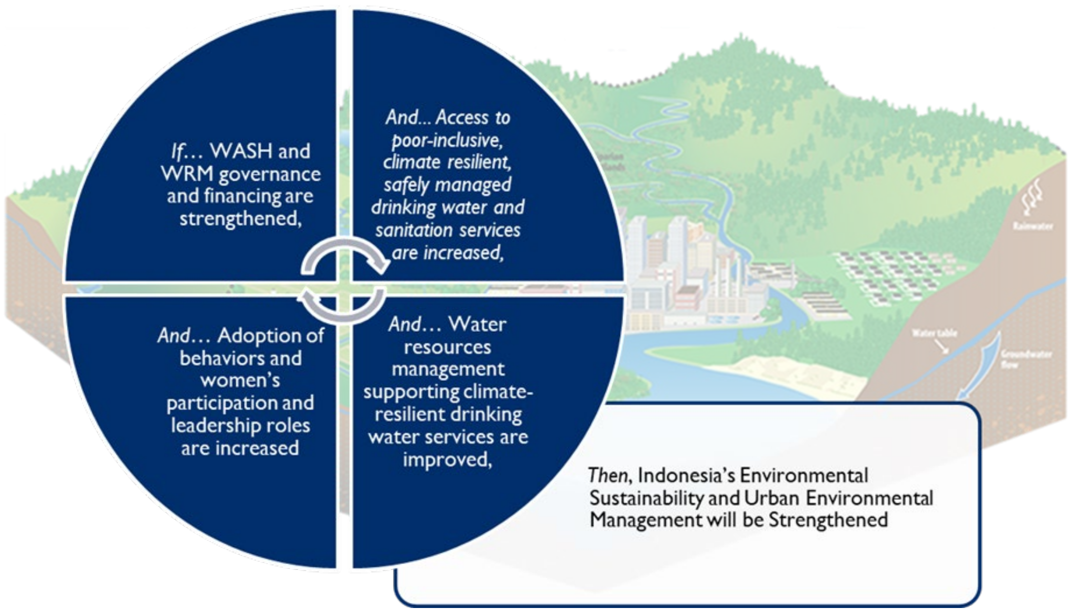
[6] Note that the districts included among IUWASH Tangguh's implementation sites are "*kabupaten*." *Kabupaten* is sometimes translated to English as "regency" instead of "district." The evaluation team uses the term "districts" in this report to mirror IUWASH Tangguh's own language. Cities ("*kota*") and districts ("*kabupaten*") are at an equivalent administrative level in Indonesia, though cities are larger and more population dense than districts as a rule.

[7] See Annex 2: Additional Background Information

2.2 ACTIVITY THEORY OF CHANGE AND KEY CAUSAL PATHWAYS

The theory of change presented in IUWASH Tangguh's Project Year 1 Work Plan (Figure 2) demonstrates the relationship between the project's four main Objectives and the goal of strengthening Indonesia's environmental sustainability and urban environmental management.

Figure 2*: *USAID IUWASH Tangguh Theory of Change



IUWASH Tangguh's results framework (Figure 3) details the interventions, outputs, outcomes, and anticipated impacts from the Activity. The following sections describe key causal pathways through which IUWASH Tangguh interventions could affect household water security, bulk water availability, and city-wide water service resilience—the three higher order impacts on which the EQs focus. These causal pathways provide the bases for causal hypotheses for each EQ that the evaluation can either validate or refute.

As of November 2022, IUWASH Tangguh is still refining its specific implementation approaches to support increased bulk water availability for PDAMs in treatment cities and districts. Until this implementation approach is finalized, URBAN WASH cannot describe causal pathways for IUWASH Tangguh to support increased bulk water availability to inform an evaluation design for EQ2.

Figure 3: USAID IUWASH Tangguh Results Framework

Interventions	Outputs	Outcomes	Impact	
Objective 1: Strengthened WASH and WRM sector governance and financing			Increase access to safely managed drinking water, sanitation and hygiene (WASH) in Indonesia's vulnerable urban areas and strengthening climate-resilient WASH services and water resources management	**DO3** Environmental Stability Improved **IR 3.2** Urban Environmental Management Strengthened
• Provide technical support to LGs, CSOs, and user groups to advocate for revisions of WASH/WRM legislation • Develop financial flow tracking system and pipeline of potential WASH/WRM investment opportunities • Build sector capacities through mentorship, training, technical exchanges, and tech enhancements • Create/expand mechanisms to promote social accountability and inclusivity	• **65** new and/or revised laws, policies, regulation, or agreements in place that promote access to improved WASH services • **50** PPP established with USG support • **35** social inclusion and public accountability measures implemented	• **$300** million mobilized to the water and sanitation sectors with USG assistance (HL.8.4-1) • **20%** increase to overall budget appropriations for WASH, WRM by targeted local governments • **100** institutions strengthened to manage the urban environment		
Objective 2: Increased access to safely managed drinking water and sanitation services				
• Implement workforce development framework for service providers • Assist BAPPENAS, MoPWH, and MoHA to further develop the PDAM data reporting system with data-driven KPIs, targets and definitions • Expand WASH and WRM service models through a consultation process with communities and potential • Integrate data systems for use by service providers and LGs	• **5,000** WASH/WRM workers with improved skills • **10** new financial/operational management tools developed by service providers	• **1.5 million** people gaining access to safely managed drinking water services with USG assistance (HL.8.1-2) • **1 million** people gaining access to safely managed sanitation services with USG assistance (HL.8.2-3) • **100** water and sanitation sector institutions strengthened to manage water resources or improve water supply and sanitation services with USG assistance (HL.8.3-3)		
Objective 3: Improved WRM to support climate-resilient drinking water services				
• Conduct stakeholder consultations, public-private dialogues, community outreach, and youth education to raise awareness of climate change • Collect data on the climate-resiliency of watersheds • Prepare climate resilient water resource vulnerability assessments and action plans for PDAMs • Promote private sector investment to protect natural assets and invest in green infrastructure • Implement protection measures for watershed catchment areas • Integrate GESI into WRM policies and budgeting • Collect and share WRM, ground/surface water data and climate information • Develop climate vulnerability assessment reports	• **60** WRM policies, guidance, and programs to support resilient drinking water services • **40** institutions with management information systems and/or data management tools adopted as a result of USG assistance • **35** climate vulnerability assessment reports produced to support improvement of raw water sources of assisted PDAM	• **$10 million** investment mobilized for climate change adaptation as supported by USG assistance (EG.11-4 and NRSS CC1.2) • **100** institutions with improved capacity to assess or address climate change risks supported by USG assistance (EG.11-2 and NRSS CC1.1) • **300,000** people benefiting from adoption and implementation of measures to improve water resources management as a result of USG assistance (HL.8.5.1 and NRSS O3.5)		
Objective 4: Increased adoption of behaviors and improved women's participation & leadership roles in WASH and WRM				
• Implement SBC strategies targeted to sustainable sanitation systems and payments for water and sanitation services • Publish a gender analysis report • Design Inclusive WASH facilities • Implement strategy to accelerate women's participation and agency in WASH and WRM institutions • Coach and train women to manage WASH facilities • Establish GRBP regulations and KPIs	• **5,000** persons trained with USG assistance to advance outcomes consistent with gender equality or female empowerment through their roles in public or private sector institutions or organizations (GNDR-8) • **1 million** people reached through SBC campaigns promoting WASH and WRM resilience • **35** cities/districts implement improved desludging services	• **30%** of households with soap and water at a handwashing station commonly used by family members (HL.8.2-5) • **20%** of participants reporting increased agreement with the concept that males and females should have equal access to social, economic, and political resources and opportunities (GNDR-4)		

2.2.1 KEY CAUSAL PATHWAYS: HOUSEHOLD WATER SECURITY (EQ1)

The IUWASH Tangguh results framework emphasizes access to safely managed drinking water as the chief household-level impact related to water supply. This aligns with the GOI's commitments to the United Nations (UN) 2030 Agenda, which encourages all UN member countries to reach 100 percent coverage of safely managed drinking water for their citizens by 2030.[8] Safely managed drinking water is defined as an improved water source which is located on premises, available when needed, and free of fecal and priority chemical contamination.[9] Beginning with the 2015-2019 National Medium-Term Development Plan (RPJMN), the GOI expanded this commitment to say that safely managed drinking water should meet minimum standards for quantity, quality, continuity, and affordability.[10] This expansion aligns with developing international commitments to "household water security" as a minimum level of access to water required for household wellbeing.

There are two separate pathways through which the Activity seeks to improve household water security. The first is by improving the quantity, quality, reliability, and/or affordability of water supply for households that are already connected to PDAM piped networks (i.e., "improved quality of water services from PDAMs"). The second pathway is promoting access to piped PDAM water for households that do not currently have it (i.e., "increased access to basic water services").

Figure 4 depicts the specific links between the activities, outputs, and outcomes for the Objectives in the results framework and improved household water security through improved quality of water services and increased access to basic water services, described further in the remainder of this section.

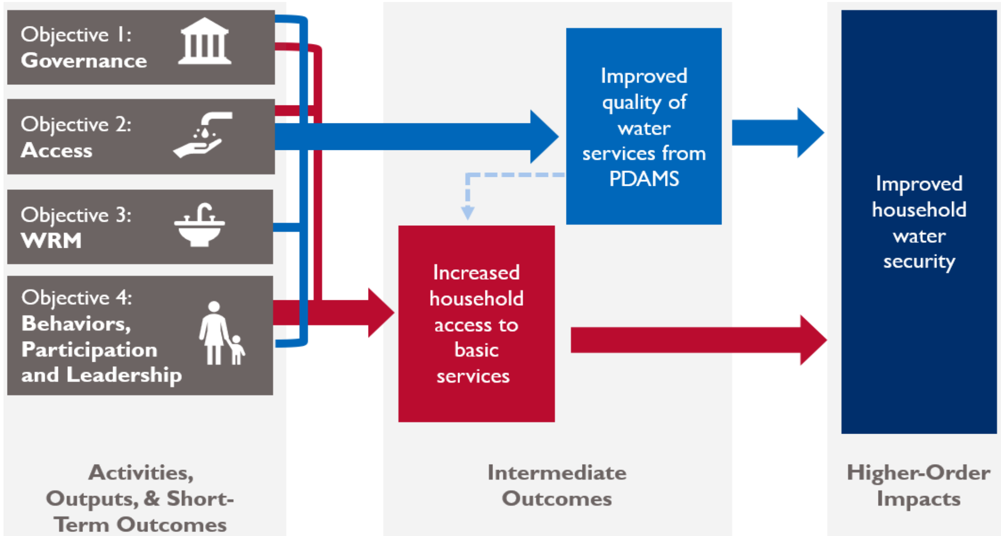
---

8 UN Department of Economic and Social Affairs, "Transforming Our World: The 2030 Agenda for Sustainable Development," accessed October 20, 2022, https://sdgs.un.org/2030agenda.

9 'Improved' sources are those that are potentially capable of delivering safe water by nature of their design and construction. These include piped water, boreholes or tubewells, protected dug wells, protected springs, and rainwater. World Health Organization and United Nations Children's Fund (UNICEF), *Safely Managed Drinking Water: Thematic Report on Drinking Water 2017* (Geneva: World Health Organization, 2017), https://apps.who.int/iris/handle/10665/325897.

10 These are referred to as the "4K" principles for their Bahasa Indonesia translation: *Kuantitas*, *Kualitas*, *Kontinuitas*, and *Keterjangkauan*. Specific standards include a minimum water supply of 40 liters per capita per day, which is free from priority fecal and chemical contamination and available 24 hours per day, at a cost of no more than 4 percent of total household income. "Presidential Regulation of The Republic of Indonesia Number 2 of 2015 About Medium-Term Development Plan (RPJMN) 2015-2019"

Figure 4: Causal Pathways for Improved Household Water Security



IUWASH Tangguh interventions should contribute to improved quality of water services in all four of the Activity's Objectives. TA under Objective 2 should enhance the capacity of WASH and WRM personnel and the performance of WASH and WRM institutions, which should contribute to improving water services. Objective 1 interventions create a favorable legal environment for WASH and WRM service provision. Objective 3 interventions strengthen water services through improved water quality, quantity, and reliability in watersheds which supply raw water for drinking water services.[11] Finally, Objective 4 interventions strengthen water services by increasing gender integration and women's agency in WASH and WRM sectors.

The 2020-2024 RPJMN acknowledges that although 20.1 percent of households in Indonesia had a piped connection in 2018, only 6.8 percent had a water supply that was adequate in terms of quantity, safe to drink, and reliable.[12] For households with access to a piped PDAM connection—some of which are nonetheless water insecure—a more favorable legal and regulatory environment, more sufficient financial resources, better WRM, and improved WASH services from improved and increasingly gender-integrated PDAMs should improve the quantity, quality, and reliability of service without becoming unaffordable.

Where IUWASH Tangguh's interventions to promote improved quality of water services target change among institutional actors, the Activity also includes household-level interventions to promote increased household access to basic services. Specifically, under Objective 4, it seeks to promote households' demand and willingness to pay for water services. However, some interventions from Objectives 1 and 2 will also contribute to increased access to basic services at the institutional level by attracting investment for or promoting service models that favor connections by new PDAM customers.

Access to a piped water supply with a tap on the household premises is one of the core components of household water security because sources that are further away and/or categorized by the UNICEF/World Health Organization (WHO) Joint Monitoring Programme (JMP) as "unimproved" carry

---

[11] The evaluation team refers to this concept as "improved bulk water availability" in this report

[12] Republic of Indonesia. 2020. "Appendix Presidential Regulation No. 18 of 2020 Concerning the National Medium-Term Development Plan For 2020-2024" (Republic of Indonesia, 2020), https://perpustakaan.bappenas.go.id/e-library/file_upload/koleksi/migrasi-data-publikasi/file/RP_RKP/Narasi-RPJMN-2020-2024-versi-Bahasa-Inggris.pdf.

inherent risks associated with the quantity, quality, or reliability of supply.[13] Access to a new piped water connection should improve overall household water security, even absent improvements in the quality of PDAM water services, if households' previous sources provided less, lower quality, or less reliable water than the PDAM provides. Therefore, households that obtain access to a new piped connection may experience greater improvement in household water security than households for which IUWASH Tangguh interventions improve the quality of existing piped water services. The potential impact of increased access to basic services on household water security may also occur more quickly than the impact of improved quality of water services. It is reasonable to expect improved quality of water services to occur gradually, as upstream outputs related to PDAM performance and financing gradually improve the quantity, quality, reliability, and affordability of water delivered to customers. On the other hand, the benefits of increased access would occur as soon as a household connects, provided that the quantity, quality, reliability, and affordability of the PDAM network exceeds that of their previous source(s).[14]

Both the "improved quality of water services" and "increased access to basic services" causal pathways will operate at different intensities in different portions of the PDAM service areas in which IUWASH Tangguh intervenes. Some components of the Activity (e.g., improved legal/regulatory environment, capacity building for front office and water treatment plant staff, improved WRM policies and practices, etc.) should result in improved quality of water services and increased access to basic services throughout the entire PDAM service area. However, the intervention will also identify "hotspot neighborhoods" (*kelurahan* in Bahasa Indonesia) that have a relatively larger proportion of low-income households and, likely, lower coverage rates of piped water connections.[15] In these neighborhoods, in addition to the interventions that should promote each causal pathway throughout the PDAM service area, IUWASH Tangguh will pursue targeted TA to improve the quality of the water services and increase access. IUWASH Tangguh's social behavior change campaigns to promote demand for piped water will target these neighborhoods specifically. These neighborhoods will likely experience greater improvements in water security than the PDAM service area on average due to the intensity of the treatment and the likelihood that they started with lower baseline water security than other neighborhoods. There will be up to two hotspot neighborhoods in each city and district focused on improved water services, though this number will increase continuously over the course of activity implementation, such that more and more of each treatment city and district participates in the hotspot intervention by the end of the activity.

2.2.2 KEY CAUSAL PATHWAYS: CITY-WIDE RESILIENT WATER SERVICE (EQ3)

This IE defines resilience of city-wide water service as reducing the incidence and duration of disruptions to the water supply caused by shocks and stressors. Increasing resilience requires promoting an improved ability to address and reduce risk and increased adaptive capacity on the part of institutions responsible for water service provision at the city level. This section highlights the link between IUWASH Tangguh activities and outputs and improvements in how water service providers address and

---

[13] Guy Howard et al., "Domestic Water Quantity, Service Level and Health: Second Edition" (World Health Organization, 2020), https://www.who.int/publications/i/item/9789240015241.

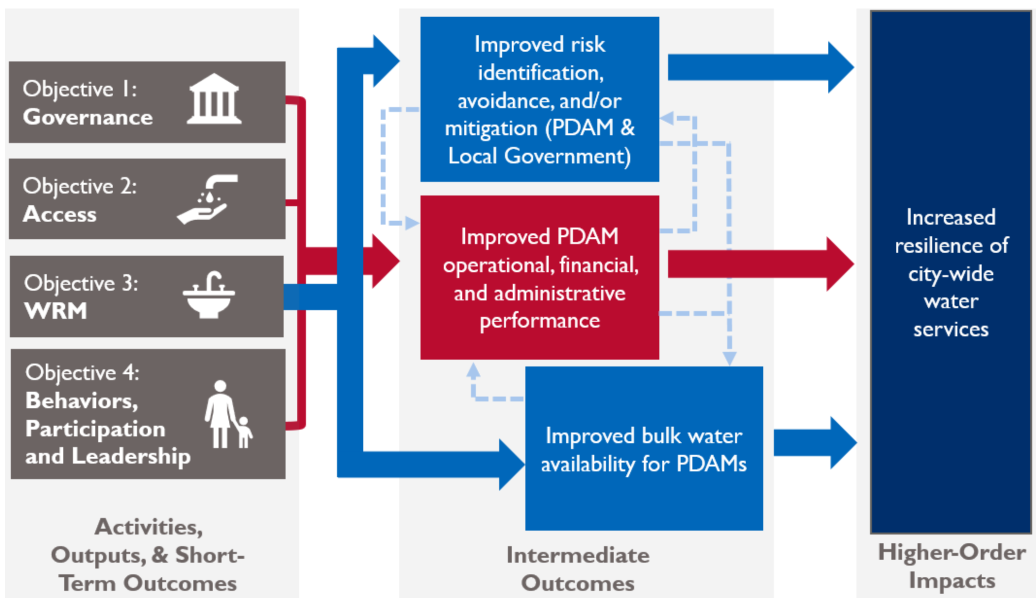
[14] This is an assumption underlying the theory of change which has a reasonable basis in literature cited from the WHO/JMP, but which would significantly reduce potential program impacts if it does not hold (i.e., water from the PDAM is actually less reliable, lower quality, more expensive, etc. than water from alternative sources).

[15] Some of IUWASH Tangguh's hotspot neighborhoods will focus on WRM interventions in the catchment area or sanitation interventions. This section refers to the subset of IUWASH Tangguh's hotspot neighborhoods which focus on water supply. Though these neighborhoods are expected to be lower income and have lower access to basic WASH services than the city average, they will not necessarily be the very poorest neighborhoods in the city/district.

reduce risks to water service disruptions as well as improvements in service providers' capacity to adapt to disruptions as they occur.

Figure 5 depicts the specific links between the activities, outputs, and outcomes for the Objectives in the IUWASH Tangguh results framework and improved city-wide water service resilience, which is described in the remainder of this section. IUWASH Tangguh's interventions to improve WRM under Objective 3 primarily address (i) improving risk identification, avoidance, and/or mitigation related to shocks and stressors associated with climate change and (ii) improving the quality, quantity, and reliability of PDAM raw water sources (i.e., "bulk water availability"). However, interventions across all other Objectives meant to improve PDAMs' operational, financial, and administrative performance are also expected to contribute to increased resilience by strengthening the quality of water service prior to shocks and stressors occurring, and thus reducing the probability that service might be disrupted when a shock or stressor challenges the water supply (whether caused by climate change or not). These three causal pathways are likely to have indirect reinforcing effects, as demonstrated by the dashed blue arrows in Figure 5.

Figure 5: Causal Pathways for City-Wide Resilient Water Supply



Within the Indonesian institutional framework, city and district governments are ultimately accountable for water service provision within their administrative boundaries. There is no single agency responsible for water service provision within LGs. Instead, there is typically a functional working group (*Pokja* in Bahasa Indonesia) that comprise various line agencies in the government with responsibilities for specific aspects of water service provision.[16]

The LG's degree of involvement in water service provision depends on the source of water. PDAMs are city/district-owned enterprises which provide piped water to paying customers in the city or district. While PDAMs are mostly managed and financed independently, they are accountable to meeting LG standards for water service provision and are often partially dependent on LG investment. A LG official

[16] These include representatives from the city government agencies for the Environment and Forestry (*Dinas Lingunan Hidap*, responsible for raw water quality), Health (*Dinas Kesehatan*, responsible for household water quality), Public Works (*Dinas Pekerjaan Umun*, responsible for physical infrastructure), and Planning (*Bappeda*).

typically sits on each PDAM's supervisory board. For private water sources (e.g., dug wells, boreholes, springs, surface water, bottled water providers, etc.), LGs' role in service provision is typically limited to monitoring water quality. Finally, LGs may also oversee BPSPAMS—community-owned organizations, which fulfill a similar role to PDAMs for the operation and management of community water supply systems. Although these organizations are rare in cities, they occur in some districts with substantial rural areas.

IUWASH Tangguh's interventions most directly target PDAMs. Therefore, the Activity should contribute to greater improvements in the resilience of PDAM water service provision than in the resilience of water supplied from BPSPAMS or private sources. Under Objective 3, IUWASH Tangguh will produce climate vulnerability assessments for PDAMs. The assessments will provide PDAM decision makers with localized information on how projected climate change interacts with the characteristics of their catchment, water supply infrastructure, and customer base. This analysis will identify the risks PDAMs face to their ability to provide continuous, high-quality water that meets demand without exhausting their water sources. Additionally, IUWASH Tangguh will provide TA to build the capacity of PDAM management to integrate data on risks into decision making. The Activity will further support this decision making by providing management information systems with real-time data on water availability and ongoing climate risks. Beyond identifying risks and integrating data on risks in PDAM decision making, IUWASH Tangguh will also provide TA to remediate identified risks to PDAM water services. Specific actions intended to remediate risk will include, but are not limited to:

1) TA and advocacy to ensure that the design, budgeting, and implementation of PDAM water safety and business continuity plans incorporate components to manage risk at the source, operational, and consumer levels;
2) Support to find investment to prevent or reduce risk, including to identify new water sources; and
3) Direct intervention with communities, governments, and landscapes in catchments to improve the quality and quantity of water in existing sources (i.e., improve bulk water availability).[17]

IUWASH Tangguh will share the PDAM climate vulnerability assessments with the city and district governments to increase their access to information on the risks posed by climate change to water sources in the area. The Activity will also provide some limited TA to ensure that *Pokjas* set targets associated with resilient water services in their medium- and long-term development plans (RPJMN and RPJMD) and that they incorporate climate data and hazards to their water services in their decision making and budgeting. However, the benefits to water service resilience from these interventions would not be restricted to PDAM water services—they could also improve how risks are identified and mitigated for private or community water services. However, IUWASH Tangguh will not provide TA to LGs to mitigate risks to non-PDAM water services like it will to the PDAMs. There could be indirect effects on non-PDAM water services from interventions that target the PDAM catchment (for example, LGs could use the data on water availability for raw water sources shared by PDAMs and community supplies, or benefit from groundwater recharge for aquifers shared by PDAMs and private users), but these would be much smaller in magnitude than the effects on the PDAM water services.

Finally, some of the same interventions that contribute to improved water services, like TA to improve PDAM operational performance, could also improve bulk water production and provision. This improvement could result in fewer/shorter bulk water production disruptions occurring or perhaps

---

17 While this causal pathway is important in the long -term, in practice IUWASH Tangguh does not expect this impact to occur within the evaluation timeframe.

needing larger shocks or stressors to disrupt water services. Additionally, improved financial performance and attraction of external investment can ensure that PDAMs have sufficient resources to recover from disruptions that occur, and increased access to basic services could move households from less resilient sources to more resilient sources without any change in their awareness of risks or capacity for adaptive behavior. For example, a household who previously used a shallow well may have experienced a disruption in their water supply from drought, but instead will retain a continuous supply during drought when they switch to piped water.

3.0 EVALUATION DESIGN

3.1 EVALUATION DESIGN SUMMARY

URBAN WASH proposes a mixed-methods, quasi-experimental design for this IE. The proposed design includes two household-level quasi-experiments focused on household water security in response to EQ1 and one city/district-level quasi-experiment focused on resilient water service delivery for EQ3. Each component of this design will draw its sample from the same set of treatment and comparison cities and districts, described later in this section (though one component, the incremental hotspot quasi-experiment, focuses entirely within treatment cities and districts). URBAN WASH's design will also rely on complementary thematic analysis of endline qualitative interviews with institutional personnel to help explain and contextualize quantitative IE results from the quasi-experiments. The endline evaluation will include a cost analysis to identify the unit cost of Activity impacts based on overall Activity costs.

Within the broader IE, URBAN WASH proposes to use performance evaluation (PE) methods in response to EQ2, which will combine longitudinal analysis of administrative data on bulk water availability for PDAMs in treatment areas and thematic analysis of qualitative interviews with institutional personnel regarding changes in WRM practices perceived to result from the IUWASH Tangguh intervention. Table 2 below summarizes URBAN WASH's proposed evaluation design.

Table 2: Evaluation Design Summary Matrix

IDENTIFICATION STRATEGY	ANALYSES	INDICATOR(S)	KEY DATA SOURCE(S)*
EQ1: Household Water Security			
Statistical matching at city/district and neighborhood levels for: 1.1 Households in hotspot neighborhoods compared to non-hotspot neighborhoods in treatment cities/districts 1.2 Households in non-hotspot neighborhoods in treatment cities/districts compared to households in similar neighborhoods in comparison cities/districts	▪ Difference-in-difference (DID) or analysis in covariance (ANCOVA) (IE analysis) ▪ Thematic analysis (explanatory qualitative analysis) ▪ Cost analysis	Custom Household Water Security Index comprising: 1. Access to improved source on premises 2. Days per most recent week where main drinking water source was disrupted 3. Liters per capita per day of water consumption 4. Percent of total household expenditure spent on water 5. Presence/absence of E. Coli at point of consumption	▪ Household survey, with integrated water quality testing ▪ Key informant interviews (KIIs) with institutional personnel ▪ IUWASH Tangguh cost data
EQ2: WRM and Bulk Water Availability			
N/A	▪ Longitudinal data analysis (quantitative) ▪ Thematic analysis (qualitative)	Bulk water availability indicators (TBD)	▪ PDAM MIS data on bulk water availability ▪ KIIs with institutional personnel
EQ3: City-wide Water Service Resilience			

IDENTIFICATION STRATEGY	ANALYSES	INDICATOR(S)	KEY DATA SOURCE(S)*
Statistically matched treatment and comparison cities and districts	▪ Difference-in-difference or ANCOVA (IE analysis) ▪ Thematic analysis (explanatory qualitative analysis) ▪ Cost analysis	Custom resilience Index comprising: 1. Risk identification 2. Risk understanding 3. Risk data use 4. Planning for risk mitigation 5. Finance for risk mitigation 6. Other indicators specific to PDAMs and LGs	▪ PDAM survey ▪ LG survey ▪ Document review ▪ PUPR annual PDAM performance data ▪ KIIs with institutional personnel ▪ IUWASH Tangguh cost data

*Throughout the evaluation, URBAN WASH will also rely on IUWASH Tangguh Activity monitoring and evaluation data and secondary datasets from GOI sources, where available, to support explanatory analysis.

3.1.1 SELECTION OF TREATMENT AND COMPARISON CITIES AND DISTRICTS

IUWASH Tangguh selected a set of 38 cities and districts (referred to for this evaluation as treatment sites) for the Activity's interventions based on three core considerations. First, the Activity aggregated data on objective criteria related to WASH service provision and access, watershed conditions, government commitment and capacity, and socioeconomic vulnerability and equity, which resulted in a long list of potential treatment sites. IUWASH Tangguh's site selection report describes how the long list prioritized large population centers with potential for improvement on the site selection criteria that also had nearby cities or districts which provided opportunities for coordination in WRM and cost effectiveness in Activity implementation.[18] Second, the Activity team completed a subjective analysis of this list to identify a short list of sites that they believed would be capable of meeting Activity performance targets. Finally, the Activity team worked with BAPPENAS and LGs in each of these sites to determine whether there was adequate local commitment and made minor alterations to the short list where government commitment was suboptimal.

IE designs must construct a counterfactual to estimate program impacts. The "counterfactual" is the hypothetical state of outcomes of interest that would have existed for the treatment group had the intervention never occurred. Randomly assigning intervention sites to treatment and control groups (i.e., an experimental design) produces the most valid and rigorous counterfactual, since it virtually guarantees balance in characteristics between treatment and control groups. However, random assignment is not possible when the treatment group has already been selected prior to the evaluation as in the case of IUWASH Tangguh.

As an alternative approach, URBAN WASH conducted statistical matching analysis to identify a set of comparison cities and districts that most closely resemble treatment cities and districts on observable characteristics associated with both treatment selection and intended outcomes.[19] Most of the indicators selected are related to characteristics or performance of the PDAMs serving the cities and districts, though there are some others related to the cities and districts' sociodemographic characteristics. The matching characteristics are important because, so long as they are equivalent between the two groups, there are very few rival explanations for observed changes in outcomes of

---

18 IUWASH Tangguh Site Selection Report, pgs. 4-5

19 Note that the appropriate verbiage switches from "control" sites to "comparison" sites when non-random (i.e., quasi-experimental) methods are deployed to select the groups.

interest.[20] Using IUWASH Tangguh's site selection dataset and historical data from the Ministry of Public Works and Housing (PUPR) on PDAM characteristics and performance, URBAN WASH selected the variables in Table 3 for inclusion in the statistical matching exercise. URBAN WASH matched cities and districts on these variables using a genetic matching algorithm without replacement (i.e., one comparison site can match to only one treatment site), explained in further detail in Annex 3.

Table 3: Statistical Matching Variables, Cities, and Districts

VARIABLE	UNITS	SOURCE	MATCHING CRITERION
Urban area classification (city or district)	Categorical	IUWASH Tangguh	Exact
Province	Categorical	IUWASH Tangguh	Distance (Mahlanobis)
Poverty rate	Percent	IUWASH Tangguh	Distance (Mahlanobis)
Households with access to improved sanitation	Percent	IUWASH Tangguh	Distance (Mahlanobis)
Domestic customer coverage rate for PDAM, 2020	Percent	PUPR	Distance (Mahlanobis)
Population in the PDAM working area, 2020	People	PUPR	Distance (Mahlanobis)
Average tariff rate for PDAM customers, 2020	Rupiah/m^3	PUPR	Distance (Mahlanobis)
Solvency of the PDAM, 2020	Rupiah	PUPR	Distance (Mahlanobis)
Ratio of local government contribution to total assets for PDAM, 2020	Ratio	PUPR	Distance (Mahlanobis)
PDAM receipt of World Bank National Urban Water Supply Project (NUWSP) investment, 2020	Binary (0/1)	World Bank	Distance (Mahlanobis)
PDAM production volume, 2020	Meters^3/year	PUPR	Distance (Mahlanobis)
PDAM transmission pipe length, 2020	Meters	PUPR	Distance (Mahlanobis)
PDAM water loss rate (i.e., non-revenue water), 2020	Percent	PUPR	Distance (Mahlanobis)
Overall PDAM performance score for the 2020 fiscal year, 2020	Points	PUPR	Distance (Mahlanobis)
PDAM customer growth rate year over year, 2020	Percent	PUPR	Distance (Mahlanobis)
PDAM operating hours as proportion of 24-hour day, 2020	Percent	PUPR	Distance (Mahlanobis)
Volume of water abstracted from surface water sources	Liters/second	PUPR	Distance (Mahlanobis)
Volume of water abstracted from spring sources	Liters/second	PUPR	Distance (Mahlanobis)
Volume of water abstracted from groundwater sources	Liters/second	PUPR	Distance (Mahlanobis)

URBAN WASH's matching algorithm selected 31 pairs of cities and districts for inclusion in the study, presented in Table 4. This selection includes all of IUWASH Tangguh's treatment sites from the Banten, East Java, Central Java, West Java, South Sulawesi, and North Sumatra provinces and excludes its treatment sites from DKI Jakarta, Papua, East Nusa Tenggara, and West Kalimantan, which do not have

[20] As discussed later, characteristics which explain selection into treatment and outcomes that are not observable can be one alternative explanation for observed changes in outcomes of interest.

reasonably similar cities or districts to serve as comparisons for the evaluation. Prior to matching, the analysis finds that treated cities and districts have lower poverty rates, higher access to improved sanitation, and better performing PDAMs serving larger populations than untreated cities and districts. Over half of PDAMs in treated cities and districts received investment from the World Bank NUWSP program, compared to less than one fifth of those in untreated cities and districts. URBAN WASH's matching exercise substantially reduces most of these differences, though some differences remain and must be controlled for in the analysis of program impact.

Although all comparison sites come from the same set of provinces as the treatment sites, there are relatively more cities and districts from West Java and South Sulawesi in the comparison group and relatively fewer in Banten, East Java, and North Sumatra than in the treatment group. Annex 3 presents detailed output from URBAN WASH's statistical matching exercise. The designs specified for each of the EQs use the same set of treatment and comparison sites.

Table 4: Treatment and Comparison Cities and Districts[21]

PAIR ID	TREATED SITE	COMPARISON SITE	PROVINCE(S)
1	Kabupaten Wonogiri	Kabupaten Pati	Jawa Tengah
2	Kota Binjai	Kota Mojokerto	*Sumatera Utara/Jawa Timur*
3	Kota Blitar	Kota Semarang	*Jawa Timur/Jawa Tengah*
4	Kota Depok	Kota Bekasi	Jawa Barat
5	Kota Magelang	Kota Probolinggo	*Jawa Tengah/Jawa Timur*
6	Kota Makassar	Kota Parepare	Sulawesi Selatan
7	Kota Malang	Kota Sibolga	*Jawa Timur/Sumatera Utara*
8	Kota Medan	Kota Tebingtinggi	Sumatera Utara
9	Kota Pasuruan	Kota Pekalongan	*Jawa Timur/Jawa Tengah*
10	Kota Pematangsiantar	Kota Palopo	*Sumatera Utara/Sulawesi Selatan*
11	Kota Salatiga	Kota Tegal	Jawa Tengah
12	Kota Surabaya	Kota Bogor	*Jawa Timur/Jawa Barat*
13	Kota Surakarta	Kota Bandung	*Jawa Tengah/Jawa Barat*
14	Kota Tangerang	Kota Banjar	*Banten/Jawa Barat*
15	Kabupaten Bogor	Kabupaten Bandung	Jawa Barat
16	Kabupaten Deli Serdang	Kabupaten Asahan	Sumatera Utara
17	Kabupaten Gowa	Kabupaten Sinjai	Sulawesi Selatan
18	Kabupaten Gresik	Kabupaten Magetan	Jawa Timur
19	Kabupaten Karanganyar	Kabupaten Magelang	Jawa Tengah
20	Kabupaten Malang	Kabupaten Banyuwangi	Jawa Timur
21	Kabupaten Maros	Kabupaten Luwu Utara	Sulawesi Selatan
22	Kabupaten Pasuruan	Kabupaten Mojokerto	Jawa Timur
23	Kabupaten Sidoarjo	Kabupaten Bojonegoro	Jawa Timur
24	Kabupaten Simalungun	Kabupaten Langkat	Sumatera Utara
25	Kabupaten Barru	Kabupaten Toraja Utara	Sulawesi Selatan
26	Kabupaten Sragen	Kabupaten Kendal	Jawa Tengah
27	Kabupaten Sukoharjo	Kabupaten Demak	Jawa Tengah

21 Rows which are italicized in the province column match cities or districts in different provinces. URBAN WASH attempted an alternate specification which enforced matching within province. However, this method was not selected because it required either dropping three unmatched treatment sites or accepting very poor matches for these sites. The selected method performs best in terms of overall balance (average mean standardized difference is less than 0.19) and retains the most treatment sites in the study.

PAIR ID	TREATED SITE	COMPARISON SITE	PROVINCE(S)
28	Kabupaten Takalar	Kabupaten Luwu Timur	Sulawesi Selatan
29	Kota Tangerang Selatan	Kota Kediri	*Banten/Jawa Timur*
30	Kabupaten Tangerang	Kabupaten Pandeglang	Banten
31	Kabupaten Temanggung	Kabupaten Batang	Jawa Tengah

URBAN WASH is in the process of securing agreement from the proposed comparison sites to participate in the study (see Section 5.1 for an associated timeline) and it is possible that some of these sites will not agree to participate. Once URBAN WASH knows which sites decline to participate, it will re-run its statistical matching algorithm exclusively for the treatment sites that have lost their match. Existing comparison sites will be excluded from this second matching exercise (i.e., each comparison site can match to only one treatment site). URBAN WASH will begin data collection with the set of comparison sites who initially agree to participate and will sequence data collection with replacement sites later in the data collection calendar. In the baseline report, URBAN WASH will present updated outputs from its statistical matching exercise with the final set of treatment and comparison sites, including replacements.

3.2 EQ1 (HOUSEHOLD WATER SECURITY) DESIGN

3.2.1 HYPOTHESES OF INTEREST

As discussed in Section 2.2, the IUWASH Tangguh theory of change purports to increase household water security by improving the quality of existing water service to the piped PDAM network and increasing access to basic water services for households who connect to the piped PDAM network for the first time. Each of these causal pathways exists throughout the whole PDAM service area and in hotspot neighborhoods alike, but they operate at different intensities. The IE separately tests two hypotheses related to these causal pathways.

PDAM Service Area Hypothesis

If IUWASH Tangguh TA improves PDAM capacity and performance related to WASH and WRM service provision, *then* households throughout the PDAM service area will experience improvements in the quantity, quality, reliability, and affordability of their water supply (i.e., household water security) through increased access to basic water services and/or improved quality of water services from the PDAM.

Hotspot Neighborhood Hypothesis

In addition to the changes in household water security experienced throughout the PDAM service area, if targeted IUWASH Tangguh TA promotes specific improvements in PDAM performance in hotspot neighborhoods, promotes investment in increased household connections in hotspot neighborhoods, and promotes household willingness and desire to connect to the piped network and pay for water provided by PDAMs, then households in these hotspot neighborhoods will experience even greater improvements in the quantity, quality, reliability, and affordability of their water supply than are experienced in the service area writ large through increased access to basic water services and/or improved quality of water services from the PDAM.

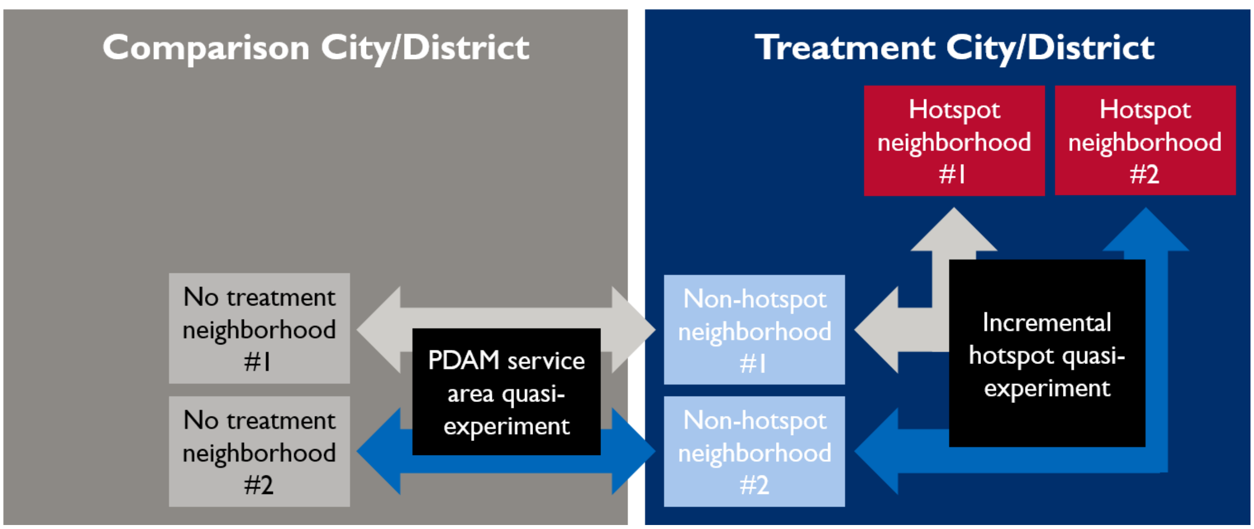
3.2.2 DESCRIPTION OF TREATMENT AND COMPARISON GROUPS

URBAN WASH's EQ1 design includes two quasi-experiments. In the first quasi-experiment, labeled the "incremental hotspot experiment," the treatment group comprises IUWASH Tangguh's hotspot neighborhoods, which receive the most intense (and complete) version of IUWASH Tangguh's combined interventions to improve household water security. This treatment group includes all households who reside in hotspot neighborhoods at the time of baseline data collection. IUWASH Tangguh envisions selecting two hotspot neighborhoods per treatment site by the evaluation baseline, though they will continuously expand the number of hotspots over activity implementation; at minimum to four per site by the end of the second project year. So, URBAN WASH proposes a hotspot treatment group which includes the 62 hotspot neighborhoods as of baseline data collection from the 31 treatment cities and districts included in the study.

For the incremental hotspot experiment, the comparison group includes urban neighborhoods in cities and districts treated by IUWASH Tangguh which resemble hotspot neighborhoods but were not selected for the hotspot treatment as of the evaluation baseline. These neighborhoods benefit from IUWASH Tangguh's institutional interventions at the national, provincial, local, and PDAM levels, but do not receive the hotspot interventions. URBAN WASH will select at least two neighborhoods in each treatment city and district using statistical matching to construct the comparison group. [22]

Comparing outcomes for the treatment and comparison groups represents the average treatment effect of IUWASH Tangguh's hotspot intervention on hotspot neighborhoods relative to the effect of the non-hotspot component of the Activity.

Figure 6: EQ1 Experimental Design Construction



The second experiment, labeled the "PDAM service area quasi-experiment," compares non-hotspot neighborhoods in treatment cities/districts to matched neighborhoods in comparison cities/districts. The treatment group comprises the comparison group selected for the incremental hotspot quasi-experiment. URBAN WASH will construct the comparison group by statistically matching treatment group neighborhoods to two neighborhoods in comparison cities and districts on the same set of characteristics used to match the hotspot and non-hotspot neighborhoods in treatment cities. Comparing outcomes for these two groups represents the average treatment effect of IUWASH

[22] At the time of this design report, the specific 62 hotspots are not yet known. It is possible some of these hotspots may focus on WRM and fall in the catchment area outside the city, rather than within the city itself. Such hotspots would be declared ineligible for the study, and replaced with non-treatment hotspots (i.e., if 5 of 62 hotspot neighborhoods were ineligible for the study, the baseline sample would include 57 hotspot/red neighborhoods and 67 non-hotspot/light blue treatment neighborhoods).

Tangguh's non-hotspot interventions on households in neighborhoods which resemble hotspot neighborhoods.[23] This impact, together with the incremental hotspot impact, represents the full impact of IUWASH Tangguh on hotspot neighborhoods.

3.2.3 USING STATISTICAL MATCHING TO CONSTRUCT COMPARISON GROUPS

IUWASH Tangguh clarified that most of its hotspot neighborhoods will be within PDAM service area boundaries at the time of baseline data collection. However, some neighborhoods will be currently outside the service area and are where local PDAMs envision expanding their services in the near term. URBAN WASH will seek to duplicate this situation in selecting comparison neighborhoods – neighborhoods which are currently inside the PDAM service area and those to which PDAMs envision expanding services in the near term would both be eligible for inclusion in the study. If it is possible to identify in time for statistical matching (i.e., no later than January 30), location in or outside the service area will be an exact matching criteria for comparison group selection (i.e., hotspots outside service areas would only be matched with non-hotspot neighborhoods outside service areas).

IUWASH Tangguh is currently selecting its hotspot neighborhoods using a combination of Village Potential Survey (PODES) data, Master File Village data, satellite imagery, and geospatial data from PDAMs related to piped network coverage. Selection will prioritize urban neighborhoods with a high proportion of low-income households and low access to WASH services in or near PDAM service areas.[24] Neighborhoods which have the potential to increase access to safely managed drinking water supply will be prioritized. URBAN WASH proposes to select indicators from as many of these same data sources as it can access to use as matching variables in a genetic matching algorithm to select comparison neighborhoods. At minimum, URBAN WASH will use variables from the PODES dataset related to poverty and access to WASH services. URBAN WASH will match neighborhoods without replacement (i.e., a comparison neighborhood may match to only one treatment neighborhood). The matches will be selected from the paired city or district from the city/district-level matching exercise. So, for example, comparison neighborhoods for the treatment neighborhoods in Kabupaten Wonogiri will be selected from Kabupaten Pati.

In advance of the statistical matching exercise, URBAN WASH will seek to obtain a list of neighborhoods from each PDAM which are currently included in its service area or where the PDAM envisions expanding in the next two years. If URBAN WASH cannot obtain these lists before the statistical matching exercise, its data collection teams will verify whether the best matches are eligible for the study (i.e., within or soon to be within the service area), upon arriving in each city and district. If the best match is not eligible, the data collection team will proceed to the next best eligible match.

3.2.4 HOUSEHOLD SAMPLING DESIGN

The IE will administer household surveys at baseline and at endline. The intent is to collect panel (longitudinal) data at the household level (i.e., the same households are interviewed in each round of data collection). This requires a baseline sample size large enough to account for attrition (assumed to

---

[23] As discussed later on, this does not represent IUWASH Tangguh's average treatment effect in the PDAM service area more broadly, as the treatment group is not representative of neighborhoods in the broader service area. It is only representative of neighborhoods similar to hotspot neighborhoods. If URBAN WASH's assumptions underlying causal pathways hold, the treatment effect on neighborhoods like hotspot neighborhoods will be larger than the effect in the broader service area. Thus, this likely represents an upper bound on IUWASH Tangguh's service area-wide impact, which is likely smaller.

[24] All neighborhoods in cities are urban neighborhoods, but some neighborhoods in districts are rural.

be no greater than 12.5 percent) that will produce an *endline* sample size sufficient to achieve targeted precision.

URBAN WASH will draw the household sample from two hotspot neighborhoods in each of the 31 treatment cities and districts, two non-hotspot neighborhoods in each of the treatment cities and districts, and two neighborhoods similar to hotspot neighborhoods in the 31 comparison cities and districts.

The household sample frame is thus the population of households that reside within the 186 neighborhoods selected for inclusion in the study. Households from these neighborhoods will be selected randomly, or quasi-randomly using the random walk method or similar, to ensure representativeness within the sampling frame. [25] Before interviewing a household, enumerators will confirm that the household possesses a *Kartu Keluarga* (KK, a government-issued identification card) to ensure they are an official resident. This verification should help reduce endline attrition. Any household that does not meet these conditions will be replaced with another randomly selected household.

URBAN WASH calculated a required household sample size for this study using the assumptions for the two household-level quasi-experiments in Table 5. URBAN WASH will sample 1,674 households for baseline—558 hotspot households and 558 non-hotspot households in treatment cities/districts and 558 no-treatment households in comparison cities/districts (Figure 6). The assumed 12.5 percent level of attrition will leave a sample of 1,488 households for endline. The parameters underlying URBAN WASH's power calculations are approximations within a standard range typically used for ex ante power calculations. However, URBAN WASH will update its power analysis with baseline data, which may increase or decrease the minimum detectable effect size (MDES) depending on the true values for parameters such as R^2 and blocking covariance capture in the evaluation's sample and for the evaluation's outcome variables. This power calculation makes an assumption that the size of the hotspot, treatment non-hotspot, and comparison groups will be the same (62 neighborhoods each). However, as previously described, the size of these groups may change at baseline (as certain hotspots are declared ineligible for the study) and endline (as certain non-hotspots end up receiving the hotspot intervention). At each phase of the evaluation, URBAN WASH will describe how the construction and size of these groups has changed and what effects these changes have on the study's statistical power.

Table 5: Power Calculations for Household Sample

INCREMENTAL HOTSPOT TREATMENT		PDAM-ONLY TREATMENT	
Design parameter	*V*	*Design parameter*	*V*
Number of sites (all treated)	31	Treated sites	31
Hotspot neighborhoods / treated site	2	Comparison sites	31
Non-hotspot neighborhoods / treated site	2	Neighborhoods/site	2
Households/neighborhood	8	Households/neighborhood	8
MDES	0.205	MDES	0.220
Endline Sample size (#Cs = #Ts)	992	Endline Sample size (#Cs = #Ts):	992*
R^2 (2nd stage)	0.5	R^2 (3rd stage)	0.5
Level of significance (α)*	0.05	Level of significance (α)	0.05

25 See UNICEF Multiple-Indicator Survey Handbook, Chapter 6, for a description of such methods. https://mics.unicef.org/files?job=W1siZiIsIjIwMTUvMDQvMDMvMDYvNDIvNDgvMzgyL2NoYXAwNi5wZGYiXV0&sha=6509e495a61af931

INCREMENTAL HOTSPOT TREATMENT	
Blocking covariance capture	0.3
Power	0.80
**Assumes one-tailed significance test*	

PDAM-ONLY TREATMENT	
Between neighborhood cluster variability	Fixed effects
Power	0.80
**Half already accounted for in first quasi-experiment*	

3.2.5 OUTCOME INDICATORS

The IE defines household water security as "reliable access to a quantity and quality of water adequate to maintain wellbeing." This comprises:

- A **source** which is improved and accessible;
- from which water is **reliably** available;[26]
- in the **quantity** needed to meet basic needs;
- with sufficient **quality** to pose no risk to the health of household members; and
- whose cost is **affordable** in the context of household income and other expenditures required to meet basic needs.

To measure household water security, the IE assigns an indicator to each of the five components and estimates changes in each indicator caused by IUWASH Tangguh from baseline to endline. These indicators are listed in Table 6.

Table 6: Outcome Indicators for Household Water Security

COMPONENT	INDICATOR[27]	UNITS
Access	Household's main drinking water source is "improved" as defined by JMP service ladder, on premises, and available when needed	Categorical, aligned with JMP service ladder
Reliability of supply	Number of days in last seven days when regular availability of main drinking water source has been disrupted[28]	Number of days in the last seven days
Quantity	Total water consumed by household members (for any purpose across all sources)	Liters per capita per day of water consumed from all sources combined[29]

26 We use the term "reliable" here, though in a way that we believe is consistent with the GOI's goals for "continuity." This choice reflects that, though an ideally reliable source would be available continuously, more reliable availability reflects stronger water security than less reliable availability, even if the source is not continuously available.

27 URBAN WASH proposes to use estimates of water supply over the most recent seven or thirty days for its measures of impact, where applicable, because these are where respondents will have the most accurate recall and/or most recent water bill from their PDAM. However, URBAN WASH also proposes to collect estimates of typical water supply by season and assess the sensitivity of evaluation results to these measures.

28 The questionnaire will be clear this is referring to abnormal changes to water availability or pressure for more than one hour. Though this theoretically would include disruptions to water quality, these are not always perceptible to household members and can be falsely identified. So, the questionnaire will be clear this does not include changes to the taste, odor, or appearance of the water.

29 Note this will be approximated differently for different sources of water. For PDAM customers, the most recent monthly bill divided by days per month would yield the estimate of liters per day. For sources outside the household, consumption will be approximated based on the container normally used to collect water and the frequency with which water is normally collected in the current season, normalized to the day.

COMPONENT	INDICATOR[27]	UNITS
Quality	Presence/absence of *E. coli* in drinking water at the household point of consumption[30]	Binary – household either has absence of *E. coli* at point of consumption or presence of *E. coli* at point of consumption
Affordability	Household expenditure on water (from all sources for any purpose) over the last month divided by approximate total household monthly expenditure[31]	Percent of total expenditure

In addition to measuring change on each of these indicators individually, URBAN WASH will construct an index which scores a household's overall water security by assigning values between 0-100 corresponding to each of the components of water security and averaging the scores across the components (i.e., assigning each component an equal weight) to compute an overall score. These scoring criteria affect only the way the data are analyzed, not how they are collected. URBAN WASH can thus test the sensitivity of IE estimates on household water security to different specifications, including assigning different weights to each component or adjusting the scoring criteria. URBAN WASH proposes to score the components of household water security as indicated in Table 7 on the following page. These criteria are based on a combination of international and GOI standards for water supply in each component.

Table 7: Scoring of Household Water Security Index

COMPONENT	SCORING CRITERIA	STANDARDS
Access (A)	**100:** Improved source, on premises, available when needed **60:** Improved source, within 30 minutes roundtrip of household including queuing time, available when needed **50:** Improved source, within 30 minutes roundtrip of household including queuing time, not available when needed **40:** Improved source, more than 30 minutes roundtrip from household including queuing time, available when needed **30:** Improved source, more than 30 minutes roundtrip from household including queuing time, not available when needed **20:** Access only to unimproved sources, available when needed **10:** Access only to unimproved sources, not available when needed **0**: Access only to surface water	100% access to "safely managed WASH" per JMP Service Ladder (GOI and WHO)

30 Note that the absence of *E. coli* for this component combined with the considerations for access in the first component constitute all the requirements for a household to have "safely managed access" to drinking water according to WHO definitions.

31 URBAN WASH intends to ask households to select which category their expenditure belongs in from a set of ten deciles of Indonesian expenditure in urban areas. For the purposes of this indicator, URBAN WASH will divide the household WASH expenditure by the lower bound of the category of expenditure the household selects. This means the actual percentage calculated is an upper bound—actual WASH expenditure may be slightly lower as a percentage of expenditure than is calculated, but will be no higher than is calculated.

COMPONENT	SCORING CRITERIA	STANDARDS
Reliability (R)	**100:** No disruptions last seven days **80:** Service disrupted in one of last seven days **55:** Service disrupted in two of last seven days **25:** Service disrupted in three of last seven days **15:** Service disrupted in four of last seven days **10:** Service disrupted in five of last seven days **5:** Service disrupted in six of last seven days **0:** No service in last seven days	24 hours per day of water service (GOI)
Quantity (Q)	**100:** 100 liters per capita per day or more **90:** 60-99 liters per capita per day **50:** 50-59 liters per capita per day **30:** 20-49 liters per capita per day **0:** fewer than 20 liters per capita per day	At least 60 liters per capita per day (GOI) 20 required for basic needs, 50 for intermediate needs, 100 for all needs (WHO)
Quality (E)	**100:** Absence of *E. coli* at point of consumption **33:** Presence of *E. coli* at point of consumption, absent at point of collection **0:** Presence of *E. coli* at point of consumption and point of collection	Absence of priority fecal contamination (GOI and WHO)
Affordability ($)	**100:** Water expenditure no more than 4.0% of total income **80:** 4.0 - 4.5% **60:** 4.5 - 5.0% **40:** 5.0 - 5.5% **20:** 5.5 – 6.0% **0:** More than 6.0%	Water tariffs must not exceed 4% of customer income (GOI) 2 – 6% of total income (JMP)
Total	**HWS Index Score = (A+R+Q+E+$)/5**	

Annex 3 illustrates how two hypothetical households would score on this index based on different assumptions regarding their water supply, and how their scores would change over time based on changes to their supply.

3.2.6 EQ1 ANALYSIS PLAN

To generate causal estimates of impact from the matched treatment and comparison groups, URBAN WASH will test and select the best performing of two possible analysis methods: difference-in-difference (DID) or analysis in covariance (ANCOVA). Annex 3 describes these methodologies in more detail. Each of these methods can account for baseline differences in outcomes of interest between treatment and comparison groups, which are likely given the differences in some important pre-intervention characteristics in the proposed treatment and comparison cities. These methods will perform best with panel data, which URBAN WASH is incorporating in its plans for data collection and sample size calculations. Although URBAN WASH proposes to use genetic matching to select comparison neighborhoods for each EQ1 quasi-experiment, URBAN WASH may update the matching exercise with the baseline and endline data. Specifically, an updated matching exercise would test whether alternative matching methods improve the balance between the treatment and comparison groups—including methods which might match multiple comparison neighborhoods to one treatment neighborhood.

As part of the analysis to estimate causal impacts, URBAN WASH will conduct additional quantitative and qualitative analyses to understand explanatory factors for observed impacts or lack thereof. This includes assessing the influence of covariates in final fixed effects models to calculate program impact, thematic analysis of qualitative interview data regarding possible explanations for program impact, and interpretations from subject matter experts in URBAN WASH's research team. URBAN WASH will also use these analyses to generate lessons learned and recommendations for USAID.

3.2.7 ASSUMPTIONS AND LIMITATIONS

Assumptions: Testable and Non-Testable

A DID analysis assumes "parallel trends" between the treatment and comparison group. In other words, estimates of impact can be accurate even if there are baseline differences in outcomes of interest, but only if the treatment group's trend in outcomes would have been identical to the comparison group's trend absent the intervention. The parallel trends assumption is not testable. However, similarity between treatment and comparison groups on observable characteristics that influence selection and outcomes gives increased confidence that trends in outcomes would have been similar without the intervention. As such, combining statistical matching with DID reduces the risk that the parallel trends assumption poses. If feasible with the available data, URBAN WASH will also compare pre-intervention trends for key covariates between the treatment and comparison groups as a check on this assumption. [32]

ANCOVA analysis can increase statistical power relative to DID if the appropriate assumptions hold, particularly if autocorrelation is low.[33] URBAN WASH will test autocorrelation as part of its decision process to determine which method to select for final IE analysis. If autocorrelation is not low enough, DID will be a more appropriate method.

Limitations and Tradeoffs

There are four main limitations to URBAN WASH's proposed EQ1 design, described in this section in descending order of the risk the limitation poses to the evaluation. First, IUWASH Tangguh's plan to expand hotspot treatment over the course of activity implementation—combined with the reality that some hotspots may not focus on water supply interventions—risks introducing selection bias into the study's experimental groups and reducing the planned quasi-experiments' statistical power. Second, as with all quasi-experimental designs employing statistical matching, the design is vulnerable to omitted or unobserved variable bias. Third, the design will estimate an average treatment effect for a highly complex and heterogeneous set of interventions. Finally, the design excludes a key treatment group from its analysis—the set of households in neighborhoods which do not resemble hotspot neighborhoods in treatment PDAM service areas. These limitations, their implications, and associated mitigation strategies are described below.

Contamination and reduced precision due to hotspot treatment assignment: At the time of this report, IUWASH Tangguh does not yet know which 62 neighborhoods in treatment cities and districts will begin as hotspot neighborhoods, what will be the focus of their hotspot intervention (e.g., water supply, sanitation, or WRM), or how many additional neighborhoods will become hotspot neighborhoods by the impact evaluation's endline (2026). IUWASH Tangguh anticipates most hotspot neighborhoods will fall within or near the treatment PDAM service area, and that most will have at least a partial water supply focus in their intervention. However, an unknown minority of hotspots will be in the treatment city's catchment, totally outside the PDAM service area and without any plans for water supply interventions over the course of the activity. URBAN WASH must exclude these neighborhoods from the study. As the number of treatment hotspots decreases according to this justification, the

---

32 These are the same variables listed in Tables 3, 14, and 15.

33 While either method produces the same impact estimates, one will be more precise.

statistical power (and/or precision of estimates of impact) for URBAN WASH's proposed incremental hotspot experiment also decreases. Meanwhile, as IUWASH Tangguh expands its hotspot intervention to additional neighborhoods, it may target some neighborhoods which the study has selected as treatment non-hotspot neighborhoods for water supply hotspot interventions (i.e., "contaminate" neighborhoods that the study assumes are untreated with the hotspot treatment). Such expansion, particularly if it is very extensive, renders the hotspot and treatment non-hotspot groups less and less comparable (i.e., introduces selection bias under the assumption that the most similar neighborhoods are selected first and least similar neighborhoods remain) and also reduces the statistical power/precision of both planned quasi-experiments (which rely on comparisons to a consequentially smaller non-treatment hotspot group).

URBAN WASH will take multiple measures to mitigate against the risks that these complications pose to the study. First, URBAN WASH proposes to increase the number of neighborhoods in the non-treatment hotspot group in a one-to-one ratio for each hotspot neighborhood that is excluded from the study at baseline. Second, URBAN WASH proposes to reclassify the quasi-experimental group of some neighborhoods at endline based on the treatment received. For example, hotspot neighborhoods which receive only sanitation interventions would be reclassified as non-hotspot neighborhoods (since they only received water supply interventions at the city/institutional level), and non-hotspot neighborhoods which receive the water supply-focused hotspot intervention would be reclassified as hotspot neighborhoods.[34] These measures will build a buffer in the size of the non-treatment hotspot group to accommodate the expansion of the hotspot intervention, and also allow the size of the hotspot group to recover by endline using neighborhoods named as hotspots in the interim. If the extent of hotspot intervention expansion is limited, any selection bias should be minimal (i.e., many reasonably similar non-hotspot treatment neighborhoods should remain) and the statistical power/precision of the two quasi-experiments should be more or less preserved (i.e., close to 62 neighborhoods will end up in each group). However, if the hotspot intervention expands so significantly that there are very few non-hotspot treatment neighborhoods remaining, the two proposed quasi-experiments will suffer from selection bias and low statistical power/precision. If this occurs, URBAN WASH proposes to fall back to a single quasi-experiment, comparing the hotspot group (which would be very large) with the comparison/no-treatment group. This maneuver forfeits the ability to estimate separate treatment effects for the hotspot and non-hotspot components of the intervention, but retains a valid impact estimate for the whole of treatment effect on hotspot neighborhoods.[35] In either case, URBAN WASH will plan to use econometric techniques in analysis to assess the effect of the duration of exposure to treatment on impacts, given that some hotspot neighborhoods will have been exposed to treatment for three years and others will have been exposed for a shorter duration.

Omitted variable bias: Any characteristics which influence both selection for the Activity and Activity outcomes, but which are not measured by the study, would cause differential trends in outcomes between the treatment and comparison groups and bias study results. This phenomenon is referred to as "omitted" or "unobserved" variable bias. Given that IUWASH Tangguh's site selection process included a subjective step along with objective criteria, there is a non-negligible probability that at least

---

[34] There is one exception to this strategy: in Kota Blitar, the institutional intervention is planned to be sanitation only. If this plan remains in place, and hotspots in this city only receive sanitation interventions, they would be reclassified as comparison/no-treatment neighborhoods, since there would be no water supply intervention at the institutional or neighborhood level.

[35] Note that the ideal strategy to avoid these risks entirely would be to pursue a pipeline design, where any expansion of the IUWASH Tangguh intervention to non-hotspot neighborhoods in the study is delayed until after the study's completion. However, IUWASH Tangguh has made clear that a pipeline design is not feasible in this case.

some factors which explain both treatment and outcomes are not captured by available data sources. The only way to avoid this bias would have been to randomly assign sites to treatment and control groups, which was not possible for this evaluation. To substantially bias impact estimates, unobserved variables would need to be strongly correlated with Activity selection and impacts and poorly correlated with observable covariates selected by the study. While some unobserved variables might exist, the likelihood of these being poorly correlated with observable covariates, and thus of this bias substantially affecting the validity of evaluation results, is low. Nevertheless, to mitigate against this bias, URBAN WASH will attempt to render as many characteristics associated with Activity selection and outcomes "observable" as possible, either directly or by proxy, and control for their influence in estimating treatment effects.

Average treatment effect: URBAN WASH's EQ1 design will estimate an average treatment effect across all hotspot households in all geographies covered by the study. While the proposed design will yield reliable estimates of average treatment effects across all treated sites, it will sometimes be difficult to assess which element of treatment is most responsible for observed effects, or if all elements are necessary. This limitation stems from the substantial complexity and heterogeneity of interventions that comprise the IUWASH Tangguh activity. As it stands, the proposed design chooses to isolate the local and neighborhood treatment effects from the national and provincial ones, which will apply equally across treatment and comparison groups and thus not appear in estimates of program impact. Design-based estimates of the contribution of different elements of the intervention to the impacts observed would be informative but would require much larger samples. More importantly, it would violate the assertion of IUWASH Tangguh that its theory of change requires tailored packages of interventions to work together, which would not produce the same effects if they were implemented separately. Regardless, to facilitate learning and considerations for scaling the intervention in the future, URBAN WASH will employ outcome equations and qualitative data analysis to attempt to characterize how different aspects of the intervention contribute to results observed. Incorporating IUWASH Tangguh's AMELP data on outputs and outcomes, ideally disaggregated by treatment site (i.e., city/district, PDAM, or neighborhood), into endline analysis will be particularly helpful in this regard.

Exclusion of parts of the treatment group: Finally, since the treatment neighborhoods selected outside hotspot neighborhoods and the neighborhoods selected from comparison areas for the EQ1 design are both selected purposively to resemble hotspot neighborhoods, the evaluation will not yield estimates of impact on the population of households in the entire treatment PDAM service areas, which are the majority of Activity participants. As such, it will be important to interpret the impact estimates yielded by the study appropriately—they are the impact of IUWASH Tangguh on a specific type of neighborhood, and not the impact of the activity on its full population of beneficiaries. [36] This limitation emerges from the need to prioritize evaluation resources toward specific learning objectives. Evaluation resources cannot accommodate separate, representative, and precise estimates of impact for the full set of households inside and outside hotspot neighborhoods in treatment sites. This presents a need to prioritize between an evaluation of the whole-of-service area treatment effect, which would understate hotspot effects, and an evaluation of the hotspot effect, which would overstate whole-of-service area effects. Selecting the latter specification in consultation with co-design stakeholders ensures that the evaluation captures the effect of the most complete version of the IUWASH Tangguh interventions on the set of program participants' whose household water security stands to improve the most. Any expansion of the IUWASH Tangguh approach by USAID or GOI would likely target a similar population

[36] The proposed design also omits impacts experienced by households in the seven treatment cities and districts in DKI Jakarta, Papua, East Nusa Tenggara, and West Kalimantan. However, if there are no potential matched cities and districts in the same provinces, there is no way to generate reliable estimates of impact for these households in any case.

(i.e., socioeconomically vulnerable with low access to safely managed WASH), which further establishes this group as a priority focus for the evaluation. Nonetheless, URBAN WASH will seek opportunities in endline qualitative data collection for informed stakeholders to comment on perceived differences in activity effectiveness for the full set of households who live outside hotspot neighborhoods.

3.3 EQ3 (CITY-WIDE WATER SERVICE RESILIENCE) DESIGN

3.3.1 HYPOTHESES OF INTEREST

As discussed in Section 2.2, the IUWASH Tangguh theory of change posits that the Activity will improve the climate-resilience of two types of city-wide water services: most directly those provided by the PDAM and, less directly, private and community water services overseen by the city or district government. URBAN WASH's proposed IE tests two specific hypotheses.

PDAM City-Wide Water Service Resilience Hypothesis

If IUWASH Tangguh (i) supports locally relevant climate vulnerability assessments for use by PDAM decision makers; (ii) provides TA to build PDAMs' capacity for integrating data on risks to their water supply into management and decision-making; (iii) provides additional TA to improve water safety planning and identify external investment for avoiding or mitigating risks; (iv) supports direct interventions to protect the quality and quantity of bulk water sources; and (v) supports broad improvements in PDAM operational, financial, and administrative performance; *then* PDAMs will better identify, understand, and avoid or mitigate risks to their water services through improved planning, improved financial resources, and improved baseline water service provision *and then* disruptions to the quantity and quality of PDAM water services will be less frequent and shorter.

Non-PDAM City-Wide Water Service Resilience Hypothesis

If IUWASH Tangguh supports locally relevant climate vulnerability assessments for use by local decision makers in city and district government *Pokjas* for water service provision and provides TA to build their capacity for integrating data on risks to their water supply into water safety plans and budgets, *then* city and district governments will better identify, understand, and avoid or mitigate risks to private and community water services through improved planning and financial resources *and then* disruptions to the quantity and quality of private and community water services in city and district boundaries will be less frequent and shorter.

3.3.2 DESCRIPTION OF TREATMENT AND COMPARISON GROUPS

The treatment group includes the PDAM and the city or district government in each of the 31 cities and districts included in the study.

The comparison group will consist of the cities and districts in the same provinces as the treatment cities and districts selected by statistical matching as described in "Selection of treatment and comparison cities and districts" section.

3.3.3 CITY SAMPLING DESIGN

Each PDAM and city or district government in the treatment and comparison groups for the household-level design will be included in the sample for the city-level design. Intended respondents from PDAMs will be Research and Development Managers or their designee. URBAN WASH proposes surveying the *Bappeda* representative to the *Pokja* (or their designee) for the LG survey. Assuming a one-stage trial design with a level of significance equal to 0.05, an R^2 value of 0.50, and power equal to 0.80, URBAN WASH's design is powered for a MDES of 0.520.

3.3.4 OUTCOME INDICATORS

Table 8 includes the set of indicators which URBAN WASH proposes to use to quantify city-wide water service resilience. For the purposes of an overall city-wide resilience index, each of these indicators are pass/fail, meaning that the PDAM or LG achieves the required condition or does not. Nine of these indicators apply to PDAMs, while six apply for LGs. Thus, a city's PDAM resilience index score would be a value from 0 to 9 corresponding to the number of PDAM-specific resilience conditions which it satisfies, and its LG resilience index score would be a value from 0 to 6 corresponding to the number of LG-specific resilience conditions which it satisfies.

The proposed measures combine expert review of water safety planning documentation, survey results, and secondary PDAM performance data aggregated by PUPR. URBAN WASH proposes to characterize impact by analyzing the proportion of treatment vs. comparison cities for which the conditions described in the table are true or false. However, for many of the indicators, URBAN WASH will collect more detailed information that permit an analysis of the magnitude of change. For example, URBAN WASH can analyze the change in raw financial performance scores for treatment and comparison PDAMs in addition to analyzing the change in the proportion which qualify as health or not healthy.

Table 8: EQ3 City-Wide Water Supply Resilience Measures

COMPONENT	INDICATOR	SOURCE	PDAM/LG
Risk identification	Institution's planning document (either Water Safety Plan [RPAM], Business Continuity [BC] Plan, or Water Supply System Master Plan [RISPAM]) identifies hazards to water supply based on localized climate projections	RPAM/ RISPAM expert scoring	Both
Risk understanding	Institution's planning document (RPAM, BC Plan, or RISPAM) incorporates scenarios no more than five years old for most likely and severe hazards with instructions for use, and identified intervals for updates no longer than five years	RPAM/ RISPAM expert scoring	Both
Risk data use	[PDAM/LG] monitors real-time data from each of the following, as relevant: - bulk water source quantity and quality - early warning systems for hydrometeorological disasters - early warning systems for geological disasters	Survey*	Both
Planning for risk mitigation and avoidance	[RPAM/RISPAM] includes objectives and measures to prevent and/or mitigate risks to water service provision, including target indicators and timeframes for risk avoidance/mitigation	RPAM/ RISPAM expert scoring	Both
Finance for risk mitigation and avoidance	[PDAM/LG] budget includes separate allocations for risk avoidance/mitigation and disaster response/recovery that cannot be used for other purposes	Survey*	Both
Water quality monitoring	LG asserts that water quality at point of use is adequately monitored for PDAM, community, and private water supplies per GOI regulations	Survey*	LG

COMPONENT	INDICATOR	SOURCE	PDAM/LG
Operational performance (quantity)	Production volume is less than installed capacity, while maintaining at least 16 hours per day of water service	PUPR Performance Data	PDAM
Financial performance	PUPR Financial Performance score is sufficient to qualify as "healthy"	PUPR Performance Data	PDAM
Adequate staffing	Respondent perceives PDAM staffing as adequate to maintain operational performance	Survey	PDAM
Infrastructure safety	Respondent states that PDAM water abstraction, treatment, and distribution infrastructure is designed to withstand disruptions from most likely hazards	Survey	PDAM

* For indicated survey measures, URBAN WASH will assess during piloting if it is possible for PDAMs and/or LGs to produce proof of this assertion which an enumerator could observe and document.

3.3.5 EQ3 ANALYSIS PLAN

As with the EQ1 design, URBAN WASH proposes to use the higher performing of DID or ANCOVA analysis to detect impacts on city-wide resilience for PDAMs and LGs. URBAN WASH proposes to conduct separate analyses to estimate changes in the resilience of PDAMs and the resilience of LGs as city-wide water service providers. However, the analysis has not been designed to attribute the independent impact of IUWASH Tangguh's PDAM-targeted interventions on PDAM resilience, or the independent impact of its LG-targeted intervention on LG resilience. It could be, for example, that interventions targeting the PDAM improved LG resilience, or vice versa. Therefore, each of these analyses will estimate the whole-of-program impact on PDAMs and LGs.

URBAN WASH intends to include a variety of other measures in the PDAM and LG surveys to help explain or contextualize changes in the outcome indicators. These will include data on the types of water sources cities and districts rely on, the specific hazards they are most concerned with, the perceived likelihood and severity of these hazards, and the perceived relationships with other institutions responsible for disaster reduction and response, among other measures. Section 4 of this report provides describes the content of these surveys in detail.

At endline only, URBAN WASH additionally intends to interview institutional actors responsible for WRM and WASH service provision in a small sub-set of treatment cities to understand perceived changes in WRM and disaster reduction and response procedures which may favor improved resilience and which may help further explain or contextualize quantitative results.

3.3.6 ASSUMPTIONS AND LIMITATIONS

All assumptions described in the previous section related to statistical matching, DID analysis, and ANCOVA analysis also apply to the city-level design, where URBAN WASH proposes to use the same methods.

The city-level design may produce imprecise estimates of impact due to the small number of cities and districts in the treatment group. URBAN WASH will collect corroborating data from qualitative and/or secondary sources to strengthen the credibility of EQ3 IE findings in case quantitative results are inconclusive (i.e., if any effect is lower than the study's MDES).

Quantitative measures of city-wide resilience are less established, and perhaps less reliable, than measures of household water security. The proposed resilience indicators are adapted loosely from the

UN Sendai Framework for disaster risk reduction and modified during the evaluation co-design process. URBAN WASH hopes this evaluation contributes to sector-wide learning of how to assess resilience in the context of an IE.

3.4 TENTATIVE QUALITATIVE METHODS AND EQ2 (BULK WATER AVAILABILITY) PERFORMANCE EVALUATION DESIGN

At the endline, URBAN WASH proposes to employ exclusively PE methods to respond to EQ2 (i.e., there will be no IE estimates of Activity impact for this question) and collect qualitative data from institutional respondents on themes relevant to all three EQs. URBAN WASH intends to define these methods following the baseline, and after at least two years of implementation monitoring, to ensure that they reflect the realities of Activity implementation and emerging learning priorities identified by evaluation stakeholders between baseline and endline. This section briefly outlines the general shape URBAN WASH intends these evaluation activities to take. The endline update to the EDR will present the final plan for PE-specific and qualitative data collection and analysis.

3.4.1 TENTATIVE EQ2 DESIGN

URBAN WASH intends to pursue a mixed-methods PE approach for EQ2. A qualitative assessment of institutional WRM processes, including any change respondents perceive that this represents from historical WRM processes, forms the foundation of this approach. URBAN WASH intends to interview PDAM, LG, and (where relevant) provincial government personnel responsible for WRM in at least four treatment cities and districts. The evaluation team will also interview national-level personnel with roles in WRM (e.g., from the Ministry of Environment and Forestry, river basin management organizations, Ministry of Energy and Mines, etc.) to gather information on WRM processes. Where possible, URBAN WASH will add questions to its endline surveys of PDAM and LGs to get more broadly representative descriptions of WRM processes from the treatment group to complement the qualitative data. Thematic analysis of the qualitative data along with descriptive analysis of the quantitative data will respond to the first portion of EQ2: *How have PDAMs participating in IUWASH Tangguh and their local government counterparts changed WRM policies and practices in response to the Activity?*

To address the second part of EQ2, (i.e., *What implications does this [changed WRM policies and practices] have for the quantity and quality (i.e., availability) of their [PDAMs] bulk water supply*?), the evaluation team intends to conduct longitudinal analysis of data from treatment PDAMs' management information systems (MIS) regarding the availability (quantity and quality) of their bulk water sources. IUWASH Tangguh intends to support PDAMs in developing these MIS, so URBAN WASH understands this data will only be available for PDAMs in treatment areas. The specific indicators available from these MIS will only be known after they have been developed and implemented. With appropriate data, URBAN WASH expects to be able to establish a trend in bulk water supply for the treatment group to complement qualitative information provided by institutional stakeholders and subject matter experts. The evaluation team will ask these stakeholders to respond to the trends observed and describe any implications they perceive for bulk water availability in the future.

However, measurable changes in bulk water availability from a raw water source require large-scale interventions and/or long time horizons, and IUWASH Tangguh does not expect its interventions to affect bulk water availability within the evaluation timeframe. Therefore, while URBAN WASH will report on trends in bulk water availability, attributing trends to changed WRM processes will likely be a prospective (and even potentially speculative) exercise.

3.4.2 ANTICIPATED QUALITATIVE METHODS FOR EQ1 AND EQ3 (ENDLINE ONLY)

URBAN WASH plans to conduct KIIs after completing the preliminary endline quantitative analysis and estimating program impact to seek institutional respondent perspectives on factors that may explain impacts, or lack thereof, detected on household water security and city-wide resilient water services. Interview respondents may include PDAM senior leadership, PDAM operational staff, and *Pokja* personnel.[37] If any cities or districts in the treatment group experience a significant shock that affects their water services prior to endline data collection, it may be interesting to select this city or district for a qualitative case study and investigate how it leveraged IUWASH Tangguh's TA to prepare for and respond to this shock.

URBAN WASH may also seek to conduct a limited number of focus group discussions with household respondents to understand explanatory factors for observed changes in household water security from the household perspective. The focus groups could also seek to understand gender and social dimensions of changes in PDAM water services for the households who experience them.

URBAN WASH anticipates it will field a pair of two-person research teams to collect qualitative data over 16 days each of field work. This would accommodate about 50 KIIs covering national stakeholders in Jakarta and at least four treatment cities and districts.

3.5 COST ANALYSIS

Based on preferences expressed by participants in evaluation co-design, URBAN WASH proposes to use a cost-effectiveness analysis for this evaluation. Cost-effectiveness analysis answers the question "What did this activity cost per *outcome* delivered?"[38] While cost-effectiveness analysis can also compare the ratio of cost to outcomes across different activities, URBAN WASH understands that this is not the objective of the cost-effectiveness analysis for this Activity. Instead, URBAN WASH's cost-effectiveness analysis will focus on establishing what it might cost to replicate this Activity and its outcomes—for example if the GOI desired to replicate IUWASH Tangguh's TA or if USAID intended to expand the interventions to additional geographies. The cost analysis could contribute to institutionalizing IUWASH Tangguh interventions, so the resources exist locally to replicate and scale up results.

URBAN WASH proposes that the evaluation's cost effectiveness analysis not only include the cost incurred by USAID, but also the costs incurred by GOI and funds mobilized by other donors and the private sector in support of IUWASH Tangguh which would not have occurred in IUWASH Tangguh's absence.[39] URBAN WASH understands that USAID/Indonesia is tracking GOI contributions, including monetizing in-kind contributions, and can share data on this source of external funding. IUWASH Tangguh will track mobilized funds, but it will also include additional GOI allocations that would not have occurred without IUWASH Tangguh. To populate relevant measures in its PDAM Index, IUWASH Tangguh will be tracking PDAM budgets to report on its indicator of percentage increase in local department allocations for water.

Because the sanitation aspect of IUWASH Tangguh's intervention is not included in the scope of this evaluation, URBAN WASH will work with IUWASH Tangguh to understand how this cost may be excluded from the calculation of overall Activity cost. Where it is not feasible to exclude (e.g., for time

---

37 Some of these respondents will also be targeted to respond to questions relevant to EQ2. In these cases, corresponding interview protocols will include questions and probes relevant to all three EQs within the same interview

38 https://www.edu-links.org/sites/default/files/media/file/USAID%20Cost%20Analysis%20Guidance_Final%20Feb20.pdf

39 This is an important distinction—for example, a local government may have contributed a certain value of grants or equity to its local PDAM even without the intervention. So, the full value of their contribution should not be counted toward the analysis, only the portion which was facilitated by the intervention.

spent by core management personnel), URBAN WASH will request that IUWASH Tangguh make a rough approximation for how much cost should be attributed to the sanitation interventions and excluded.

URBAN WASH will work with IUWASH Tangguh to track all its costs and all funds contributed by external sources in support of the Activity based on Activity administrative and M&E data. In its analysis of cost-effectiveness, URBAN WASH can only present the ratio of detected impacts to the whole-of-Activity costs, given that all of URBAN WASH's evaluation designs detect whole-of-Activity impact. However, where it is possible to do so with minimal additional effort, URBAN WASH will also request that IUWASH Tangguh disaggregate its costs where feasible in the following categories:

1. Cost by Objective
2. USAID vs. GOI vs. private sector cost
3. Hotspot vs. non-hotspot cost
4. Cost by service-package (i.e., full interventions, water supply, WRM).

It will not be possible to estimate cost effectiveness for specific disaggregation within these categories, since the evaluation will not yield estimates of impact at this level (i.e., even if the cost of Objective 1 is known on its own, the impacts of Objective 1 on its own will not be known). However, estimates of the costs of specific Activity components and evidence of their potential contribution to impacts may inform lessons learned for taking the program to scale.

3.6 GENDER COMPONENTS OF ANALYSIS

Throughout the evaluation, URBAN WASH will seek opportunities to contextualize evaluation results from a gender perspective. Since this evaluation is focusing on outcomes that are at the household- and city-level, and not at the individual level, it will not be possible to analyze Activity impact data in a gender-disaggregated fashion. However, given the important gender roles that often exist in managing a household's water supply, and further given IUWASH Tangguh's explicit intention to increase the agency of women in WASH institutional settings, URBAN WASH will conduct analysis to understand how women and men might experience Activity impacts differently. Specific measures URBAN WASH will pursue to this end include:

- Emphasizing that the individual responsible for and most knowledgeable about the household water supply respond to the household survey. URBAN WASH expects this person will be a woman in most cases;
- Analyzing the gender of household members who collect water from sources outside the home and integrating this information in the analysis of any source switching that occurs over the course of the program;
- Including the volume of gender-related outputs in Objective 4 as a covariate in IE analysis to assess the relationship between intensive gender-focused interventions of the IUWASH Tangguh Activity and Activity impact;
- Including questions in the PDAM and LG surveys inquiring as to perceptions related to the participation of women in WASH institutions, for example as PDAM management or operational staff; and

- Ensuring women are among the respondents selected for qualitative data collection, where feasible.[40] Qualitative instruments will include specific questions regarding the gender dimensions of impact at the institutional and household levels.

[40] Note URBAN WASH does not set a specific goal for the proportion of qualitative respondents who are women because it is possible relatively few women are employed with WASH institutions in the roles which URBAN WASH would target for interviews (e.g., executive directors of PDAMs, etc.). In such cases, it is not possible to guarantee a certain proportion of respondents are women.

4.0 DATA COLLECTION, QUALITY ASSURANCE, AND MANAGEMENT

4.1 BASELINE DATA COLLECTION INSTRUMENTS

URBAN WASH will deploy three survey data collection instruments for the baseline, one for the household survey, one for the PDAM survey, and one for the LG survey. The evaluation team will revise each of the instruments, if necessary, for the endline. This section summarizes the basic structure and content of each instrument. URBAN WASH will develop the instruments and submit them for review by USAID and IUWASH Tangguh following USAID review and approval of this EDR. In its endline update to the EDR, URBAN WASH will describe any changes to these instruments and will add a description of any qualitative data collection tools deployed at endline only.

4.1.1 HOUSEHOLD SURVEY AND WATER QUALITY TESTS

URBAN WASH will administer a 60-minute, face-to-face questionnaire with households that will provide most of the data for the evaluation's main EQ1 outcome indicators. The questionnaire will also collect data on important household-level covariates for the evaluation. Data collectors will test water quality at the point of collection and the point of consumption in a sample of surveyed households.[41] URBAN WASH will use tablet-based, computer-assisted personal interviewing (CAPI) techniques for this survey. Table 9 summarizes the basic household survey structure and modules.

Table 9: Summary of Household Survey Questionnaire

MODULE NAME	MODULE CONTENT
A. Consent	Informed consent script, identification of intended respondent
B. Household roster	Head of household identification, plus gender, age, education, primary occupation, and relationship to head of household for each household member.
C. Water sources	Household indicates each of the water sources it uses, including in which season(s) the source is regularly used and for what purpose(s). For sources of drinking water, households will indicate which of the sources it uses is its main source.
D. Water collection, consumption, and expenditure	**This module repeats once for each of the household's sources of water.** For each source, the household will indicate round-trip collection time (including queuing), amount collected, and associated expenditure. For sources of drinking water, household will indicate availability and reliability. Questions to be asked for most recent seven or thirty days, and again for each season (i.e., rainy and dry). There will be minor differences in questions asked according to source (e.g., PDAM customers will be asked to present bills, while households will estimate costs based on recall for other sources).
E. Water storage and treatment	Household will describe its water storage and treatment practices. If household consents, the enumerator will observe any vessels the household uses to store water, recording whether these are open, sealed, etc. Households will further describe any water treatment practices they regularly use.
F. Experiences of water insecurity	This module will include the questions from the standard Household Water Insecurity Experiences (HWISE) questionnaire, presented verbatim, with some additional questions in

[41] For households with a piped connection to PDAM water services and multiple taps, the point of collection will be the tap closest to the customer meter.

MODULE NAME	MODULE CONTENT
	a similar style associated with household expenditure on water.[42] These questions are meant to understand the psychosocial consequences of water insecurity on the household (e.g., on the way household members feel, spend time, etc.).
G. Household information and economic status	Enumerator observes and inquires as to various aspects of the household construction, access to sanitation, and hygiene behaviors. Household also shares information regarding its household ownership status, approximate total expenditure, and approximate assets.
H. Water quality testing	**For only a sub-set of households**, enumerator reads a consent script specific to water quality testing. If granted consent, enumerator asks household for a glass of water that it would normally drink (i.e., "point of consumption") and takes a sample for *E. coli* testing. If the source of this water is on the premises, the enumerator additionally takes a sample from the source (i.e., "point of collection") for *E. coli* testing. Each of these samples is incubated for at least 24 hours with results interpreted and recorded thereafter.
I. Conclusion and case disposition	Respondent is thanked for their time and requested to share contact details in case follow-up data validation is required. Enumerator summarizes the outcome of the survey (e.g., complete interview, partial interview, refusal, non-contact, etc.).

4.1.2 PDAM AND LOCAL GOVERNMENT OFFICIAL SURVEYS

Both the PDAM and LG official surveys will be implemented face-to-face over the course of approximately 30 minutes using tablet-based CAPI techniques. These interviews will provide some of the evaluation's main EQ3 outcome indicators. The interviews will also collect information on city- and PDAM-level characteristics that will be important covariates for EQ1 and EQ3. The PDAM and LG questionnaires will have a similar structure and content. Table 10 presents their contents together in one table for the sake of efficiency. There will be some minor differences in the instruments as indicated below.

Table 10: Summary of PDAM and Local Government Official Survey Questionnaires

MODULE NAME	MODULE CONTENT	SURVEY
A. Consent	Informed consent script, identification of intended respondent and their role within their institution.	Both
B. Raw water source profile	Respondent will indicate the type (e.g., river, spring, aquifer, etc.) and name of each raw water source their institution relies on for water service provision. Respondent will also approximate the volume of water abstracted from each source over the previous 12 months, if possible.	Both
C. Risk identification	Respondent will indicate which hazard(s) most threaten to disrupt each of their raw water sources, characterizing the most probable and most severe consequences to their water supply if these hazards occur.	Both
D. Risk data use	Respondent will indicate which data sources, if any, their institution uses, and at which frequency, to monitor hazards to their water services. These may include bulk water source quantity or quality monitoring, early warning systems for natural disasters, etc. During instrument piloting, URBAN WASH will assess if it is possible for enumerators to observe documentary evidence of any sources monitored to corroborate survey responses.	Both

42 Young SL, Boateng GO, Jamaluddine Z, et al. The Household Water In Security Experiences (HWISE) Scale: development and validation of a household water insecurity measure for low income and middle-income countries. BMJ Global Health 2019;4:e001750. doi:10.1136/ bmjgh-2019-001750

MODULE NAME	MODULE CONTENT	SURVEY
E. Planning for risk mitigation and avoidance	Respondent will indicate which of the hazards they named, if any, are targeted by formal strategic planning for risk mitigation/avoidance. Respondent will also share perceptions regarding the adequacy of their institution's plans. Respondent will share with which institutions they partner, and how often they communicate, to manage their water sources. Respondent will also indicate their satisfaction with each of these partnerships.	Both
F. Finance for risk mitigation and avoidance	Respondent will indicate whether their institutions' budget includes separate allocations for risk avoidance/mitigation and disaster response/recovery that cannot be used for other purposes. They will also share details regarding public and/or private investment they have received for these purposes, including the source, amount, and purpose of these investments. During instrument piloting, URBAN WASH will assess if it is possible for enumerators to observe documentary evidence of institutions' budgets or financial policies to corroborate survey responses.	Both
G-LG. LG-specific performance questions	Respondent will indicate institution's compliance with requirements for monitoring water quality at the point of consumption, separated by service provider (PDAM, private, community-based). During instrument piloting, URBAN WASH will assess if it is possible for enumerators to observe documentary evidence of compliance with requirements to corroborate survey responses.	LG only
G-PDAM. PDAM-specific performance questions	Respondent will share basic details about the profile and tenure of PDAM executive leadership. They will also share perceptions regarding adequacy of staffing for maintaining operational performance and inclusion of considerations for withstanding hazards in the design of PDAM water abstraction, treatment, and distribution infrastructure.	PDAM only
I. Government commitment	Respondent will indicate their perception regarding commitment of their city or district government to PDAM performance, including in terms of financial contributions.	PDAM only
J. Gender in WASH institutions	Respondents will share perceptions related to the participation and agency of women in WASH institutions. Will also ask about perceptions of inclusive service	Both
K. Conclusion and case disposition	Respondent is thanked for their time and requested to share contact details in case follow-up data validation is required. Enumerator summarizes the outcome of the survey (e.g., complete interview, partial interview, refusal, non-contact, etc.).	Both

4.1.3 RPAM/RISPAM SCORING MATRIX

In addition to its three survey instruments for baseline, URBAN WASH will develop standard templates with defined criteria where expert reviewers will score the compliance of RPAMs and RISPAMs with standards for resilience established in Section 3.3. Reviewers will also document the justification for their score.

4.2 DATA QUALITY ASSURANCE

URBAN WASH will integrate measures for data quality assurance (DQA) in the design of its survey instruments, training of data collection personnel, and oversight of data collection activities (including data transmission, storage, and cleaning) as described in the ensuing sections.

4.2.1 SURVEY TRANSLATION, PROGRAMMING, AND PRE-TEST

URBAN WASH will prepare paper versions of each of its survey instruments following USAID review and approval of this EDR. Once approved by USAID, URBAN WASH will rely on data collection subcontractor Article 33 to translate the instruments into Bahasa Indonesia. Article 33 will hire two independent translators to separately translate the instruments, with differences reconciled between the two translators to produce a final translation. Article 33 will pre-test the paper versions of the translated instruments with up to eight respondents in low-income neighborhoods of DKI Jakarta. The purpose of this paper instrument pre-test is to ensure that the instruments consistently elicit the expected information from respondents. This paper instrument pre-test will also give Article 33's survey leadership team an opportunity to familiarize themselves with planned survey instruments, sampling procedures, and logistic prior to enumerator training.

Concurrently with translation, URBAN WASH will program the instrument for a CAPI interface using the KoboToolbox suite of software. URBAN WASH will integrate measures in the survey programming that eliminate, where feasible, the possibility of user error in data entry. This includes, for example, constraining the range of numeric indicators to a set of logical values (e.g., 0 to 24 hours in a day), constraining downstream response options based on upstream responses (e.g., only asking about water collection for the sources indicated in the source module), and imposing skip logic automatically throughout the survey so that the enumerator does not read erroneous questions to the respondent. Where relevant, responses to questions will be deemed mandatory so that they cannot be skipped over.

Once programmed, URBAN WASH will pre-test the programmed instrument. Pre-testing programmed instruments involves mocking surveys on an identical device to one that will be used in data collection, ensuring the survey programming operates as intended, and attempting to challenge the survey logic to ensure that illogical values cannot be entered inadvertently. This component of the pre-test exercise ensures the programming logic and quality checks work as intended, there are no fundamental errors in the programming (e.g., no sequence of responses brings the tool to a premature end), and that the survey appears on the screen in a way that is intuitive for the enumerator to complete. URBAN WASH will iteratively update the programming as pre-testing uncovers bugs.

4.2.2 ENUMERATOR TRAINING AND PILOTING

With final programmed data collection instruments in-hand, URBAN WASH will develop training materials for field data collection personnel. Training manuals will be developed for reference throughout training and over the course of data collection. These manuals will include:

- Background information on the study and the IUWASH Tangguh Activity;
- Roles and responsibilities for the various data collection personnel (enumerators, supervisors, data managers, central office personnel, etc.);
- A glossary of key terms for the study, including photos or graphics where needed (for example to define different kinds of water sources, water storage vessels, etc.);
- Expectations for responsible, ethical, and professional behavior from field data collection staff; and
- Description of data quality assurance and data management procedures, including methods for data validation, expectations for data storage and transfer, etc.

URBAN WASH subcontractors Article 33 and NORC will hold one enumerator training per round of data collection. Trainings will last six days in total, with four days in the classroom followed by a pilot data collection exercise and associated debriefing session implemented over the ensuing two days. The classroom sessions will focus on the study background, best practices for face-to-face data collection, content of the questionnaires, and role playing using the same tablets which will be used for field data collection. One of the classroom sessions will allow enumerators to practice the water quality testing procedures. URBAN WASH anticipates training about 8-10 supervisors and 35-45 enumerators. For the pilot exercise, each enumerator will conduct at least two household surveys. URBAN WASH anticipates supervisors will conduct the LG and PDAM survey and will ensure each supervisor either conducts or observes a PDAM or LG survey during the pilot exercise. The enumerators and supervisors will follow sampling protocols exactly as designed for the evaluation in an area similar to the geography selected for fieldwork but not included in the evaluation sample. Supervisors and field managers will observe enumerator performance and dismiss any enumerators whose knowledge and/or execution of the data collection tools is deemed insufficient. Following the pilot exercise, URBAN WASH will debrief with enumerators to ensure there are no remaining areas of concern prior to full-scale implementation of the survey, and to catch any bugs in programming that may have evaded the pre-testing exercise.

4.2.3 DATA QUALITY CONTROL AND ASSURANCE

URBAN WASH will monitor enumerator quality in the field by checking the quality of a minimum of 20 percent of interviews. Quality checks may include having supervisors observe interviews, visiting respondents to verify the interview, and conducting telephone back checks to verify key variables. Additionally, the team will design customized data quality monitoring scripts in Stata to automatically run several quality control checks after data are uploaded from tablets. These checks will flag inconsistent responses, exceptionally high non-response within surveys or by enumerator, excessively long or short interview durations, checks on GPS locations to ensure interviews were conducted within expected geographic boundaries, and custom checks on survey items identified as being potentially problematic or exceptionally important. The results of these checks will be fed back to field data collection personnel to take necessary corrective action and ensure all quality issues are resolved.

Table 11 presents specific DQA issues, checks, and corresponding response strategies URBAN WASH will undertake to address them during survey implementation.

Table 11: DQA Approach

DQA TYPE /DESCRIPTION	RESPONSE STRATEGY
Date/time verification. Ensures start and end times of the survey are logical (i.e., sequential and within the field period) and that survey duration is not abnormally short or long.	Flag anomalous interview dates/times/durations for field supervisors who verify whether the survey actually took place. Investigate and re-enumerate cases of suspected data falsification.
"Don't know"/"No Response" frequencies. Flags variables for which the don't know/no response rate is 5% or more, and cases where a given enumerator has at least 5 don't know/refused responses.	If particular enumerators are registering abnormally high don't know/refused responses, the supervisor will talk to them to understand the source of the issue and provide coaching on administering the questions correctly.
Outlier review. Flags continuous numerical variables of over 2 standard deviations from the mean value.	Flag questionable outliers to the supervisor to verify whether they are correct or need to be cleaned.
GPS verification. For a given sampling unit, evaluates GPS coordinates for all interviews to assess if they were conducted within the sampling unit.	Alert supervisors to interviews at anomalous locations. Investigate and re-enumerate cases of suspected data falsification.

DQA TYPE /DESCRIPTION	RESPONSE STRATEGY
Back checks. Back check data are analyzed using *bcstats*, a user-written Stata command and .do file that outputs a .csv file with all discrepancies between original and back check data, by variable and enumerator.	When discrepancies are observed, enumerators may be temporarily suspended until an investigation is complete. If data falsification is confirmed through the investigation, the enumerators are immediately terminated and their surveys are re-enumerated.

4.3 HUMAN SUBJECTS PROTECTION AND DATA MANAGEMENT PLAN

4.3.1 INFORMED CONSENT AND HUMAN SUBJECTS PROTECTION

URBAN WASH subcontractor NORC at the University of Chicago (NORC) has an in-house Institutional Review Board (IRB) which is registered with the United States Federal Office for Human Research Protections (Federal Wide Assurance #00000142). This registration demonstrates that the NORC IRB complies with the United States government's strict regulations for research on human subjects. NORC's IRB requires that research protocols provide sufficient detail to ensure that (1) the selection of subjects is equitable, subjects' privacy is protected, and data confidentiality is maintained; (2) informed consent is written in language that study participants can understand and is obtained without coercion or undue influence; and (3) appropriate safeguards protect the rights and welfare of vulnerable subjects. URBAN WASH will submit a package to NORC's IRB for this study which will include a full description of the study sample, the data collection methodology, the use of incentives, the informed consent statements, all contact materials, and the survey instruments. This package will also include the curriculum used for training interviewers and other project staff on human subject protections, as well as the security plan to safeguard data security and the information technology infrastructure for collecting and transmitting data. URBAN WASH will await approval from NORC's IRB prior to commencing enumerator training and will follow the protocols approved by IRB to ensure that the rights of the respondents in the IUWASH Tangguh IE will be fully protected.

URBAN WASH subcontractor Article 33 will concurrently work to obtain ethical approval from Indonesia's National Research and Innovation Agency (BRIN), adapting a version of the US IRB package to BRIN's specifications. Article 33 will also send letters and follow up in person to the National Unity, Politics and Community Protection Agency (*Badan Kesbangpol*) in the LG office of each city and district where survey data collection will take place to obtain administrative permits for data collection.

4.3.2 DATA STORAGE AND TRANSFER

All data will be collected on tablets using the KoboCollect software. All tablets will be password protected, and URBAN WASH will include protocols for physical tablet security in its training materials for enumerators. Data will be uploaded each day from tablets using Secure Sockets Layer (SSL) encryption protocol to a server managed by Kobo, Inc. and hosted by Amazon Web Services. This password-protected server will be accessible only by a minimal set of credentialed users on URBAN WASH's research team. A backup version of the data will be securely downloaded from this server and stored on URBAN WASH subcontractor NORC's internal server, where it will be accessed only by members of the URBAN WASH research team for analysis purposes. Once evaluation activities are complete, URBAN WASH will further submit primary datasets collected in service of the evaluation to USAID's Development Data Library (DDL) in accordance with ADS 579.3.2.6.

5.0 EVALUATION MANAGEMENT PLAN

5.1 EVALUATION TIMELINE AND USAID PARTICIPATION

Table 12 presents the timeline for key baseline evaluation milestones. Aside from standard review of deliverables during the allocated dates, URBAN WASH requests the following support from USAID:

1) **Remote final co-design:** In addition to providing written feedback on this document, URBAN WASH requests that USAID/Indonesia and USAID/RFS join the evaluation team and the IUWASH Tangguh team to discuss any points where the stakeholders may have divergent points of view and/or where clarification of key elements of the evaluation design may be required. The final co-design discussion is intended to reach final decisions on the evaluation design and permit a rapid turnaround of a final EDR. It may also be useful to discuss key points related to the evaluation instruments during this meeting, the drafts of which will be near complete. If necessary, URBAN WASH will schedule a separate session devoted to discussing the instruments once drafts of these have been completed.

2) **Securing administrative approvals for data collection:** URBAN WASH's data collection subcontractor will likely require letters of support on USAID letterhead to request permissions to collect data. Once their specific needs are known, URBAN WASH will share draft letters for USAID/Indonesia to revise, print on letterhead, and sign. These requests will likely come to USAID in early January 2023.

3) **Outreach to selected cities, districts, and PDAMs for data collection:** Aside from administrative approvals, URBAN WASH may also request that BAPPENAS, PERPAMSI, PUPR, or other Indonesian institutions help secure agreement from PDAMs and LGs to participate in the study. The study may also require certain inputs from these organizations, such as spatial and administrative data required to sample neighborhoods and households. URBAN WASH would appreciate USAID/Indonesia facilitating these requests, should any be required. Any such requests will come to USAID in early January 2023.

4) **Support for procuring and shipping materials:** Depending on the specific materials procured for water quality testing, USAID support may be helpful to ensure that the material can be shipped internationally expeditiously, avoiding any undue costs or delays in the customs process. URBAN WASH may also request USAID's support storing these materials prior to data collection, depending on the storage capabilities of URBAN WASH's data collection subcontractor. The procurement and shipment of water quality testing materials will likely occur between December 2022 and January 2023.

5) **Support with progress updates and dissemination for GOI:** URBAN WASH requests that USAID/Indonesia coordinate as necessary with GOI stakeholders to ensure their participation in progress updates and dissemination activities. URBAN WASH proposes to offer these stakeholders a virtual presentation regarding the final study design prior to data collection training (late January 2023) and a virtual presentation of baseline evaluation findings in June 2023.

6) **Implementation Monitoring:** Concurrently with baseline data collection and continuing for three years thereafter, the evaluation team will monitor program implementation and ensure that any updates to program implementation which imply challenges or opportunities for the established evaluation design are documented and communicated with all evaluation stakeholders. URBAN

WASH requests USAID/Indonesia's assistance to ensure that Evaluation Director Miguel Albornoz and Deputy Team Lead Trimo Pamudji receive regular updates on implementation, at minimum by having the opportunity to review regular IUWASH Tangguh progress reporting deliverables. Annex 4 includes a proposed Memorandum of Understanding between URBAN WASH and IUWASH Tangguh that details the specific information the evaluation requires from IUWASH Tangguh over the course of Activity implementation.

Table 12: Timeline for Key Baseline Evaluation Milestones

TASK	ANTICIPATED DATE(S)*
EDR submitted	Nov. 30, 2022
EDR reviewed	Dec. 1 - Dec. 14, 2022
Remote final co-design meetings(s)	Dec. 16 – 21
Draft baseline data collection instruments submitted	Within 2 business days of co-design meeting
NORC (International) IRB Submission	Within 2 business days of draft instrument submission
Outreach to PUPR/BAPPENAS to secure comparison site participation	Dec. 30, 2022
Revised/final EDR	Jan. 6, 2023
Outreach to comparison sites to request participation	Jan. 6, 2023
Final EDR approved	Jan. 13, 2023
NORC IRB approval of initial submission	Jan. 13, 2023
Submit amendment to NORC IRB approval with final design and instruments	Jan. 16, 2023
NORC IRB approval of amendment	Jan. 23, 2023
Comparison sites agree to participate, or replaced	Jan. 30, 2023
Enumerator training materials finalized, English and Bahasa Indonesia	Feb. 6, 2023
Baseline data collection training and piloting	Feb. 13 - 20, 2023
Baseline data collection and ongoing data quality assurance	Feb. 20 – Mar. 22, 2023
Draft baseline report submission	May 26, 2023
Baseline draft findings debrief with USAID	Week of May 29, 2023
Draft baseline report reviewed	May 29-June 2, 2023
Final baseline report submission	June 30, 2023
Baseline final findings debrief with GOI (if desired)	Week of July 3, 2023

* Anticipated dates will depend on USAID review and approval of the EDR and the survey instruments as well as the local ethical and administrative approvals obtained by the data collection firm.

By October of 2025, URBAN WASH will submit revisions to endline instruments, if any, together with a brief update to the EDR covering elements of the evaluation design specific only to endline data collection. These might include qualitative methods, sample designs, and instruments for EQ1 and EQ3 together with PE methods for EQ2. The update to the evaluation design report will also include a timeline of key milestones for the endline evaluation, which URBAN WASH anticipates will unfold between January and September of 2026. The extended endline evaluation timeline allows for qualitative data collection once preliminary quantitative data analysis has been completed, which is not envisioned for the baseline.

5.2 EVALUATION PERSONNEL

URBAN WASH will implement this evaluation together with two subcontractors. NORC at the University of Chicago will design the evaluation, lend technical oversight to survey data collection, and lead data analysis, reporting, and dissemination. The local data collection firm, Article 33, will collect and ensure the quality of survey data from households, PDAMs, and LGs. The URBAN WASH Deputy Chief of Party (DCOP) Miriam Otoo will coordinate across the various parties engaged in the evaluation, with the support of URBAN WASH evaluation consultant Doug Krieger, URBAN WASH Project Manager Zach Borrenpohl and other URBAN WASH project management and operational staff. Mr. Borrenpohl is the Tetra Tech Buy-in Manager and will provide overall management with support from the DCOP. NORC Evaluation Director Miguel Albornoz will report to Dr. Otoo and coordinate NORC's role in evaluation design and implementation. Article 33's Executive Director Santoso will report to Dr. Otoo and coordinate Article 33's role in survey data collection, receiving technical guidance and oversight from Mr. Albornoz. The relationships between these organizations and associated personnel are depicted in Figure 7, with additional details on roles and responsibilities included immediately below in Table 13.

Figure 7: Organogram for Impact Evaluation

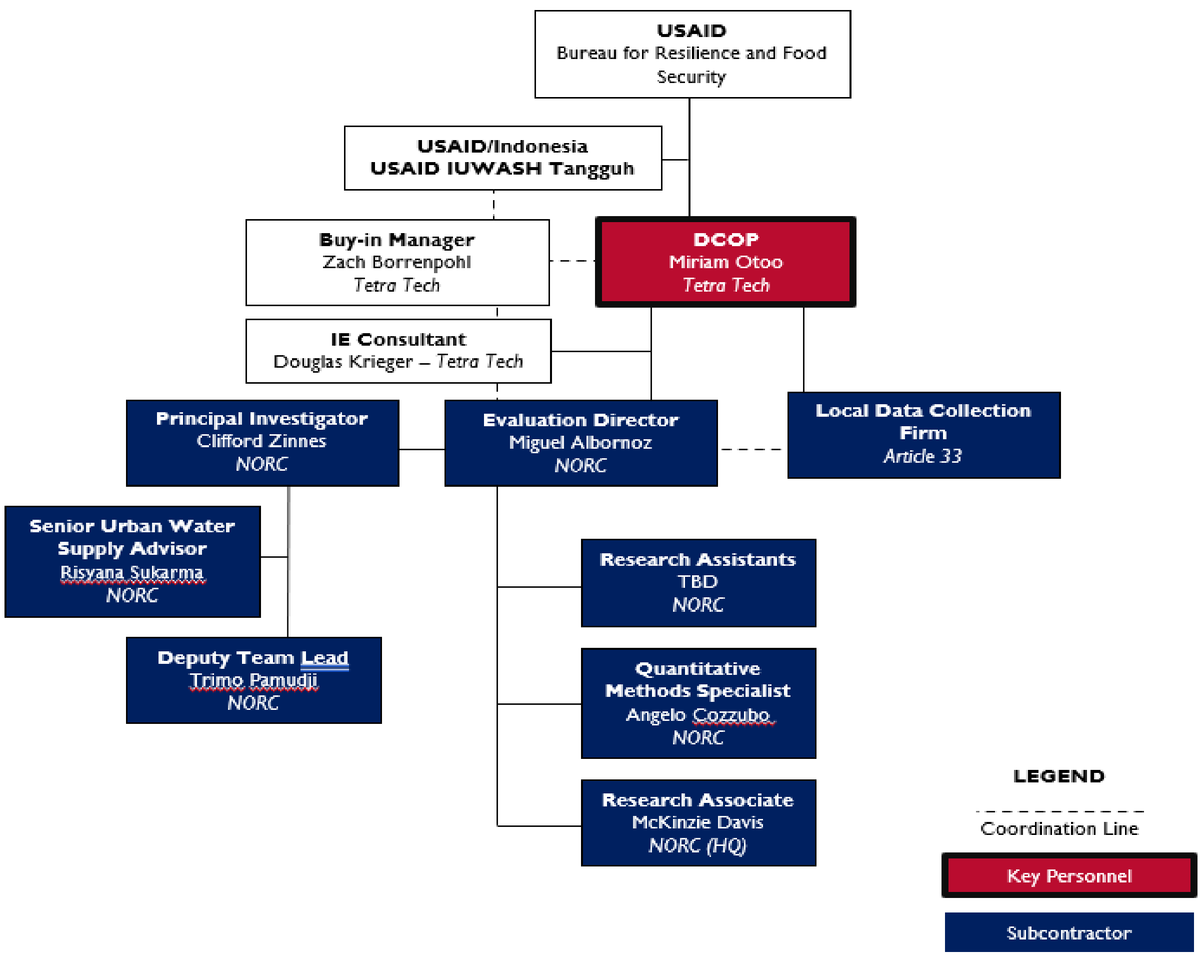


Table 13: Evaluation Team Roles and Responsibilities

TEAM MEMBER	ROLE	RESPONSIBILITIES
Miriam Otoo	Deputy Chief of Party	Responsible for overall delivery of the evaluation design, data analysis plan, and reporting and dissemination. Supervise and support the work of the entire evaluation team, with direct oversight of the Evaluation Director and the data collection firm. Provide inputs and support to the final IE products. Support the review and quality control process for deliverables.
Zach Borrenpohl	Buy-in Activity Manager	Manages and coordinates the evaluation design, data analysis plan, and reporting and dissemination. Provide inputs and support to the final IE products submitted by the DCOP. Support the review and quality control process for deliverables.
Douglas Krieger	Impact Evaluation Consultant	Provide inputs and support to the final IE products. Support the review and quality control process for deliverables.
Clifford Zinnes	Evaluation Team Lead/Principal Investigator	Responsible for providing methodological leadership on all technical aspects of the evaluation design, data analysis plan, and reporting and dissemination. Supervise and support the work of the entire evaluation team, with direct oversight of the deputy team lead. Provide inputs and data, as requested, to the URBAN WASH DCOP, and support the synthesis and interpretation of information. Support the review and quality control process for NORC deliverables.
Trimo Pamudji	Evaluation Deputy Team Lead	Responsible for providing methodological input and leadership on all technical aspects of the evaluation design, data analysis plan, and reporting and dissemination. Ensure that evaluation findings, conclusions, and recommendations are properly contextualized given local (Indonesian) policies, regulations, and practical considerations. Lead implementation monitoring activities in-country and conduct primary qualitative data collection with WASH institutional personnel.
Risyana Sukarma	Senior Urban Water Supply Advisor	Supports in co-design for determining indicators and data collection methods for key outputs, outcomes, and impacts associated with expertise. Further supports interpretation of results during data analysis, reporting, and dissemination, as needed and relevant to subject matter expertise.
Miguel Albornoz	Evaluation Director	Main point of contact between external stakeholders and NORC. Coordinates all aspects of the evaluation team and effort (technical, management, partnership, etc.). Contributes heavily to evaluation design, analysis, reporting, and dissemination under leadership of principal investigator.
Santoso	Data Collection Team Lead	Executive for data collection activities, reporting to the URBAN WASH DCOP and coordinating with the evaluation director. Coordinates all personnel engaged in field data collection and leads the survey management team, which consists of a survey research specialist, two regional field managers, and a data manager/programmer.
Angelo Cozzubo	Quantitative Methods Specialist	Supports team lead, or delegated responsibility as appropriate, for counterfactual identification strategy, power calculation, sampling, and quantitative measurement, data quality assurance, analysis, and visualization.
McKinzie Davis	Research Associate	Supports all research tasks as necessary and delegated by other team members, supports in design and analysis of qualitative instruments and data as delegated by evaluation director.

TEAM MEMBER	ROLE	RESPONSIBILITIES
TBD	Local Research Assistant(s)	Supports field data collection and data quality control. May contribute to collection of primary qualitative data, as needed.

5.3 COMMUNICATIONS AND COORDINATION PLAN

Mr. Borrenpohl is responsible for coordinating between the DCOP, USAID stakeholders in Indonesia, and Washington D.C., the IUWASH Tangguh project team, and evaluation team members working for URBAN WASH and its subcontractors.

When communicating regarding this evaluation with USAID stakeholders, and consistent with ADS 203 definitions and requirements, URBAN WASH and its subcontractors will refer to this evaluation as an "impact evaluation" or "program evaluation" of IUWASH Tangguh. However, to facilitate smooth collaboration with stakeholders in Indonesia, all stakeholders will refer to this evaluation as an "independent study" of IUWASH Tangguh to stakeholders outside USAID and the evaluation team. The evaluation team will be referred to as a "study team" in these contexts.

Occasionally, evaluation team members employed by NORC or Article 33 may need to interact with government or civil society stakeholders in Indonesia to request data, provide updates regarding the evaluation, or disseminate evaluation results. In all cases, this interaction will be facilitated by and under the discretion of the USAID/Indonesia Mission. The USAID/Indonesia Mission may request assistance from the IUWASH Tangguh team to facilitate such interactions. URBAN WASH and its subcontractors will not contact these stakeholders directly without prior approval of USAID/Indonesia.

URBAN WASH and its subcontractors will rely on USAID/Indonesia Monitoring, Evaluation, and Learning (MEL) Specialist Ade Darmawansyah and IUWASH Tangguh Contracting Officer's Representative (COR) Trigeany Linggoatmodjo in their capacity as evaluation Activity managers for day-to-day technical direction. URBAN WASH will submit all evaluation deliverables and any critical decisions regarding scope to URBAN WASH COR Ryan Mahoney for approval, who will consult with Mr. Darmawansyah and Ms. Linggoatmodjo as needed. In its ongoing process of evaluation co-design, URBAN WASH will also solicit and consider feedback from the IUWASH Tangguh project team, the USAID/Asia Bureau, and stakeholders in Indonesia in finalizing its evaluation design.

In all communications, URBAN WASH, URBAN WASH's subcontractors, USAID, and IUWASH Tangguh are to be clear that the evaluation (or "study") team is independent of the IUWASH Tangguh project team. Although the evaluation team and the IP will collaborate, this collaboration is only for the purposes of facilitating the evaluation and does not compromise the independence of the study. To ensure smooth collaboration for an effective evaluation while maintaining independence, URBAN WASH has drafted a memorandum of understanding (MOU) to be signed by IUWASH Tangguh, URBAN WASH, and its subcontractors codifying their mutual responsibilities for the success of the impact evaluation. The draft MOU is presented in Annex 4.

5.4 EVALUATION DISSEMINATION AND USE

Baseline evaluation reporting and analysis will focus on characterizing the baseline status of outcomes of interest for the treatment group and assessing the similarity between the treatment and comparison groups on outcome variables and key covariates (i.e., do any dissimilarities between the groups pose challenges to the study's validity?). In turn, endline evaluation reporting and analysis will accomplish the evaluation's objectives related to measuring and explaining program impact, analyzing the unit costs of impacts achieved, and generating recommendations and lessons learned for USAID's consideration. Evaluation dissemination activities will vary between baseline and endline according to the differences in objectives for reporting and analysis.

5.4.1 BASELINE DISSEMINATION AND USE

At baseline, URBAN WASH proposes to hold two remote dissemination sessions. The first, a Draft Findings Debrief for USAID, would target USAID/Indonesia, USAID/RFS, USAID/Asia, and IUWASH Tangguh personnel. This session would follow the submission of the draft baseline report and make the evaluation team available to summarize baseline findings and conclusions, especially as relevant to the baseline's two analytical objectives (i.e., assessing balance between quasi-experimental groups and characterizing the status of pre-intervention outcomes). The discussion from this session will help ensure efficient and effective review of the draft baseline evaluation report and ensure that key questions are resolved in the final baseline evaluation report. URBAN WASH does not envision holding a recommendations workshop at baseline, as the baseline analysis will yield few (if any) recommendations for USAID.

Once the baseline evaluation report is finalized, URBAN WASH proposes offering a courtesy Findings Debrief to interested GOI stakeholders, including BAPPENAS, PUPR, and PERPAMSI at minimum. This session would also include a USAID audience and incorporate updates to findings and conclusions made in revisions to the draft baseline report. URBAN WASH proposes to frame this session more as a progress update than a learning event. This is because both baseline analytical objectives are only tangentially useful to these stakeholders—the sampling design for the evaluation is meant to be representative for IUWASH Tangguh's treatment group only, and so URBAN WASH must be careful in this session to avoid implying that findings associated with this group can be extrapolated to apply to any broader group in which these stakeholders might be interested. Still, to the extent the baseline status of any of the evaluation's survey measures are of interest to this group, URBAN WASH proposes to offer this opportunity to discuss them. The main objective of this session will be to keep these stakeholders engaged and aware of the plans for more substantive endline findings.

5.4.2 ENDLINE EVALUATION USE

Following the evaluation endline, URBAN WASH proposes a more robust set of dissemination sessions intended to drive learning from evaluation results.

As with baseline, URBAN WASH proposes a Findings Debrief with USAID and IUWASH Tangguh stakeholders immediately following the submission of the endline evaluation report. The debrief will focus on conclusions regarding Activity impacts relevant to each EQ, potential explanatory factors for these impacts, and lessons learned for future programs. URBAN WASH will request that stakeholders complete their detailed review of the endline evaluation report following this debrief, and then reconvene for a recommendations workshop following the review. During the recommendations workshop, URBAN WASH will facilitate a discussion of the evaluation's draft recommendations focused on USAID and IUWASH Tangguh perceptions regarding the recommendations' relevance and practicality. The workshop will aim to provide all inputs needed for URBAN WASH to refine and finalize the recommendations for the final endline evaluation report.

Once the endline evaluation report is completed, URBAN WASH proposes two learning events to disseminate final evaluation findings, conclusions, and recommendations. First, URBAN WASH proposes a session devoted to a wide set of USAID personnel from the RFS Bureau plus personnel from any regional bureaus responsible for WASH and WRM programming to discuss lessons learned relevant to future USAID programming. Second, URBAN WASH proposes a session devoted to GOI and other

WASH sector stakeholders in Indonesia to discuss lessons learned relevant to advancing Indonesia's goals and action plans for ensuring safely managed access to WASH and climate-resilient cities.[43]

Depending on interest and feasibility, URBAN WASH may pursue additional dissemination in broader sector channels, for example by seeking to publish an academic journal article based on evaluation findings and/or presenting evaluation findings and conclusions at conferences such as World Water Week or the UNC Water and Health Conference.

When final evaluation deliverables are approved by USAID, URBAN WASH will publish them on USAID's Development Experience Clearinghouse and on the Global Waters knowledge management website, further announcing their publication via the institutional websites of the various organizations which participate in the evaluation. URBAN WASH will further submit primary datasets collected in service of the evaluation to USAID's DDL in accordance with ADS 579.3.2.6. This publication will allow other researchers to use the anonymized evaluation datasets for other research and learning purposes. URBAN WASH will ensure that all personal identifying information is removed from the data uploaded to the DDL.

[43] For example, the *Rencana Aksi Nasional* and associated *Rencana Aksi Daerah* for achieving SDGs.

REFERENCES

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in differences estimates?" *The Quarterly Journal of Economics*, Volume 119, Issue 1, February 2004, Pages 249–275, https://doi.org/10.1162/003355304772839588

DAI Global, LLC. 2022. "USAID/Indonesia Urban Resilient, Sanitation, and Hygiene (IUWASH Tangguh): Project Year 1 Work Plan" (USAID/Indonesia, July 2, 2022).

Diamond, Alexis and Jesjeet S. Sekhon. 2013. "Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies." *Review of Economics and Statistics*, 95(3), 932-945.

Ham, Dae Woong and Luke W. Miratrix. 2022. "Benefits and costs of matching prior to a Difference in Differences analysis when parallel trends does not hold." arXiv preprint arXiv:2205.08644

Howard, Guy, Jamie Bartram, Ashley Williams, Alycia Overbo, David Fuente, and Jo-Anne Geere, 2020, "Domestic Water Quantity, Service Level and Health: Second Edition," Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO. https://www.who.int/publications/i/item/9789240015241.

Huntington-Klein, Nick. 2021. *The Effect: an introduction to research design and causality*. CRC Press, a Chapman & Hall Book.

Khandker, Shahidur R.; Gayatri B Koolwal, and Hussain A. Samad,. 2010. *Handbook on Impact Evaluation : Quantitative Methods and Practices*. World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/2693 License: CC BY 3.0 IGO.

McKenzie, David J., and John Gibson. 2010. "The Development Impact of a Best Practice Seasonal Worker Policy," *World Bank Policy Research Working Paper No. 5488*.

Miller, Gregory A. and Jean P. Chapman. 2001. "Misunderstanding Analysis of Covariance," Journal of Abstract Psychology 2002, Vol. 110, No. 1, -40-48. Miller_Chapman_JAP Miller_Chapman_JAP_2001.pdf (arizona.edu)_2001.pdf (arizona.edu)

Prabhakaran, S. 2019. "Mahalanobis Distance–Understanding the math with examples (python)." *Machine Learning Plus*.

Republic of Indonesia. 2015. "Presidential Regulation of The Republic of Indonesia Number 2 of 2015 About Medium-Term Development Plan (RPJMN) 2015-2019."

Republic of Indonesia. 2020. "Appendix Presidential Regulation No. 18 of 2020 Concerning the National Medium-Term Development Plan For 2020-2024," https://perpustakaan.bappenas.go.id/e-library/file_upload/koleksi/migrasi-data-publikasi/file/RP_RKP/Narasi-RPJMN-2020-2024-versi-Bahasa-Inggris.pdf.

Rosenbaum, Paul R. 2010. "Design of Observational Studies," *Springer Series in Statistics*. New York: Springer.

Rubin, Donald B. 1980. ""Bias Reduction Using Mahalanobis-Metric Matching." Biometrics 36 (2): 293–98. https://doi.org/10.2307/2529981

UN Department of Economic and Social Affairs. n.d. "Transforming Our World: The 2030 Agenda for Sustainable Development," accessed October 20, 2022, https://sdgs.un.org/2030agenda.

UNICEF. 1995. *UNICEF Multiple-Indicator Survey Handbook*, Chapter 6.

https://mics.unicef.org/files?job=W1siZiIsIjIwMTUvMDQvMDMvMDYvNDIvNDgvMzgyL2NoYXAwNi5wZGYiXV0&sha=6509e495a61af931

USAID. 2020. "Country Development Cooperation Strategy 2020 - 2025 | Indonesia | U.S. Agency for International Development," Accessed February 24, 2021. https://www.usaid.gov/indonesia/cdcs.

USAID/E3. 2020. "Cost Analysis Guidance for USAID-Funded Education Activities," Office of Education, Bureau for Economic Growth, Education, and Environment (E3), 2020. https://www.edu-links.org/sites/default/files/media/file/USAID%20Cost%20Analysis%20Guidance_Final%20Feb20.pdf

World Health Organization (WHO) and United Nations Children's Fund (UNICEF). 2017. *Safely Managed Drinking Water: Thematic Report on Drinking Water 2017* (Geneva: World Health Organization, 2017), https://apps.who.int/iris/handle/10665/325897

Xiang, S., F. Nie, and C. Zhang. 2008. "Learning a Mahalanobis distance metric for data clustering and classification." *Pattern recognition*, 41(12), 3600-3612.

Young, Sera L., Godfred O. Boateng, Zeina Jamaluddine, Joshua D. Miller, Edward A. Frongillo, Torsten B. Neilands, Shalean M. Collins et al. 2019. "The Household Water In Security Experiences (HWISE) Scale: development and validation of a household water insecurity measure for low income and middle-income countries." *BMJ Global Health* 2019;4:e001750. doi:10.1136/ bmjgh-2019-001750

ANNEX 1: ACKNOWLEDGEMENTS

Authors: Dr. Clifford Zinnes, Risyana Sukarma, Trimo Pamudji Al Djono, Miguel Albornoz, Angelo Cozzubo, and McKinzie Davis; Doug Krieger and Zach Borrenpohl.

The authors would like to acknowledge valuable support, guidance, and feedback from Ryan Mahoney, Jesse Shapiro, Miriam Otoo, and all other members of the URBAN WASH team. They would additionally like to acknowledge and express gratitude to the members of DAI's IUWASH Tangguh Team in Washington, D.C. and Jakarta, Indonesia, along with Jean Paul Petraud from the USAID/Asia Bureau, who participated in co-design sessions and a scoping exercise to ground feasible evaluation designs as presented in this report. The authors extend special thanks to Trigeany Linggoatmodjo, Ade Darmawansyah, and the rest of the team at USAID/Indonesia who hosted the scoping trip and continue to support evaluation co-design and coordination with IUWASH Tangguh and the Government of Indonesia. Finally, the team thanks the various personnel affiliated with PUPR, BAPPENAS, PERPAMSI, the Ministry of Environment and Forestry, and the PDAMs in Bogor District and Depok City who graciously hosted evaluation personnel during a scoping visit and provided information regarding their respective roles in supporting water security and climate-resilient water supply in urban areas of Indonesia.

ANNEX 2: ADDITIONAL BACKGROUND INFORMATION

The table below, reproduced from a physical copy shared by IUWASH Tangguh with the evaluation team during a scoping trip in Indonesia, provides a summary of the 38 treatment sites included in IUWASH Tangguh and the type of support they will receive from the Activity.

NO.	PRO-VINCE	SHORTLISTED LOCATION		WATER-SHED AREA (BOLD = PRIORITY DAS)	TYPE OF SCENARIO SUPPORT				
					Cities and districts will receive full support	Cities and district will receive water supply focus	Cities and district will receive sanitation focus	Cities and district will receive WRM focus	Remarks
1	North Sumatra	1	Medan city	Deli			√	√	• Implementation of regular desludging program • Encourage to implement the KKMA developed under IPLUS
2	North Sumatra	2	Binjai city	Deli		√			• Improve PDAM Performance and expansion of piping network under MEBIDANG
3	North Sumatra	3	Deli Serdang district	Deli		√	√		• Expand the operation of regular desludging • Improve PDAM Performance
4	North Sumatra	4	Pematang Siantar city	Bah Bolon	√				
5	North Sumatra	5	Simalungun district	Bah Bolon		√		√	• Improve PDAM Performance • Conduct climate vulnerability assessment and actions plan
6	Banten	1	Tangerang city	**Cisadane**		√	√		• Expansion of piping network under Karian Regional Water Supply system • Implementation of regular desludging program
7	Banten	2	Tangerang district	**Cisadane** and Ciujung		√	√		• Expansion of piping network under Karian Regional Water Supply system • Implementation of regular desludging program

NO.	PRO-VINCE	SHORTLISTED LOCATION		WATER-SHED AREA (BOLD = PRIORITY DAS)	TYPE OF SCENARIO SUPPORT				
					Cities and districts will receive full support	Cities and district will receive water supply focus	Cities and district will receive sanitation focus	Cities and district will receive WRM focus	Remarks
8	Banten	3	Tangsel district	**Cisadane**		√	√		• Expansion of piping network under Karian Regional Water Supply system • Implementation of regular desludging program
9	DKI Jakarta	1	DKI Jakarta province	**Citarum, Ciliwung**			√	√	• Implementation of regular desludging program • Develop partnership upstream and downstream area
10	West Java	1	Bogor district	**Ciliwung, Cisadane**		√	√	√	• Improve PDAM Performance (continuation of non-revenue water (NRW) & EE Program under PBG) • Implementation of regular desludging program • Replication of KKMARA to other raw water source
11	West Java	2	Depok city	**Ciliwung, Cisadane**		√	√		• Improve PDAM Performance (continuation of NRW & EE Program under PBG) • Implementation of Regular desludging program
12	West Kalimantan (satellite)	1	Pontianak city	**Kapuas**	√				
13	West Kalimantan (satellite)	2	Kubu Raya district	**Kapuas**		√			• Improve PDAM Performance
14	Central Java	1	Surakarta city	**Bengawan Solo**		√	√		

NO.	PRO-VINCE	SHORTLISTED LOCATION		WATER-SHED AREA (BOLD = PRIORITY DAS)	TYPE OF SCENARIO SUPPORT				
					Cities and districts will receive full support	Cities and district will receive water supply focus	Cities and district will receive sanitation focus	Cities and district will receive WRM focus	Remarks
15	Central Java	2	Sukoharjo district	**Bengawan Solo**		√	√		• Improve PDAM Performance (continuation of NRW & EE Program under PBG) • Implementation of regular desludging program
16	Central Java	3	**Karanganyar district**	**Bengawan Solo**		√		√	• Improve PDAM performance and expansion of piping network under WOSOSUKAS • Conduct climate vulnerability assessment of PDAM raw water source
17	Central Java	4	Wonogiri district	**Bengawan Solo**		√	√		• Improve PDAM performance and expansion of piping network under WOSOSUKAS • Conduct climate vulnerability assessment of PDAM raw water source
18	Central Java	5	Sragen district	**Bengawan Solo**		√	√		• Improve PDAM performance and expansion of piping network under WOSOSUKAS • Implementation of regular desludging program
19	Central Java	6	Magelang city	Progo	√				
20	Central Java	7	Temanggung district	Progo		√		√	• Improve PDAM Performance • Conduct climate vulnerability assessment of PDAM raw water source
21	Central Java	8	Salatiga city	Progo	√				

NO.	PRO-VINCE	SHORTLISTED LOCATION		WATER-SHED AREA (BOLD = PRIORITY DAS)	TYPE OF SCENARIO SUPPORT				
					Cities and districts will receive full support	Cities and district will receive water supply focus	Cities and district will receive sanitation focus	Cities and district will receive WRM focus	Remarks
22	East Java	1	Surabaya city	**Brantas (hulu dan hilir)**	√				
23	East Java	2	Sidoarjo district	**Brantas (hulu dan hilir)**		√	√		• Improve PDAM performance and expansion of piping network under UMBULAN Water Supply System • Implementation of regular desludging program
24	East Java	3	Gresik district	**Brantas (hulu dan hilir)**		√	√		• Improve PDAM performance and expansion of piping network under UMBULAN Water Supply System • Implementation of regular desludging program
25	East Java	4	Malang city	**Brantas (hulu dan hilir)**	√				
26	East Java	5	Malang district	**Brantas (hulu dan hilir)**		√		√	• Improve PDAM performance • Conduct climate vulnerability assessment of PDAM raw water source
28	East Java	7	Blitar city	**Brantas (hulu dan hilir)**			√		• Implementation of regular desludging program
29	East Java	8	Pasuruan city	**Brantas (hulu dan hilir)**		√	√		• Improve PDAM performance and expansion of piping network under UMBULAN Water Supply System

NO.	PRO-VINCE	SHORTLISTED LOCATION		WATER-SHED AREA (BOLD = PRIORITY DAS)	TYPE OF SCENARIO SUPPORT				
					Cities and districts will receive full support	Cities and district will receive water supply focus	Cities and district will receive sanitation focus	Cities and district will receive WRM focus	Remarks
30	East Java	9	Pasuruan district	**Brantas (hulu dan hilir)**		√		√	• Improve PDAM performance and expansion of piping network under UMBULAN Water Supply System • Conduct climate vulnerability assessment of PDAM raw water source
31	East Nusa Tenggara (Satellite)	1	Kupang city	Manikin				√	• Conduct climate vulnerability assessment of PDAM raw water source
32	East Nusa Tenggara (Satellite)	2	Timor Tengah Selatan	Manikin				√	• Conduct climate vulnerability assessment of PDAM raw water source
33	South Sulawesi	1	Makassar city	**Jeneberang**		√	√		• Improve PDAM Performance and expansion of piping network under MAMMINASATA Water Supply System • Implementation of regular desludging program
34	South Sulawesi	2	Maros district	**Jeneberang**		√	√		• Improve PDAM Performance and expansion of piping network under MAMMINASATA Water Supply System • Implementation of regular desludging program
35	South Sulawesi	3	Gowa district	**Jeneberang**		√	√		• Improve PDAM Performance and expansion of piping network under MAMMINASATA Water Supply System • Implementation of regular desludging program

NO.	PRO-VINCE	SHORTLISTED LOCATION		WATER-SHED AREA (BOLD = PRIORITY DAS)	TYPE OF SCENARIO SUPPORT				
					Cities and districts will receive full support	Cities and district will receive water supply focus	Cities and district will receive sanitation focus	Cities and district will receive WRM focus	Remarks
36	South Sulawesi	4	Takalar district	**Jeneberang**		√	√		• Improve PDAM Performance and expansion of piping network under MAMMINASATA Water Supply System • Implementation of regular desludging program
37	South Sulawesi	5	Barru district	Karajae		√	√		• Improve PDAM Performance (continuation of NRW & EE Program under PBG) • Implementation of regular desludging program
38	Papua (satellite)	1	Jayapura city	Memberamo		√	√		• Improve PDAM Performance • Implementation of regular desludging program
39	Papua (satellite)	2	Jayapura district	Memberamo		√	√		• Improve PDAM Performance • Implementation of regular desludging program

ANNEX 3: TECHNICAL ANNEXES

3.1 OUTPUT FROM CITY/DISTRICT-LEVEL STATISTICAL MATCHING

BACKGROUND, DATA SOURCES, AND SPECIFICATIONS

URBAN WASH deployed statistical matching techniques to identify cities and districts, and their corresponding PDAMs, which will comprise the comparison group to evaluate the IUWASH Tangguh Activity in Indonesia. From this set of cities and districts, neighborhoods and households will be selected to complete the comparison group selection.

The data used for this exercise comes from two sources:

- The Kinerja BUMDES Air Minum Reports 2016-2021: these annual reports issued by Indonesia's Ministry of Public Works and Housing (PUPR) aggregate and report on administrative, financial, and operational performance of PDAMs throughout the country.
- IUWASH Tangguh's site selection dataset, based on which treatment cities and districts were selected for the Activity. This dataset aggregates data from multiple sources, including the 2020 Kinerja BUMDES Air Minum report, BPS and World Bank survey statistics, and others.

To assemble the statistical matching dataset, URBAN WASH transformed the PUPR datasets from a PDF format to a machine-readable Excel format, translated and filtered common variables across at least the 2018-2021 datasets, and merged the datasets between years using the PDAM names. This yielded a consolidated PDAM performance dataset at the PDAM level, with repeated measurements for variables between 2016-2021, where available. Minimal errors were present in the PDAM name variable from year to year, which URBAN WASH corrected manually. PUPR instructed URBAN WASH that the 2021 dataset was not final and should not be used. So, URBAN WASH treated the 2020 variables as the values of record for the statistical matching exercise. URBAN WASH cleaned the dataset and verified that the distribution of key matching variables was similar between 2019 and 2020.

URBAN WASH merged this consolidated PDAM dataset into the IUWASH Tangguh site selection dataset using the name of the city or district where each PDAM was located. In two cases (DKI Jakarta and Tangerang Selatan), a single PDAM matched to multiple cities or districts in the site selection dataset. The final matching dataset comprises 241 cities and districts with their corresponding PDAMs and 755 variables, many of which are repeated annual measurements from 2016-2021.

URBAN WASH selected the variables in Table 14 for statistical matching.

Table 14: Statistical Matching Variables, Cities, and Districts

VARIABLE	DEFINITION	UNITS	SOURCE
Urban Area Classification	Urban area is a city or a district	Categorical	IUWASH Tangguh
Province	Province in which city or district resides	Categorical	IUWASH Tangguh
Poverty rate	Proportion of city or district population whose income falls below the poverty line	Percent	IUWASH Tangguh
Households with access to improved sanitation	Proportion of city or district households whose sanitation is defined as "improved" according to WHO/JMP service ladder	Percent	IUWASH Tangguh

VARIABLE	DEFINITION	UNITS	SOURCE
Domestic customer coverage rate for PDAM, 2020	The proportion of households in the PDAM service area who are connected to the PDAM	Percent	PUPR
Population in the PDAM working area, 2020	Population in the PDAM working area	People	PUPR
Average tariff rate for PDAM customers, 2020	Average tariff rate for all PDAM customers	Rupiah/m3	PUPR
Solvency of the PDAM, 2020	PDAM assets less PDAM equity and liabilities	Rupiah	PUPR
Ratio of local government contribution to total assets for PDAM, 2020	Ratio of grants and/or equity contributed to PDAM budget from local government to total PDAM assets	Ratio	PUPR
PDAM receipt of World Bank NUWSP investment, 2020	PDAM has been the target of investment (in any amount) from the World Bank NUWSP program as of 2020	Binary (0/1)	World Bank
PDAM production volume, 2020	The total volume of water which PDAM has abstracted from all of its sources combined over the year	$Meters^3/year$	PUPR
PDAM transmission pipe length, 2020	Length of the PDAM's transmission pipelines	Meters	PUPR
PDAM water loss rate (i.e., non-revenue water), 2020	Proportion of PDAM water produced that is lost before arriving to end users	Percent	PUPR
Overall PDAM performance score for the 2020 fiscal year, 2020	Aggregated overall PDAM performance score using PUPR methodology. Range is 0.00 – 5.00 points. PDAMs scored below 2.2 are considered "sick," between 2.2 and 2.8 are considered "less healthy," and above 2.8 are considered "healthy."	Points	PUPR
PDAM customer growth rate year over year, 2020	The percentage increase in PDAM customers in the current year relative to the previous year	Percent	PUPR
PDAM operating hours as proportion of 24-hour day, 2020	Self-reported hours per 24-hour day that PDAM provides water service, expressed as a percentage (i.e., 18 hours = 75%)	Percent	PUPR
Volume of water abstracted from surface water sources, 2020	Volume of water abstracted from sources such as rivers, streams, lakes, or reservoirs	Liters per second	PUPR
Volume of water abstracted from spring sources, 2020	Volume of water abstracted from springs	Liters per second	PUPR
Volume of water abstracted from groundwater sources, 2020	Volume of water abstracted from groundwater sources, for example through boreholes or deep wells	Liters per second	PUPR

URBAN WASH tested matching according to multiple specifications. As discussed during evaluation co-design and proposed in the inception report, URBAN WASH excluded the six treatment sites in the Papua, East Nusa Tenggara, and West Kalimantan provinces from the matching exercise since there were no reasonably similar cities or districts within IUWASH Tangguh's intervention provinces. Initial matching attempts included treatment sites in the DKI Jakarta province. However, URBAN WASH found that including these sites in matching damaged the balance between treatment and comparison groups. Indeed, Jakarta is by far the largest city in Indonesia and the only one that is also a province. Accordingly, in URBAN WASH's final matching exercise, Jakarta is excluded.

Across all its different matching specifications, URBAN WASH constrained the algorithm to exact matching on urban area classification (i.e., cities may only match with cities and districts may only match with districts). After excluding Jakarta from matching, the main variation in matching methods URBAN WASH tested was whether or not Province should be treated as an exact matching characteristic (i.e., sites may only be matched to sites within the same province). URBAN WASH's matching algorithm uses Mahlanobis distance metrics for all other variables. URBAN WASH found that restricting matches to fall within the same provinces either required removing additional cities/districts from the study or substantially reducing the balance between the treatment and comparison groups. In other words, allowing some cities and districts to match with cities and districts in other provinces permits the study to include the remaining 31 treatment sites in the study and improves the overall balance between the treatment and comparison groups. As such, URBAN WASH elected this specification (i.e., exact matching on urban area classification type and Mahlanobis distance matching on all other variables, including province) as its final matching algorithm and yielded the outputs in this report using the "MatchIt" package on R. The next section describes these methods and outputs in more detail.

METHODS

URBAN WASH employed pre-treatment characteristics of the treated cities, districts, and their corresponding PDAMs to choose the most similar untreated sites within the same set of provinces. The selected comparison cities and districts are the ones that minimize a multivariate distance metric, making them the "most similar." The distance measure is used to define how close two units are, and in nearest neighbor matching, this is used to choose the nearest control unit to each treated unit. URBAN WASH considered using Euclidean or Manhattan distance metrics, but ultimately selected the Mahlanobis distance metric given its ability to handle (i) variables with potentially different distributions and (ii) dimensions that are far from being independent of each other. The Euclidean and Manhattan metrics fail when there is high correlation between variables. On the contrary, the Mahalanobis distance can transform the dimensions into uncorrelated indicators, scale back their variance to one, and compute the Euclidean distance over these indicators.[44] Nearest neighbor matching algorithms use functions like the following:

$$\delta(x_i, x_j) = \sqrt{(x_i - x_j)' \, S^{-1} \, (x_i - x_j)}$$

where x is a p×1 vector containing the value of each of the p included covariates for that unit and S^{-1} is the (generalized) inverse of a scaling matrix. For the Mahalanobis distance, S is the pooled covariance matrix of the covariates, while the robust Mahalanobis distance is calculated using the ranks of the covariates and uses a correction for ties.[45] When using this distance metric, it is no longer necessary to standardize the data as the scaling matrix takes care of this.

Using the Mahalanobis distance in URBAN WASH's application is a considerable improvement as the algorithm computes distances from a set of treatment site characteristics, which are highly likely to correlate between them. URBAN WASH restricts choosing comparison units without replacement,

---

[44] Xiang, S., Nie, F., & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. Pattern recognition, 41(12), 3600-3612.
Prabhakaran, S. (2019) Mahalanobis Distance–Understanding the math with examples (python). Machine Learning Plus.

[45] Rubin, Donald B. 1980. "Bias Reduction Using Mahalanobis-Metric Matching." Biometrics 36 (2): 293–98. https://doi.org/10.2307/2529981.
Rosenbaum, Paul R. 2010. Design of Observational Studies. Springer Series in Statistics. New York: Springer.

ensuring that an untreated city or district cannot be selected as a match for more than one treated city or district.

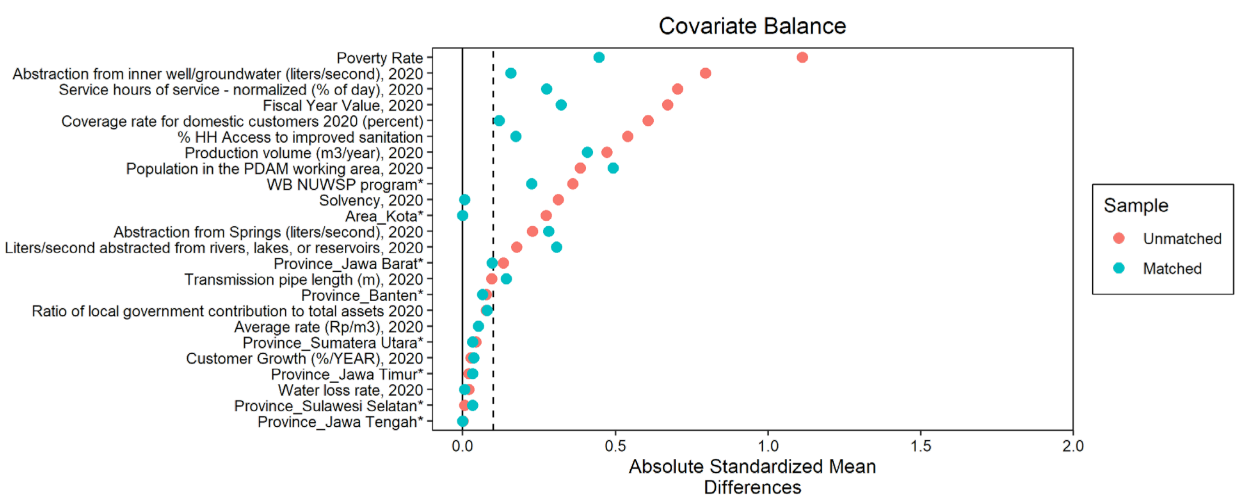
URBAN WASH employed a genetic matching model rather than a standard nearest neighbor matching model. Genetic matching algorithms search a range of distance metrics to find the measure that optimizes post-matching covariate balance, where each distance metric considered corresponds to a particular assignment of weights *W* for all matching variables.[46] This algorithm extends the previously presented matching by minimizing a generalized version of the Mahalanobis distance. In this extended version, the distance includes a variable-specific weight parameter that values each variable according to its relative importance for achieving the best overall balance.

Finally, URBAN WASH elected to only use matching covariates which resemble the treatment selection criteria and which are correlated with intended outcomes. URBAN WASH omits variables which approximate intended outcomes (such as proportion of the city/district with access to safely managed water, distribution volume per customer, and domestic sales in Indonesian rupiah per customer) as these may generate bias.[47]

RESULTS

Figure 8 plots the absolute standardized mean differences in URBAN WASH's final matching algorithm prior to and following matching. The standardized mean difference is an indication of the difference between two groups' mean values for a covariate divided by an estimate of the within-group standard deviation for that variable. It expresses how different the groups are taking into account the variability in the underlying metric. Perfectly balanced groups would have a standardized mean difference of 0. The chart shows that the matching algorithm selects groups which substantially reduce differences between treatment and comparison sites on nearly all metrics.

Figure 8: Love Plot Pre- and Post-Matching



46 Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. Review of Economics and Statistics, 95(3), 932-945.

47 Ham, D. W., & Miratrix, L. (2022). Benefits and costs of matching prior to a Difference in Differences analysis when parallel trends does not hold. arXiv preprint arXiv:2205.08644.

Table 15 shows the values in natural units for matching covariates before and after the matching exercise, along with the corresponding standard mean differences.

Table 15: Comparison of Covariate Means for Treatment and Comparison Cities and Districts

COVARIATE MEAN		UNMATCHED		MATCHED	
	Treated	*Comparison*	*Standardized Mean Difference*	*Comparison*	*Standardized Mean Difference*
Poverty Rate	7.47	10.45	-1.11	8.67	-0.45
% HH Access to improved sanitation	77.00%	67.22%	0.54	73.86%	0.17
Domestic coverage rate	30.74	20.54	0.61	28.73	0.12
Population in PDAM area	1134458.52	779392.74	0.39	680346.81	0.49
Average tariff (Rp/m^3)	4825.19	4748.04	0.05	4748.39	0.05
Solvency	4365.07	6948.41	-0.31	4415.04	-0.01
Ratio of LG contribution to assets	0.83	0.76	0.08	0.76	0.08
% which received NUWSP investment	54.84%	18.81%	0.72	32.26%	0.45
Prod. volume (millions m^3/year)	49.31	12.67	0.47	17.70	0.41
Transmission pipe length (m)	118982.03	89885.28	0.09	75359.84	0.14
Water loss rate (NRW)	30.92	30.69	0.02	30.84	0.01
PDAM Performance Score	3.38	3.01	0.67	3.20	0.32
Customer growth rate	7.28%	7.50%	-0.03	6.98%	0.04
% of day with water service	92.84	87.56	0.70	90.77	0.28
% in Banten	9.68%	1.98%	0.26	3.23%	0.22
% in Jawa Barat	6.45%	19.80%	-0.54	16.13%	-0.39
% in Jawa Tengah	25.81%	25.74%	0.00	25.81%	0.00
% in Jawa Timur	25.81%	23.76%	0.05	22.58%	0.07
% in Sulawesi Selatan	16.13%	16.83%	-0.02	19.35%	-0.09
% in Sumatera Utara	16.13%	11.88%	0.12	12.90%	0.09
Surface water abstraction (l/second)	6123.94	3380.15	0.18	1353.77	0.31
Spring abstraction (l/second)	988.10	407.22	0.23	275.03	0.28
Groundwater abstraction (l/second)	105.23	311.37	-0.80	146.10	-0.16
Area type: Kabupaten	54.84%	82.18%	-0.55	54.84%	0.00
Area type: Kota	45.16%	17.82%	0.55	45.16%	0.00

Prior to matching, treated cities and districts have lower poverty rates, higher access to improved sanitation, and better performing PDAMs serving larger populations than untreated cities and districts. Untreated cities and districts are much more reliant on groundwater sources than treated cities and districts. Over half of PDAMs in treated cities and districts received investment from the World Bank National Urban Water Supply Project (NUWSP), compared to less than one-fifth of those in untreated cities and districts. URBAN WASH's matching exercise substantially reduces most of these differences, though some differences remain and must be controlled for in the analysis of program impact. Although all comparison sites come from the same set of provinces as the treatment sites, there are relatively more cities and districts from West Java and South Sulawesi in the comparison group and relatively fewer in Banten, East Java, and North Sumatra than in the treatment group. Table 16 presents the final set of treatment and comparison cities and districts selected by URBAN WASH's statistical matching algorithm.

Table 16: Treatment and Comparison Cities and Districts

PAIR ID	TREATED SITE	COMPARISON SITE	PROVINCE(S)
1	Kabupaten Wonogiri	Kabupaten Pati	Jawa Tengah
2	Kota Binjai	Kota Mojokerto	Sumatera Utara/Jawa Timur
3	Kota Blitar	Kota Semarang	Jawa Timur/Jawa Tengah
4	Kota Depok	Kota Bekasi	Jawa Barat
5	Kota Magelang	Kota Probolinggo	Jawa Tengah/Jawa Timur
6	Kota Makassar	Kota Parepare	Sulawesi Selatan
7	Kota Malang	Kota Sibolga	Jawa Timur/Sumatera Utara
8	Kota Medan	Kota Tebingtinggi	Sumatera Utara
9	Kota Pasuruan	Kota Pekalongan	Jawa Timur/Jawa Tengah
10	Kota Pematangsiantar	Kota Palopo	Sumatera Utara/Sulawesi Selatan
11	Kota Salatiga	Kota Tegal	Jawa Tengah
12	Kota Surabaya	Kota Bogor	Jawa Timur/Jawa Barat
13	Kota Surakarta	Kota Bandung	Jawa Tengah/Jawa Barat
14	Kota Tangerang	Kota Banjar	Banten/Jawa Barat
15	Kabupaten Bogor	Kabupaten Bandung	Jawa Barat
16	Kabupaten Deli Serdang	Kabupaten Asahan	Sumatera Utara
17	Kabupaten Gowa	Kabupaten Sinjai	Sulawesi Selatan
18	Kabupaten Gresik	Kabupaten Magetan	Jawa Timur
19	Kabupaten Karanganyar	Kabupaten Magelang	Jawa Tengah
20	Kabupaten Malang	Kabupaten Banyuwangi	Jawa Timur
21	Kabupaten Maros	Kabupaten Luwu Utara	Sulawesi Selatan
22	Kabupaten Pasuruan	Kabupaten Mojokerto	Jawa Timur
23	Kabupaten Sidoarjo	Kabupaten Bojonegoro	Jawa Timur
24	Kabupaten Simalungun	Kabupaten Langkat	Sumatera Utara
25	Kabupaten Barru	Kabupaten Toraja Utara	Sulawesi Selatan
26	Kabupaten Sragen	Kabupaten Kendal	Jawa Tengah
27	Kabupaten Sukoharjo	Kabupaten Demak	Jawa Tengah
28	Kabupaten Takalar	Kabupaten Luwu Timur	Sulawesi Selatan
29	Kota Tangerang Selatan	Kota Kediri	Banten/Jawa Timur
30	Kabupaten Tangerang	Kabupaten Pandeglang	Banten
31	Kabupaten Temanggung	Kabupaten Batang	Jawa Tengah

3.2 DESCRIPTION OF DIFFERENCE-IN-DIFFERENCES ANALYSIS

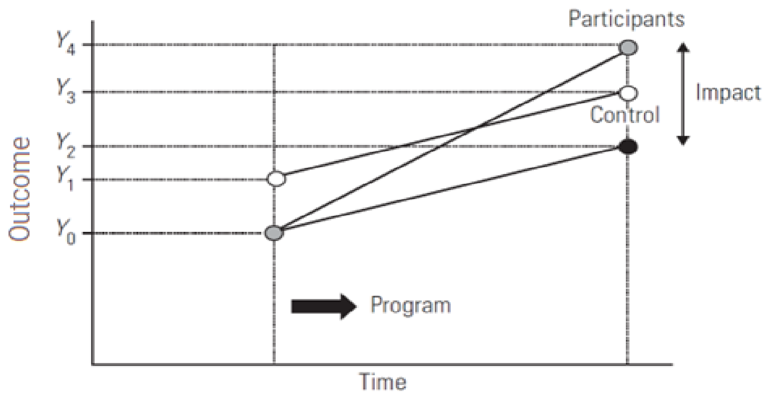
The evaluation team's preferred method, the DID model, is one of the most popular methodologies for applied research in economics. To answer the hypotheses, DID estimates causal relationships among variables by comparing the difference in outcomes before and after an intervention between groups of beneficiaries and nonparticipants (Bertrand et al. 2004). The method takes its name as the first "difference" is obtained by comparing the unit before (baseline) and after the intervention (endline), while the second "difference" is computed between the beneficiary group (treatment) and nonparticipant group (control). Thus, two or more rounds of data are required (see Figure 9).

The main advantage of this approach is that it considers both observed and unobserved factors, reducing endogeneity problems and controlling for both types of variables in the analysis (Bertrand et al. 2004; Khander et al. 2009). The main requirement of this method is that the (conditional) parallel trends assumption holds, and this assumption states that treated units would have shown the trend in the outcome as the control units if they had not been treated.

As this method relies on differentiation through time to estimate the impacts, it is robust to any pre-treatment differences in the outcome variables. Moreover, as we compute the double difference by contrasting each group against itself in the previous period, the method eliminates any time-invariant characteristics that may cause bias (i.e., culture, deep-rooted attitudes, etc.). However, the strength of this method to correctly estimate the policy impacts relies on ensuring that both treatment and control groups are similar (Khander et al. 2009). As Cunningham (2021) highlights, DID only works if the comparison group is good in terms of parallel trends, which is an assumption that cannot be tested directly. For this reason, the impact estimate benefits from the matching procedure, as we select the most similar untreated units in the pre-treatment period as the control group.

The combination of matching and DID satisfies some desired properties. First, as the DID has group fixed effects, it already controls for any differences between treated and untreated groups constant over time. However, DID is insufficient to analyze a potential treatment selection bias: why specific units were selected to receive the treatment and others were not. This selection may bias our results if any back door between receiving the intervention and the evolution of the outcome after the treatment remains. "That's where matching comes in. If we can pick a set of matching variables that close the back doors between which groups become treated and when and the outcome, we get parallel trends back" (Huntington-Klein 2021).

Figure 9: Graphic Representation of DID



Source: Khandker, Koolwal & Samad (2009)

For our specific application, we exploit the differential application of the intervention as it was not applied in all the PDAMs (or service areas) nor started simultaneously everywhere. This distribution of the treated and untreated (or control) units let us compare them before and after the program was applied, thus having a heterogeneous starting point for the treated PDAMs.

To identify the parameter of interest, we will employ the dataset comprised of households in treated service areas and households in the corresponding matched untreated localities. The DID can be computed following the regression specification by Wooldrige and Imbens (2007) for the case of panel data, many periods, and arbitrary treatment patterns specified as

$$Y_{i,t} = \lambda_t + c_i + \beta_1 D_{i,t} + X_{i,t}\delta + \varepsilon_{i,t}\ , \qquad t = 1, \ldots, T$$

where $Y_{i,t}$ is the outcome variable for unit i at time t, $D_{i,t}$ is the treatment status dummy, and β_1 is the treatment effect. X_i is a matrix of relevant covariates to increase the precision of β_1; while λ_t and c_i denote the time and individual fixed effects, respectively. It is essential to highlight that, while ideal, the DID model does not require the data collection effort to interview the exact same individuals (respondents panel). If we maintain the same localities for both baseline and endline data collection, the DID model can be computed as

$$Y_{ij,t} = \lambda_t + c_j + \beta_1 D_{i,t} + X_{i,t}\delta + \varepsilon_{i,t}\ , \qquad t = 1, \ldots, T$$

where the subindex j denotes the localities, and the individual fixed effect has been replaced by a locality-level fixed effect c_j. As Huntington-Klein (2011) noted, the computation of variances can be adjusted for the uncertainty introduced by the matching process by using a numerical approximation through bootstrap standard errors.

As an extension of the DID, if pre-treatment data and sample size allow, we will consider a doubly robust DID (DRDID) as a robustness test of our estimates. This method, proposed by SantAnna and Zhao (2018), leverages pre-treatment information to model a propensity score of the probability of being treated. The DRDID application exploits this propensity score function as a first-stage estimation before computing the DID estimate. As the authors demonstrate, doubly robust methods will yield unbiased estimates if one of these two regressions is estimated consistently. They will also produce efficient estimates if both are estimated consistently. Moreover, the technique can be applied either with panel or repeated cross-section data, making it "less demanding in terms of the researcher's ability to correctly specify models" (SantAnna and Zhao 2018).

3.3 DESCRIPTION OF ANCOVA ANALYSIS

An alternative method for evaluating the policy impact is ANCOVA. This technique is a statistical method based on variance, multiple regression, and correlation analysis used to increase the precision of comparison between groups and reduce the probability of Type II errors, i.e., when a false null hypothesis is not rejected (Miller and Chapman 2001; Huck 2012). ANCOVA is thought to improve statistical power as long as (i) the relationship between the dependent variable and the covariate within each group is linear and parallel, (ii) the covariate is unaffected by other independent variables, and (iii) if data are collected under a completely randomized design in the least one wave before any treatment is applied (Schwarz 2015). Moreover, McKenzie (2012) argues that non-experimental interventions may also benefit from ANCOVA. Gibson and McKenzie (2010) apply it to a matched DID of a seasonal migration program where multiple rounds of follow-up data are averaged to get more precise measurements of consumption and income. In this context, the baseline data becomes highly beneficial as it allows control for baseline differences across treated units.

When complying with these assumptions, ANCOVA can have higher explanatory power than DID only if autocorrelation is low. We will test this latter condition with the survey data to evaluate if the ANCOVA is a suitable method. In the context of this evaluation, ANCOVA takes advantage of the low autocorrelation of certain outcome variables in this study to improve power beyond what a DID approach can attain with the same sample size. Baseline data for these outcome measures have little predictive power for future outcomes, so it is inefficient to entirely correct baseline imbalances between treatment and control groups using DID. Instead, an ANCOVA model can adjust the degree of correction for baseline difference in means according to the correlation between past and future outcomes observed in the data (McKenzie 2012).

The ANCOVA specification can be written as

$$Y_{i,t} = \beta_0 + \beta_1 D_i + \beta_2 Y_{i,t-1} + \delta X_i + \varepsilon_{i,t}$$

In this case, $Y_{i,t}$ and $Y_{i,t-1}$ denote the endline and baseline value of the outcome variable, respectively; while. β_1 is the ANCOVA estimate of the treatment effect. As before, D_i is the treatment status dummy and X_i is a matrix of relevant covariates.

3.4 SIMULATION OF HOUSEHOLD WATER SECURITY INDEX CALCULATION

This annex illustrates how two hypothetical households would perform on URBAN WASH's proposed Household Water Security Index, outlined in Section 3.2 of this report. Each household scenario shows 1) the hypothetical household's water supply at baseline and at endline; 2) the calculation of the hypothetical household's index score at baseline and at endline; and 3) a visual representation of the hypothetical household's baseline and endline index score.

HOUSEHOLD 1:

Household 1 has four members with an annual income between 105 million IDR and 125 million Indonesia Rupiah (IDR). At baseline, their drinking water tests positive *E. coli* at the point of consumption and negative at the point of collection. At endline, their drinking water tests negative for *E. coli* at point of consumption and negative at the point of collection. Other characteristics of their baseline and endline water supply are described below. This hypothetical household demonstrates a causal pathway for improved household water security through improvements in the quality of existing water services.

Table 17: Household 1 Water Consumption

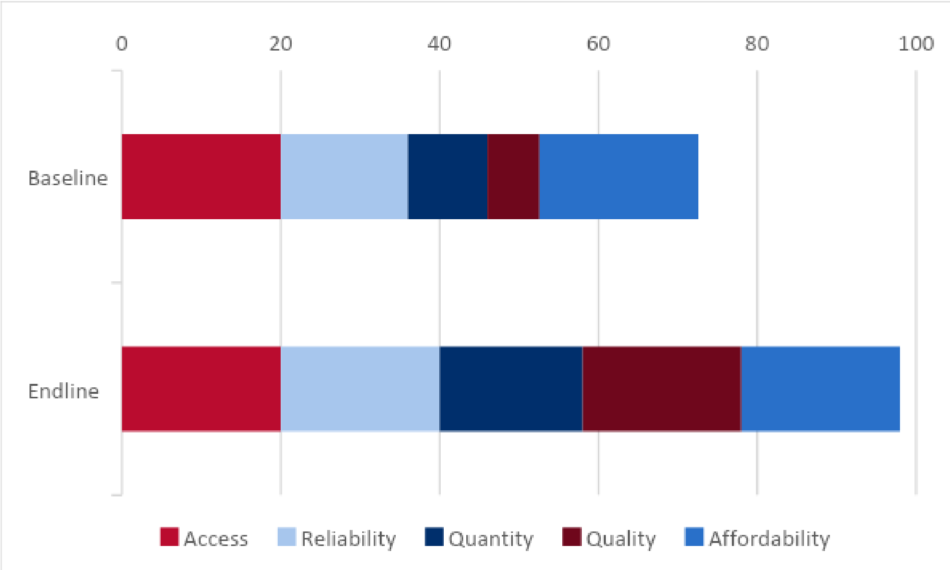
WATER SOURCE:	TIME	USES(S)	ACCESS	RELIABILITY	CONSUMPTION	EXPENDITURE
PDAM tap *(main drinking source at baseline and endline)*	Baseline	Drinking, cooking, cleaning, washing	On premises, available when needed	Disrupted once in past week	6.0 cubic meters per month	**24,000 IDR per month**
PDAM tap *(main drinking source at baseline and endline)*	Endline	Drinking, cooking, cleaning, washing	On premises, available when needed	Not disrupted in past week	7.2 cubic meters per month	**32,000 IDR per month**
Rainwater	Baseline	Gardening	On premises, available when needed	N/A	Fills 10-liter container twice per month	**0 IDR**
Rainwater	Endline	Gardening	On premises, available when needed	N/A	Fills 10-liter container twice per month	**0 IDR**
Bottled water	Baseline	Drinking	<30 minutes round trip	N/A	Buys 10 2-liter bottles per week	**5,000 IDR per bottle**
Bottled water	Endline	Drinking	<30 minutes round trip	N/A	Buys 5 2-liter bottles per week	**5,000 IDR per bottle**

Given these inputs, Table 18 and Figure 10 show how Household 1 would score on the Household Water Security Index at baseline and endline.

Table 18: Household I Household Water Security Index Scoring

COMPONENT	TIME	SCORE	RATIONALE
Access	Baseline	100	**Improved** source On premises **Available** when needed
	Endline	100	**Improved** source On premises **Available** when needed
Reliability	Baseline	80	Service disrupted in **one of last seven** days
Reliability	Endline	100	**No disruptions** last seven days
Quantity	Baseline	50	[(6 m^3 / 30 days) x 1000 L/m^3] + [(10 Lx 2 containers) / 30 days] + [(10 bottles x 2 L)/ 7 days] = 203.5 L per day for the household / 4 household members = **50.8L per capita per day**
Quantity	Endline	90	[(7.2 m^3 / 30 days) x 1000 L/m^3] + [(10L x 2 containers)/ 30 days] + [(5 bottles x 2L)/ 7 days) = 242.1 L per day for the household / 4 household members = **60.5 L per capita per day**
Quality	Baseline	33	**Presence** of e. coli at point of consumption, absent at point of collection
Quality	Endline	100	**Absence** of e. coli at point of consumption
Affordability	Baseline	100	24,000 IDR + 0 IDR + (42 bottles x 5,000 IDR) = 234,000 IDR per month *(water expenditure)* 8.75 million IDR – 10.4 million IDR per month *(income)* (234,000 IDR / 8.75 million IDR) x 100 = 2.7% **Water expenditure no more than 4.0% of total income**
Affordability	Endline	100	32,000 IDR + 0 IDR + (21 bottles x 5,000 IDR) = 137,000 IDR per month *(water)* 8.75 million IDR – 10.4 million IDR per month *(income)* (137,000 IDR / 8.75 million IDR) x 100 = 1.5% **Water expenditure no more than 4.0% of total income**

Figure 10: Household I Household Water Security Index



HOUSEHOLD 2:

Household 2 has five members with an annual income between 20 million IDR and 25 million IDR. At baseline, their drinking water tests positive *E. coli* at the point of consumption and positive at the point of collection. At endline, their drinking water tests positive for E. coli at point of consumption and negative at the point of collection. Other characteristics of their baseline and endline water supply are described below. This hypothetical household demonstrates a causal pathway for improved household water security through increased access to basic water services.

Table 19: Household 2 Water Consumption

WATER SOURCE:	TIME	USES(S)	ACCESS	RELIA-BILITY	CONSUMP-TION	EXPENDI-TURE
Shallow, unprotected well	Baseline	Cooking, cleaning, washing	On premises, not available when needed	No disruptions this week	Fills a 5L bucket about 12 times per day	**0 IDR**
Shallow, unprotected well	Endline	Cooking, cleaning, washing	On premises, available when needed	N/A	Fills a 5L bucket about 10 times per day	**0 IDR**
Protected spring *(main drinking source at baseline)*	Baseline only	Drinking	>30 min round trip, not always available	Disrupted twice in the last week	Fills a 20L jerry can 2 times per day	**0 IDR**
Rainwater	Baseline only	Drinking	On premises, not always available	Disrupted all of last week	Fills a 10L container twice per month	**0 IDR**

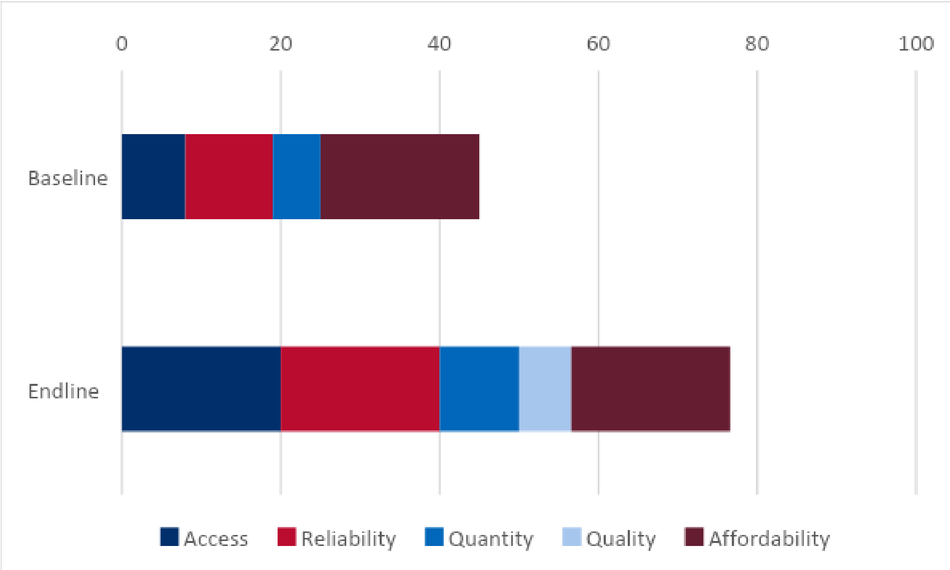
WATER SOURCE:	TIME	USES(S)	ACCESS	RELIA-BILITY	CONSUMP-TION	EXPENDI-TURE
PDAM tap *(main drinking source at endline)*	Endline only	Drinking	On premises, available when needed	No disruptions last week	6 cubic meters per month	**24,000 IDR**

Table 20 and Figure 11 show how Household 2 will score on the Household Water Security Index at baseline and endline.

Table 20: Household 2 - Household Water Security Index Scoring

COMPONENT	TIME	SCORE	RATIONALE
Access	Baseline	40	**Improved** source **Over** 30 minutes roundtrip from household **Not available** when needed
Access	Endline	100	**Improved** source On premises **Available** when needed
Reliability	Baseline	55	Disrupted **two of last seven** days
Reliability	Endline	100	**No disruptions** last seven days
Quantity	Baseline	30	(5L x 12 containers) + (20L x 2 containers) + [(10L x 2 containers)/ 30 days] = 100.7 L for the household per day / 5 household members = **20.1 L per person**
Quantity	Endline	50	(5L x 10) + [(6m^3 / 30days) x 1000 L/m^3) = 250.0 L for the household per day / 5 household members = **50.0 L per person**
Quality	Baseline	0	**Presence** of e. coli at point of consumption **and** point of collection
Quality	Endline	33	**Presence** of e. coli at point of consumption, **negative** at point of collection
Affordability	Baseline	100	0 IDR + 0 IDR + 0 IDR = 0 IDR per month *(water)* 1.6 million IDR – 2 million IDR per month *(income)* (1.6 million IDR / 20 million IDR) x 100 = 0% **Water expenditure is no more than 4.0% of total income**
Affordability	Endline	100	0 IDR + 24,000 IDR = 24,000 IDR per month *(water)* 1.6 million IDR – 2 million IDR per month *(income)* (24,000 IDR / 1.6 million IDR) x 100 = 1.5% **Water expenditure is no more than 4.0% of total income**

Figure 11: Household 2 Household Water Security Index



ANNEX 4: MEMORANDUM OF UNDERSTANDING

See PDF version for executed memorandum of understanding (MOU) between USAID URBAN WASH and USAID IUWASH Tangguh.

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20523
Tel: (202) 712-0000
Fax: (202) 216-3524
www.usaid.gov