“Impact Evaluation on the Effects of the R-WASH Program on Social Cohesion”

Baseline Report

September 29, 2023

Prepared by Prof. Dr. Anselm Hager, Prof. Dr. Carlo Koos, Bernadette Schulz, and Innocent Yekeye.

Contents

List of Figures

List of Tables

Acronyms

AMISOM African Union's Peace Support Mission for Somalia.

BMZ German Federal Ministry for Economic Cooperation and Development.

BR Baseline Report.

DAG Directed Acyclic Graphs.

DID Difference-in-Differences Methodology.

EQ Evaluation Question.

ERG Evaluation Reference Group.

EU European Union.

GDPR General Data Protection Regulation.

GOE Government of Ethiopia.

GOSO Government of Somalia.

GOSU Government of Sudan.

HRWS Human Right to Water and Sanitation.

IDP Internally Displaced Persons.

IP Implementing Partner.

KFW Kreditanstalt für Wiederaufbau.

MDG United Nations Millennium Development Goals.

NGO Non-Governmental Organisation.

PI Principal Investigator.

R-WASH Regional Program on Water and Sanitation for Refugees, Internally Displaced Persons and Host Communities in East Africa.

RCT Randomized Controlled Trial.

RIE Rigorous Impact Evaluation.

SDGs Sustainable Development Goals.

ToC Theory of Change.

ToR Terms of Reference.

UN United Nations.

UNHCR United Nations High Commissioner for Refugees.

UNICEF United Nations Children's Fund.

WASH Water, Sanitation and Hygiene.

1 EXECUTIVE SUMMARY

Access to clean water, adequate sanitation facilities, and proper hygiene practices are considered fundamental development goals and essential to achieve several Sustainable Development Goals (SDGs), including those related to health, poverty eradication, gender equality, and environmental sustainability. The UN acknowledges access to water and sanitation as a human right (HRWS), vital for the well-being, respect, and advancement of all individuals. The recognition of HRWS is reflected in various international legal documents, including human rights treaties, declarations, and established norms. Still, a significant portion of the global population lacks access to properly managed water and sanitation facilities (Beard and Mitlin, 2021). The situation is particularly dire in urban centers and regions faced with desertification and displacement, in particular at the Horn of Africa (Hirwa et al., 2021).

To tackle shortages of water in appropriate quality and to improve sanitation, development organizations implement so-called Water, Sanitation and Hygiene programs (WASH). Through which pathways such programs work and whether they can improve broader societal relations and forge peace when water is scarce, however, remains an open question (Hutton et al., 2004). Most programs have not been evaluated using rigorous modern evaluation methods. By evaluating WASH projects with modern evaluation techniques, however, policymakers can assess the effectiveness of such interventions, identify gaps in service delivery, and develop evidence-based policies and strategies to improve global WASH conditions (UNICEF, 2016). Additionally, understanding the impact of WASH initiatives on social and economic factors, such as education, incomes, and community cohesion, allows the international community to address broader development challenges and promote sustainable and inclusive societies.

The Regional Program on Water and Sanitation for Refugees, Internally Displaced Persons and Host Communities in East Africa (R-WASH)—financed by KfW and carried out by UNICEF and UNHCR—aims to enhance the WASH infrastructure in selected displacement camps and host communities in three African countries—Ethiopia, Somalia, and Sudan—through three primary outputs. R-WASH has an overall budget of 32 million Euros and is scheduled to conclude in 2027. First, R-WASH focuses on constructing integrated and climate-resilient water infrastructure for refugees, internally displaced persons (IDPs), and host communities. During the initial funding phase, KfW plans to finance various measures at three different sites. These include structural measures like well drilling, the establishment of piped drinking water systems, distribution pipelines, and standpipes. Second, R-WASH aims to strengthen the capacity of local operators to deliver sustainable, efficient, and effective services. Third, R-WASH will provide investment-related consulting services for technical operations, maintenance, financial management, personnel management, and environmental and social compatibility monitoring and evaluation.

Each of the three R-WASH country sites will receive tailored measures based on site-specific development needs. In Ethiopia, R-WASH focuses on expanding water infrastructure in the camps and nearby communities and providing operational support in establishment and capacity building of two water and sanitation service providers (utilities) for Kebrebeyay and Aw-Barre/Shedder. In Somalia, R-WASH efforts concentrate on expanding existing water infrastructure, along with operational support for the Doolow Water Management Company. In Sudan[1], R-WASH aims to improve the water supply system to align with the urban system, and provide operational support to the local state-run utility.

A key innovation of R-WASH is a rigorous impact evaluation (RIE), relying on a mixed-method difference-in-differences design (DID). The purpose of the RIE is threefold. First, the evaluation aims to rigorously assess the impact of R-WASH on social cohesion and peaceful coexistence within the camp community, the host community, and between the two communities. Second, in so doing, the RIE will develop a comprehensive Theory of Change (ToC) to inform future policy-making. Third, the RIE will aid the implementing partners and donors to ascertain accountability by providing information about the positive—and potential unintended negative—effects of R-WASH on social cohesion. The objective of the commissioned evaluation is, in short, to investigate how R-WASH can enhance social cohesion and peaceful coexistence within the camp community, the host community, and between the two communities.

The purpose of the present Baseline report (BR) in the evaluation process is to develop credible hypotheses for how R-WASH affects social cohesion as well as testing if and what adjustments may be required to make control and treatment groups comparable. This stage in the RIE serves four key purposes. First, the open-minded and exploratory nature of the baseline phase allows us to identify novel hypotheses for how WASH programs may affect social cohesion—hypotheses that may have been overlooked during the initial inception stage. Second, the baseline phase ascertains that valid and reliable measures of key phenomena, particularly indicators of social cohesion, are developed and tested. This is achieved through the analysis of two extensive quantitative baseline surveys as well as qualitative interviews. Third, the baseline phase serves the purpose to assess and validate the hypotheses formulated during the initial phase, thus conducting a preliminary correlational 'test.' The final outcome of the report is a refined ToC that has not yet undergone testing but has been carefully adjusted and improved based on the findings distilled during baseline research (Figure 1).

The BR was prepared on the basis of careful qualitative as well as quantitative research on the ground in all three countries. Baseline research began with an extended desk review of pertinent program documents and preparatory analyses as well as interviews with key stakeholders. In a

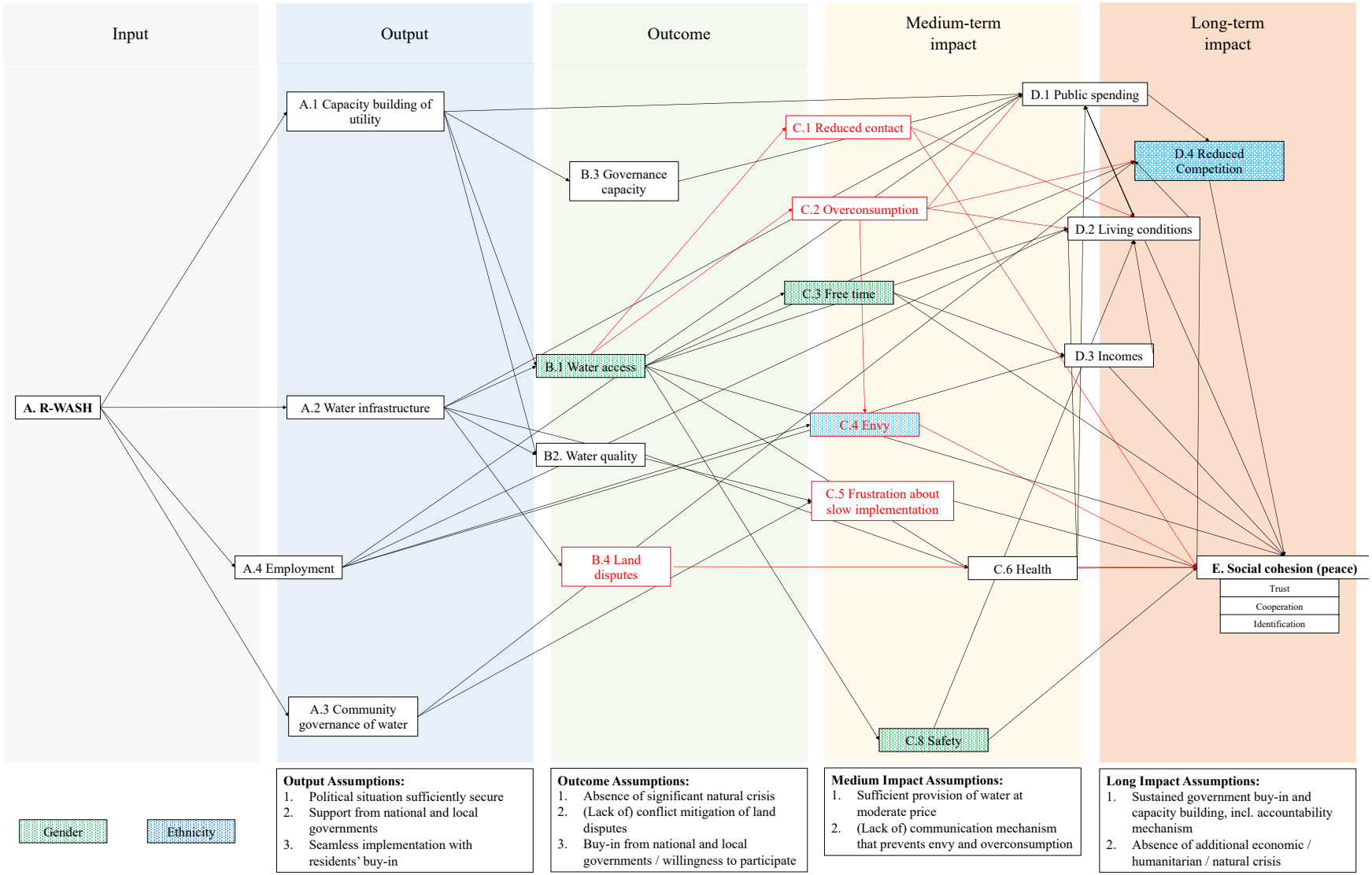
[1]NB: The 2023 conflict in Sudan has put R-WASH programming for the country site on hold and has also meant that no second baseline could be implemented in the country.

second step, the evaluation team visited all three countries and all sites (except Luuq, though the quantitative survey could be implemented there) where R-WASH is to be implemented as well as suitable control sites where R-WASH is not implemented. On the ground, the PIs used a multi-pronged qualitative research methodology, including interviews with UN staff, elders, beneficiaries, and other stakeholders as well as participant observation. In addition, the PIs implemented a large-scale population-level panel survey with the host and refugee/IDP population (total N = 11,500). The first Baseline wave was implemented in Q2/Q3 of 2022, the second Baseline in Q1 of 2023. The qualitative and quantitative data collected on the ground were then prepared and analyzed, which we outline in this report.

The main findings of the BR are as follows. First, our analysis of Baseline 1 and 2 survey data support a crucial assumption of the RIE: parallel trends in the outcome before R-WASH implementation has started. Put differently, the level and changes of social cohesion in treatment and control sites are found to be similar across the selected R-WASH and comparison sites. This finding allows the RIE to credibly attribute any potential changes in social cohesion *after* R-WASH implementation between the treatment and control sites to the program and not any other unmeasured factor. Second, our analysis showcases moderate degrees of pre-program levels of social cohesion. On a 10-point social cohesion scale (ranging from 0 = very low social cohesion to 10 = very high social cohesion), all countries hover close to the middle of the distribution at around 5 points. The moderate level suggests that we do not observe ceiling or flooring effects in our key outcome, lending credence to the measurement strategy as well as showcasing that R-WASH can meaningfully boost social cohesion. Put differently there is sufficient room to improve social cohesion through R-WASH. Third, the qualitative exploratory[2] research has allowed the research team to strengthen and update the potential causal pathways through which R-WASH will likely improve social cohesion (preliminary ToC). In particular, the BR emphasizes the following five causal pathways: i) improved safety, in particular for women; ii) more time resources for direct beneficiaries; iii) a potential for land disputes between residents (a potential detrimental pathway); iv) reduced envy from host communities toward IDPs/refugees, and v) improved governance capacity.

[2]We utilized exploratory research as an open-ended research methodology aimed at distilling key insights, generating hypotheses, and understanding the basic characteristics of the subject matter.

Figure 1: Updated Hypothesized ToC



The Figure shows the updated ToC. Arrows present causal pathways from an input (e.g., R-WASH) to an output (e.g, Employment). Assumptions capture key contextual factors that need to be ascertained in order for the respective outputs, outcomes and impacts to unfold.

2 EVALUATION CONTEXT

2.1 Problem Statement

The effectiveness of WASH programs is a critical area of study in humanitarian and development work, which warrants rigorous evaluation. A recent meta-analysis has confirmed that WASH programs improve direct outcomes pertaining to health and hygiene practices. In particular, Andres et al. (2018) find that "Water, sanitation, and hygiene interventions were found to increase the likelihood of behavior changes and the adoption of new hygiene practices by 17 percent" (2018, 1). However, studies assessing downstream societal outcomes, above all, social cohesion, remain relatively rare. While there are compelling arguments why and how WASH programs can, for instance, reduce competition for scarce resource like water, it becomes challenging to ascertain the extent to which WASH programs can effectively improve sustainable development at large without comprehensive, rigorous studies—particularly across contexts. A second gap in the literature concerns the effect of WASH programs in volatile, displaced persons-host population contexts (Husain et al., 2015). The lack of rigorous empirical evidence regarding the pathways through which WASH interventions improve outcomes such as social cohesion limits informed decision-making, hindering the ability to allocate resources efficiently and prioritize interventions that yield the greatest benefits for vulnerable displaced populations[3] while not disadvantaging host communities. Conducting rigorous evaluations to assess the effectiveness of WASH programs is thus imperative to inform evidence-based policies, improve program outcomes, and contribute to the global efforts towards achieving universal access to clean water, sanitation, and hygiene in a socially sustainable manner.

2.2 Context

Based on a comprehensive feasibility study, KfW and UNHCR / UNICEF have selected three sites for the first implementation period of R-WASH. These sites are Girba[4] in Sudan, Kebri Beyah in Ethiopia as well as Doolow in Somalia. The three sites and the accompanying control sites are provided in Table 1 below. Based on qualitative research and interviews, these sites were then matched to *highly* comparable control sites in the vicinity of the treatment sites. These are Luuq in Somalia, Kassala (Wed Sharife) in Sudan and Qoloji in Ethiopia. Based on our field work, we describe each context in turn and include preliminary evidence on likely pathways through which R-WASH may improve social cohesion and peace.

[3] Displaced populations refers to both refugees and IDPs.

[4] The precise implementation schedule in Sudan is still tentative. The first Endline survey will therefore need to be timed such that Girba can act as the Treatment site, while Kassala can function as the comparison site.

Table 1: Treatment and control sites

	Ethiopia	**Somalia**	**Sudan**
Treatment sites	Kebri Beyah	Doolow	Girba
Control sites	Qoloji	Luuq	Kassala (Wed Sharife)

Figure 2: R-WASH Implementation and Comparison Sites

Notes: The map shows the three R-WASH implementation countries—Sudan, Ethiopia and Somalia—as well as the R-WASH implementation sites (in orange) as well as the chosen comparison sites (blue).

2.2.1 Sudan

Following a period of relative peace , Sudan finds itself embroiled in a violent conflict for a decade. In 2019, an uprising resulted in the overthrow of the country's president Omar al-Bashir. Subsequently, a transitional government comprising both civilian, military and governmental representatives was established after extensive negotiations. On October 25, 2021, Army Chief Abdel Fattah al-Burhan orchestrated a military coup, leading to the ousting of civilian Prime Minister Abdalla Hamdok (Schellinger, 2022). The regime led by the army chief has since then responded to civilian protests with crackdowns and imprisonments. On April 15, 2023, an armed conflict broke out between the paramilitary Rapid Support Forces (RSF) and the Sudanese army, putting R-WASH and this evaluation on a temporary hold in the country.

The R-WASH sites in Sudan comprise two refugee camps as well as two relevant local host communities. The first is **Wed Sharife** with the host community of Kassala. Wed Sharife marks our control community though there will, later, be R-WASH programming in this site. The second is **Girba** with the neighboring host community of Khashm el Girba, which marks the treatment community. Both camps were established due to migration movements that resulted from civil conflict in the region. The Wed Sharife camp is situated about 10 km to the south of Kassala, the Girba camp 2 km from the small town Khashm el Girba. Wed Sharife is the bigger camp with 22,465 refugees (according to the prefeasibility study), while the Girba camp hosts 11,003 refugees. In both camps, the number of women exceeds the number of men and about 40 percent of camp residents are children. The average household size is 5 in both Wed Sharife and Girba.

The current conflict created significant instability and uncertainty about roles and responsibilities of government institutions in the region, and there have been isolated attacks on health facilities (Hashim, 2023). The current violent conflict increases the risk of attacks further. There are also reports on micro-level conflict lines between and across groups. Within the refugee camps, however, interviewees noted few significant conflict lines (as of Q2 2022). As one interviewee stated, "there are few conflicts, only 'normal' conflicts between persons." One reason for the absence of conflicts is the fact that most camp residents are from Eritrea who thus represent the majority group (Hovil and Jesperson, 2023).

Using the Sphere minimum standard as a benchmark, the evaluation team observes that the WASH situation across the two camps can be described as insufficient. In Girba, there are currently no boreholes. Residents use filtered and treated water pumped from the Altbarah river. At the time of the field visit (Q2 2022), a new pipeline network was being installed to connect the distribution taps in the refugee community. The WASH situation in the host community is comparatively worse. While local residents also receive water from the Altbarah river, the pumps are older and malfunctioning, and the filtering system is broken, therefore, supplying muddy water to the host community. Pipes are often blocked and people report dissatisfaction with this state of

water supply, partly because host residents are aware of the refugee population's comparatively better water supply system.

In Wed Sharife, water is available in sufficient quantity (24l/person, SPHERE standards). Water is supplied from three boreholes, managed by the Red Cross, serving 157 people per tap. The host communities are similarly served by boreholes or surface water by the state water cooperation. While there are conflicts in Wed Sharife as well, they are more common within the refugee community and not with the host community. In contrast to Girba, the Wed Sharife refugee camp is located far outside Kasalla, making direct comparisons (or potential envy) between host and refugee community with regard to their WASH status less common.

2.2.2 Ethiopia

Like Sudan, Ethiopia is facing significant political unrest and, in parts of the country, civil war. The civil war (Tigray War) that started in November 2022 between the central government under president Abiy Ahmed and the Tigray People's Liberation Front (TPLF) has led to severe casualties, leaving many hundreds of thousands of people displaced (Burki, 2022). The civil war was ended by the peace agreement of November 2022, though there have been reports of the ceasefire being broken (Crisis Group 2023). And, there are ongoing political conflicts, including between the government and the Oromo Liberation Front and, as of Q3 2023, between the government and the Amhara region.

The relevant host populations in Ethiopia are located in the cities of **Kebri Beyah** as well as the smaller community of **Quoloji** that forms the control site. The town of Kebri Beyah is situated about 60 km south-east of Jijiga and has 34,310 inhabitants with an average household size of six. Qoloji, a community which is part of the Kebele Ahnot, is about 80 km to the west of Jijiga and has about 500 households. Nearly the entire population of the two host communities are Somali.

Kebri Beyah, the treatment site, hosts 14,443 refugees. More than 55 percent of the camp population is under the age of 18, meaning that majority of residents were born in the camp. The average household size is approximately 5.5. The comparison site Qoloji is a camp divided in two sub-camps, Qoloji 1 and 2, and serves IDPs. The Qoloji camp is bigger than the Kebri Beyah camp, with a number of 79,148 of IDPs as of June 2022 and significantly more populous than the host community (which is estimated at 4.000; UNHABITAT 2021).

Across the two camps, Qoloji and Kebri Beyah, nearly all displaced persons are Somali. The majority in Kebri Beyah belong to the Darod clan. Kebri Beyah is the eldest of the three camps. Many camp residents left their home in the early 1990s, mainly as victims of draught. Qoloji camp, on the other hand, was created between 2016 and 2019, as a result of violence and the displacement of Somali Ethiopians living in other parts of Ethiopia (UNHABITAT, 2021).

Conflicts between the host and displaced persons population are, according to our document review and interviews, relatively infrequent, which is largely due to the common ethnic affiliation. Violent conflicts between clans are rare. There are occasional fightings between youth gangs. However, one woman described the relation with these words: "We are grateful that they welcomed us when we arrived. But there are differences between us, and the host community people. They treat us as people of a lower level." Scarce resources – including water – generate conflict within both the host and displaced persons communities, respectively.

The WASH situation across the two camps is currently insufficient. In Kebri Beyah, camp and host populations get their water from the same source: 7 boreholes. 2 of them are functional (as of Q2 2022). The water distribution systems in the camp and in the host community are served by separate pipelines. In both camp and host communities the current quantity of water delivered by the public water points is about 5-6 l per person per day, only about one third of the SPHERE minimum standard. Coping mechanisms include water tanks, water cisterns in the camp and wells near to the host community that are normally used for animals.

2.2.3 Somalia

Out of the three countries under study, Somalia has the longest history of violent conflict. After Siad Barre's rule from 1969 to 1991, the country descended into a brutal civil war between different clan-based warlords, which affects the country to this day. The combination of civil war and hunger resulted in the deaths of over 500,000 people and tens of thousands of people leaving the country (Nyadera, Ahmed and Agwanda, 2019). Since 2008, the African Union's Peace Support Mission for Somalia (AMISOM), currently reconfigured in the African Union Transition Mission in Somalia, has the mandate to protect the Transitional Government and the political process, to strengthen the Somalian security forces and to fight against violent radical Islamic groups, especially Al Shabaab.

The relevant host population in Somalia is located in the city of **Doolow** (treatment site), situated in the border region between Somalia, Ethiopia and Kenya as well as the city of **Luuq** (control site). Doolow has approximately 86,000 inhabitants. About 90 percent of residents are Somali of different clans. The living standard in Doolow is low (approximately 1.3 USD per day in our own representative survey evidence). Most families survive only because family members work abroad and send remittances. According to information provided by Somali Humanitarian Reflief ACTION (SHRA), there are few violent conflicts in town, despite differences in clan origin.

Luuq, similar to Qoloji in Ethiopia, hosts several IDP camps. Luuq lies in the Gedo province on the Juba River, which affords a moderate supply of water. Unfortunately, due to the adverse security situation, the PIs were not allowed to conduct on-site field research in Luuq, neither in the camp nor in the city. Luuq district hosts an estimated 8,233 IDP households, amounting to

51,028 individuals. Luuq city has an estimate population of 40,000, though precise figures are not available.

The R-WASH sites in Doolow comprise two IDP camps: **Kabasa** and **Quansaxley**. Both camps are near the center of Doolow city. The camp population in the two camps comprises 13,713 households and 81,258 individuals. The majority of IDPs have been in the camps for multiple years. 56 percent of the camp's population is younger than 18 years, 6 percent are elders of the age group above 60. Only 39 percent are in the working age of 18-59. In all age groups (with the exception of the elderly) the share of females slightly exceeds that of males.

The WASH situation across the two camps continues to be suboptimal. In Kabasa, there are 17 water points and two water tanks. In Quansaxley, there are 13 water points and three water tanks. The water points are not equally distributed (likely due to the time when the different sections of the camp were constructed). Some people must walk between 500 and 800 meters to reach the next water point (a potential source of envy), though most residents stay less than 300 meters from water points. The quantity of water is often insufficient, so that women and girls usually stand long-time in line to get water.

Across both camps, Kabasa and Quansaxley, water is provided by water committees, NGOs and Dolow Water Management Company, but mostly in insufficient quantity. Additional water is delivered by truck by UNICEF and other agencies, including local NGOs who are also involved in water management in the camp. A water committee supports the management of the scarce water resources. It is composed by representatives of the different sections of the camp (not elected, but selected by the camp leader) and the district commissioner. Representation of women in the committee is nearly 50 percent. Allegedly, according to SHRA, the agencies and the water committees distribute the scarce water equally between the camp residents.

2.3 Objectives of the Baseline Report

The main objective of the Baseline research and report is to gather practical knowledge and theoretical hypotheses in an exploratory fashion. The Baseline phase serves three core purposes. First, the open-minded and exploratory nature of the Baseline phase enabled the team to uncover additional hypotheses that may have been overlooked during the initial inception stage. Most notably, the ToC was updated significantly in line with potential pathways discovered to show how R-WASH may affect social cohesion (e.g., potential land disputes). Second, the Baseline stage aids the evaluation team in developing a comprehensive understanding of significant phenomena that make up the ToC, particularly indicators for social cohesion. Such exploration and corroboration of key indicators was accomplished through an exploratory study of the communities, but also an in-depth analysis of the panel baseline data. Third, the Baseline stage serves as a preliminary test

of the ToC. While R-WASH has not yet been implemented, the Baseline data can be used to explore whether phenomena map onto similar constructs as well as whether key causal pathways are positive predictors of social cohesion. Ultimately, the outcome of the Baseline phase is thus a more plausible and refined ToC that, importantly, has not yet undergone testing. Finally, the two baseline surveys were needed to validate a central identifying assumption for the quantitative research design: the so-called parallel trends assumption.

2.4 Purpose of the Evaluation

The purpose of the R-WASH rigorous impact evaluation is four-fold. First, the evaluation should help donors, program managers to assess the effectiveness of similar WASH interventions on social cohesion and thus make informed decisions for program improvement. It is crucial to underline that this effect—WASH on social cohesion—is a rather broad question of interest. The main aim of WASH programs is to reduce water related diseases and provide equitable service provision. In line with the BMZ's agenda, however, the evaluation scrutinizes broader developmental effects of WASH on social cohesion. Second, the RIE may be used for accountability purposes, ensuring transparency to donors, and demonstrating accountability to stakeholders, including governments, communities, and direct beneficiaries. Third, the evaluation report will hopefully aid researchers and academics as a valuable source of data and theoretical guidance for future studies. Fourth, the evaluation report can play a crucial role in influencing public opinion and raising awareness about the importance of WASH programs, generating support and advocacy for future initiatives in marginalized communities across the globe.

2.5 Stakeholders

The R-WASH evaluation (and R-WASH itself) involves a wide range of stakeholders at various levels, including international, national, subnational and local actors. Internationally, R-WASH involves the donor organizations (KfW as well as BMZ) and multilateral institutions (the participating UN agencies). Nationally, key stakeholders include the national governments in Sudan, Ethiopia and Somalia as well as government agencies responsible for development, including the ministries of finance, planning, and relevant sector-specific ministries. Regionally, stakeholders include intergovernmental organizations, regional development banks, and regional NGOs. Local government authorities and municipalities also play a crucial role in R-WASH's implementation, coordination, and the evaluation more broadly. Civil society organizations, community-based organizations, and NGOs furthermore contribute their expertise and represent the interests of local communities. Additionally, beneficiaries and community members are essential stakeholders, as their needs

and perspectives shape the program's design and outcomes. An overview of key stake holders is provided in the Matrix in Section A.5.

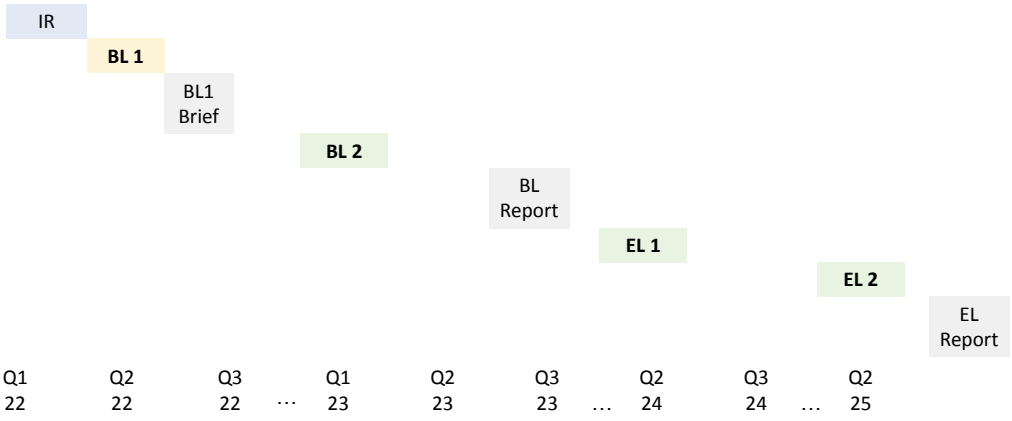
2.6 Timeline

Relative to the Terms of Reference (ToR), the timeline of the evaluation has been adapted significantly. The timeline as of August 2023 is provided in Figure 3. The first steps in the evaluation refer to the drafting of an inception report and a design stage in Q1 - Q2 2022. Upon mutually agreeing on a research design, the first Baseline phase was implemented across the three R-WASH programming countries in Q2 2022. Field work took place according to plan and led to the submission of a Baseline Policy Brief in Q3 2022. The second baseline, which we will introduce below, took place in Q1 2023. Given the turmoil in Sudan, the second (quantitative) Baseline was only implemented in Somalia and Ethiopia. The end result of both Baselines is the present Baseline report, which was first circulated in May 2023. Given that the evaluation no longer applies a cross-over design (that is, the control communities will *not* receive R-WASH programming[5]), it was agreed that the Midline is turned into a first Endline and the Endline will be turned into a second Endline. The advantage of this approach is that it allows the team to assess short- and long-term effects of R-WASH on social cohesion. The projected dates for the first and second Endline are Q2 in 2024 and Q2 in 2025, respectively.[6]

[5]There is an opportunity to use the Sudan case as a cross-over design, depending on when programming finalizes in Kassala.

[6]The projected dates may change if further unforeseen dynamics in the intervention areas occur that delay implementation.

Figure 3: Intended timeline



3 OBJECT OF EVALUATION

3.1 Scope of the Evaluation

The **scope** of the evaluation is to examine the extent and manner in which R-WASH influences social cohesion. The **programmatic** scope of the RIE encompasses all R-WASH activities starting from Q2 2023 (after the completion of the second baseline), which can shed light on how R-WASH affects social cohesion. The **geographic** scope of the RIE includes the three project and comparison sites in the Horn of Africa: Sudan, Ethiopia, and Somalia.[7] The **unit of analysis** for the RIE consists of individual camp residents and the host population. The **thematic** scope of the RIE focuses exclusively on the impact of R-WASH on social cohesion, with no analysis of other topics or evaluation criteria.

The evaluation's focus on the impact of R-WASH on social cohesion has two key implications. First, the evaluation methodology, which combines quantitative DID and qualitative methods, will solely consider outputs that contribute to explaining the connection between R-WASH and social cohesion, as outlined in Figure 1. Aspects that are highly improbable to impact social cohesion will not be examined. Second, the evaluation primarily aims to provide an abstract blueprint ToC. Consequently, it must *abstract* from local contexts. For instance, if a causal link between R-WASH and social cohesion is weak and only applicable in a specific setting, it may not be included in the overall ToC blueprint. We do, however, provide specific contextual examples below.

[7] As Sudan is undergoing a major conflict since 15 April 2023 the second Baseline could not be implemented in the country.

4 EVALUATION APPROACH AND METHODOLOGY

4.1 Evaluation Framework

The evaluation uses a DID quasi-experimental design and deploys mixed-methods where qualitative and quantitative data will be fused to derive an overarching, generalizable ToC. Before delving into the qualitative and quantitative components, we want to briefly lay out the overarching framework, which proceeds as follows. In a first step, the Evaluation team laid down its prior beliefs (i.e., experts' pre-program expectations) about how WASH programs affect social cohesion. This was done on the basis of expert interviews, desk research and a study of the academic literature. The resulting preliminary ToC was presented in the Inception Report. In a second step, we conducted qualitative and quantitative Baseline research in order to explore and distill new hypotheses that update the prior thinking laid out in the IR. This stage is described in the present report. The Baseline research included qualitative interviews and two panel surveys. The third stage will be the first Endline phase. The purpose of the first Endline phase is to put the ToC to a rigorous test and to corroborate key causal pathways using qualitative comparative process tracing. The second Endline phase extended the research of the first Endline phase to include longer-term effects, a year after the first Endline phase. All collected data will aid in the process of crystalizing a unified ToC that is, hopefully, transportable to other contexts. Depending on how the situation in Sudan evolved, the country may need to be dropped from the analysis (in case of further escalation and no R-WASH implementation) or can be included without problems (in case of a timely implementation of R-WASH).

4.2 Design and Methods

We begin by introducing the qualitative items and sampling strategy, which was used to derive new hypotheses and understand the complex cultural context in which R-WASH takes place. We then lay out the quantitative survey instruments and sampling strategy, which was used to provide robust pre-R-WASH measures for the causal pathways that most likely link R-WASH to social cohesion.

4.2.1 Instruments of the qualitative surveys

To develop a detailed understanding of the social, political, and cultural context, the R-WASH program across the sites and likely mechanisms that link R-WASH to social cohesion, we implemented a multi-pronged qualitative research strategy. While some parts of the methodology (e.g., participant observation, desk research, elaborated upon below) required no structured questionnaire,

we asked all researchers to orient the qualitative interviews along the lines of a semi-structured questionnaire. The full list of questions is provided in Section A.2.2. A first set of questions focused on describing the social context of the refugee / IDP camps (e.g., "Which ethnic groups are represented?" and "How much turnover is there?"). A battery of questions dug more deeply into the structure of the camps as well as salient conflict lines (e.g., "Are there groupings or social hierarchies between camp residents?" or "Do camp residents have the right to work? If yes, where do they work?"). A third set of questions focused on describing the social context of the host community (e.g., "Which ethnic groups are represented?" or "What livelihoods / professions do people have?"). A fourth set of questions deals with the interaction between camp residents and host communities (e.g., "Are there trade relations between camp residents and host community?" or "Do workers of the host community and camp workers compete for jobs?"). A fifth set of questions dealt with potential conflicts between camp residents and host communities (e.g., "Which are the main conflict issues?" or "To what extent does water play a role?"). A sixth set of questions dealt with the supply of water and sanitation (e.g., "How many persons does one water tap serve on average?" or "What are the main problems with the current water supply?"). A seventh set of questions focused on potential positive effects of R-WASH on social cohesion (e.g., "What do the refugees/IDPs, and the host community expect from the implementation of the R-WASH project?" or "Would water points closer to homes reduce the workload of women and girls?"). An eighth set of questions dealt with a potential negative effect of R-WASH on social cohesion (e.g., "Could higher tariffs (due to improved water quality) lead to conflicts?"

4.2.2 Sampling of the qualitative surveys

Achieving representativeness is a crucial objective when sampling for qualitative interviews. However, in this scenario, it becomes more challenging to define the appropriate "population." To approximate a representative sample, we sought guidance from key stakeholders, including local interlocutors, UN staff, and experts to identify individuals who could provide us with important clues and corroboration of the ToC as well as the specific cultural context (see Section A.1 for details). Building upon this framework, we then established a sampling frame and made contact with selected individuals. As mentioned earlier, this process did not involve simply relying on snowballing through expert networks. On the contrary, to ensure the representativeness of the expert sample it was crucial to obtain valid input from a *diverse* range of experts, rather than biased evidence from influential figures who may be more inclined to engage in discussions (such as local tribal elders). This meant, for instance, that both representatives of the refugee and IDP as well as of the host population were interviewed, thus ascertaining a representative sample along age, gender and ethnicity. Besides expert and key informant interviews, the PIs also conducted eleven Focus Group Discussions, three of which were exclusively held with women.

4.2.3 Instruments of the quantitative surveys

To rigorously measure key indicators for the ToC we implemented two panel surveys across all three countries (Sudan was only interviewed once). The quantitative large-N surveys relied on a structured questionnaire, which is provided in Section A.2.1. The primary objective of the survey was to assess if and how R-WASH improves social cohesion. We measured social cohesion on the basis of a definition provided in Chan, To and Chan (2006), which is as follows: "*Social cohesion is a state of affairs concerning both the vertical and the horizontal interactions among members of society as characterized by a set of attitudes and norms that includes trust, a sense of belonging and the willingness to participate and help, as well as their behavioral manifestations* (Chan, To and Chan 2006, 290). We condensed this definition to its quintessential elements and defined social cohesion as '**levels of trust, cooperation and identification in a community**.' Importantly, all measures are taken at the individual level. We then quasi-experimentally measured the three elements using i) the trust game (each respondent was given the opportunity to send any amount of imaginary 10 Dollars to the next respondent, thus "trusting the person"), ii) a measure for co-operation (again, each respondent was told to imagine that another respondent had invested in the trust game and had sent them money and we then measured whether the respondent would send any money back), iii) a measure for identification (each respondent was provided with an imaginary 10 USD and asked if they would like to share any of this money with the respective community). In addition, we measured the three constituent elements of social cohesion using the following three items. First, we asked subjects "*Generally speaking, to what extent do you agree with the following sentence: most people [in my ingroup / in the outgroup] can be trusted*". Second, we asked subjects "*Imagine you lost a wallet that contained 20 dollars and someone from [the ingroup / outgroup] found it. How likely is it that the money is returned to you?* Third, we asked subjects "*We have spoken to many people in this area and they have all described themselves in different ways. Some people describe themselves in terms of their religion, ethnic group, language-group or nationality. Others describe themselves in economic terms, such as farmer. Which specific group do you feel you belong to first and foremost?*" The answer choices included the relevant social reference category.

Besides these measures for social cohesion, we also created measures for each constituent part of the ToC, which is given in Figure 1. These measures, for the sake of brevity were single items per "concept in a box". The full list of items is provided in Section A.2.1. Notably, in the first step, we measured the immediate **outputs** of R-WASH, adding four items including, e.g., *'On a scale from 0 (not at all) to 10 (very much so), to what extent are you aware that the local water provider has received support from outside organizations in the last two years?'*. In a second step, we measured the **outcomes** of R-WASH, adding items such as 'On a scale from 0 (not at all) to 10 (very much so), how easy is it to fetch water in this neighborhood?'. In a third step, we measured

the **medium-term impacts** such as reduced contact (C.1; "In the last month, how often did you have a chat with a stranger from the refugee / IDP community in a given week?".). In a fourth step, we measured **long-term impacts** including, e.g., public spending (D.1) using an item "On a scale from 0 (not at all) to 10 (very much so), how much have community finances improved over the last two years?".

4.2.4 Sampling of the quantitative surveys

To obtain a representative sample of the local host and refugee/IDP population, we proceeded in four steps. First, we acquired satellite imagery of the relevant communities. Second, we superimposed a grid pattern on to the inhabited sections of each community. For instance, in Kassala, the community was overlaid with approximately 1,000 grids (see Figures 3-8). Third, we placed random points within these grids in order to achieve a random sample so as to obtain a representative sample of the population. Enumerators were then randomly assigned to grids and instructed to approach households nearest to the placed dot on the map. The overall sample sizes are provided in Table 2. As can be seen, we conducted 11,500 interviews in a two-wave panel.

Table 2: Sample sizes

Variable	**R-WASH**	**Control**	**Refugee/IDP**	**Host**	**Ethiopia**	**Sudan**	**Somalia**
N	5523	5977	5072	6428	4526	2302	4672

Figure 4: Sampling strategy in Somalia (Dolow; Host)

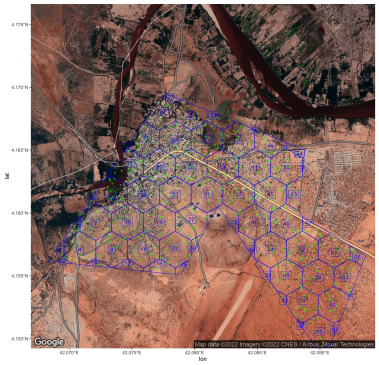


Figure 5: Sampling strategy in Somalia (Dolow; IDP)

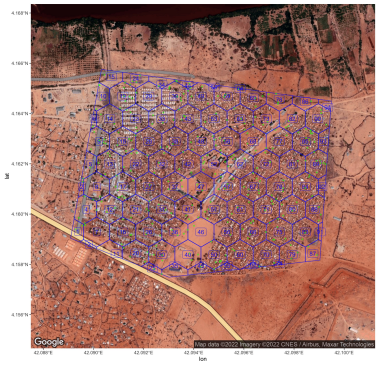


Figure 6: Sampling strategy in Ethiopia (Kebri Beyah; Host)

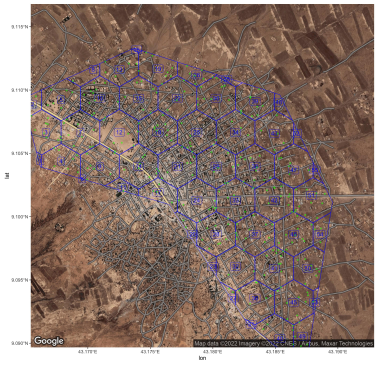


Figure 7: Sampling strategy in Ethiopia (Kebri Beyah; Refugee)

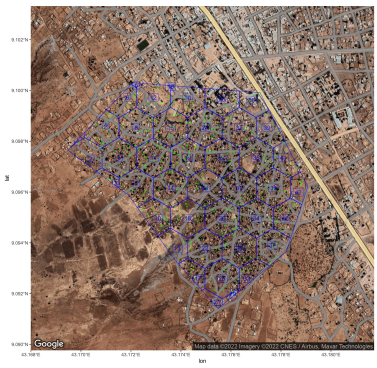


Figure 8: Sampling strategy in Sudan (Girba; Host)

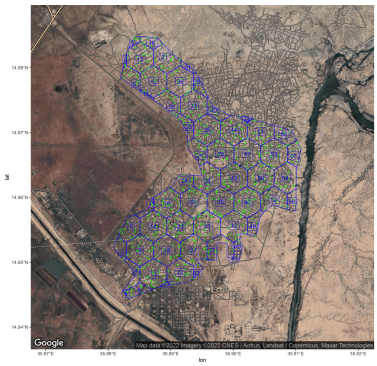
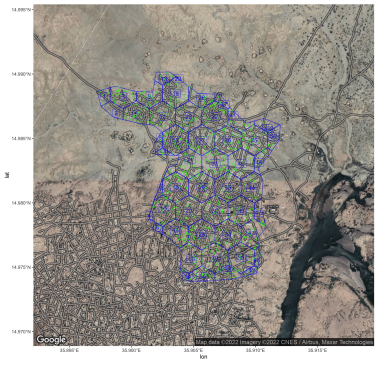


Figure 9: Sampling strategy in Sudan (Girba; Refugee)



4.3 Data Analysis

4.3.1 Qualitative data

The qualitative data were chiefly used to explore and craft new hypotheses for how R-WASH may affect social cohesion as well as to arrive at a deeper understanding of the context. To do so, we visited all three countries and collected data in three key ways. First, we interviewed local experts, including UNHCR / UNICEF staff, to explore–in an open manner–how R-WASH could affect social cohesion (see i.a. the stakeholder table). The interviews were informed by the guiding questions provided in examples, but were largely exploratory. Second, we interviewed local residents who will benefit from R-WASH. This included both displaced persons and residents in the host community. Here, too, the goal was to gain a deeper understanding of current conflict lines to eventually assess how these may be improved upon by R-WASH. In addition, the goal was to uncover alternative mechanisms, which operate independently of current conflict lines. Finally, the qualitative interviews sought to describe relevant phenomena (Mahoney, 2010) and thus test whether our constructs are measured appropriately.

The qualitative data was then analyzed using methods of iterative and recursive qualitative analysis (Srivastava and Hopwood, 2009). In a first step, we organized the qualitative data to make sure that all relevant information was captured. Second, we used the ToC as a background and then probed specific pathways or added pathways between interventions and result, where necessary. This process was done both during and after completion of the field work. (For instance, in Ethiopia, we implemented a workshop inviting key stakeholders and then re-produced the ToC step by step, thus testing whether the Theory from the Inception Report was plausible.) Third, we used a process of open coding to identify and categorize recurring themes and causal pathways as well as concepts and patterns in the data. Fourth, once the ToC was updated (both on the basis of the qualitative and the quantitative data), we then re-analyzed the data to look for evidence that supported, challenged, or even contradicted the overall causal narrative. The overall goal of the analysis, thus, was to arrive at a good balance between exploratory and structured qualitative analysis, to identify explanatory patterns that could complement existing in the ToC, develop hypotheses on causal logic, and then (preliminary) test them.

4.3.2 Quantitative data

We used the quantitative baseline data to measure key constructs of interest. To do so, we implemented two large-scale, representative population-level surveys of residents and displaced persons across all sites. The overall goal of the baseline surveys was to assess whether key phenomena are present and also to determine whether they can be reliably and validly measured. The

reliability aspect was determined on the basis of the second baseline. In particular, we compared the outcomes across the two time points. The second purpose of the baseline survey was to determine parallel trends across the treatment and control communities in order to assess comparability across the sites. We simply plot means of key outcome variables across the two phases to assure there were no differential trends across treatment and control groups.

Once the endline data is available, our estimation strategy—depending on whether Sudan will enter as a cross-over or a traditional DID—will rely on the following generalized difference-in-differences specification:

$$Y_{ijt} = \mu_i + \delta_t + \sum_{k=-10}^{10} \beta_k \left(T_j \times \mathbf{1}_{t=k} \right) + \varepsilon_{ijt}$$

Here, Y_{ijt} is the outcome of interest (e.g., trust as a core measure of social cohesion) in camp i nested in country j at time t. The terms μ_i and δ_t denote camp and time fixed effects. The main parameters of interest are a series of coefficients on the leads and lags of the treatment, denoted by β_k. This parameter gives us the difference between treated and control camps for three periods before and after the treatment is implemented. As is standard practice in leads and lags regression analyses, we leave out one interaction for the last pre-treatment period, which serve as the baseline for all estimated treatment effects.

4.4 Limitations and Ethics

4.4.1 Limitations

There are three potential limitations associated with the study overall and the Baseline report, in particular, which we want to acknowledge. These limitations can also be conceptualized as **meta risks** for the ToC such core causal pathways—and possibly the entire project implementation—that may be put in jeopardy if these identified risks materialize.

First and most pressing, the **Sudanese conflict** led to the pausing of R-WASH implementation in the country. This, in turn, presents a challenge for the ongoing evaluation. If R-WASH will nonetheless resume implementation rather soon in Sudan, the overall evaluation can proceed as planned (albeit with a lack of data on parallel trends). However, if R-WASH is delayed significantly (ca. 1 year), Sudan essentially moves into the control group or, in statistical language, becomes "non-compliant". At worst, it could mean that no data can be collected. At best, we could use Sudan as a control site (assuming parallel trends), though comparability would not be ideal. A related problem concerning the conflict is that we were unable to implement a second Baseline in

the country. Fortunately, we were able to implement a first Baseline and qualitative work in Sudan. However, the lack of a second Baseline is a problem in as much as we are unable to determine parallel trends in the country. While we will be able to estimate a panel model in Sudan, we will not be able to rule out that the treatment and control sites were on different *trends* prior to R-WASH. The lack of data on parallel trends for Sudan also further hampers the aforementioned statistical precision of the evaluation.

Second, at its core, the evaluation involves a comparison between six sites, specifically three treatment sites and three control sites (though there are subsites).[8] The relatively small number of clusters (sites) poses a concern in terms of **statistical power**. To put it simply, there is a legitimate threat that the design lacks the necessary statistical precision to estimate clear effect sizes. To mitigate this concern, we have taken two steps. Firstly, the evaluation seamlessly integrates qualitative evidence into the statistical models to better triangulate the evidence. Secondly, we have carefully chosen control sites that exhibit high comparability (further details are provided below).

Third, R-WASH is a **highly complex program**, which is unfolding in volatile contexts with a multitude of involved partners. This presents a challenge as the final programming (and the ToC derived) is a 'moving target.' This challenge is less of a problem for the *endline* stage at which point R-WASH will have been implemented and can be readily observed. However, at the Baseline stage there remained significant uncertainty about which parts of the program will be implemented exactly when, where, and how. While there were broad trends, the precise location, for example, of a well or the exact pricing of water in the utilities to be created remained somewhat elusive. Moreover, specific implementation characteristics of R-WASH—e.g., the inclusion of the communities in planning, the transparent distribution of paid employment—can influence how R-WASH affects social cohesion.

4.4.2 Ethics

To ensure adherence to rigorous ethical standards of the evaluation, we secured IRB approval from the Faculty of the Human and Social Sciences at Humboldt University. Our application outlined the research design, provided a preliminary version of the questionnaire, and delved into the ethical considerations surrounding interviews conducted with refugees, IDPs and host populations in conflict-afflicted and underdeveloped regions of the world. The main ethical concern revolves around the potential risks posed to the participants. Given the vulnerable nature of the population, we have made concerted efforts to minimize any risks that may arise from their involvement in the survey.

[8] Note also that it is currently unclear whether Sudan will be part of the Endline survey.

The following six points are of particular significance. First, participation in the survey and interviews was completely voluntary, and this aspect was explicitly communicated to all individuals. Second, no personal information, except for a means of re-contacting individuals, was collected or stored. Third, the questionnaire did not include sensitive or highly political questions. Fourth, we implemented appropriate safety precautions. Fifth, we implemented a rigorous system of communication with all involved stakeholders, notably UNICEF and UNHCR's local staff as well as KfW. Sixth, all data were stored in a doubly-encrypted manner.

4.5 Quality Assurance

In order to ensure high-quality qualitative and quantitative data, we followed seven steps. First, we defined clear objectives for all involved partners, enumerators, and respondents regarding the purpose and goal of the surveys/interviews. Doing so ensured that all involved parties clearly know the point of the exercise. Second, we defined a well-structured and clear survey instrument—both for the qualitative interviews as well as for the quantitative survey. A clean, well-structured survey significantly reduces measurement error and associated biases such as fatigue or non-comprehension. Third, we piloted the survey with a diverse set of respondents. Doing so made sure that any remaining ambiguities or inconsistencies could be removed. Fourth, we relied on a rigorous representative sampling scheme in order to ensure that the data speak to a clear population of interest (more below). Fifth, all involved enumerators, including the PIs, have received significant training in order to ensure a smooth and reliable implementation of the survey. Sixth, we monitored the data collection (particularly the quantitative part) with a survey supervisor and with local heads of survey groups. They were charged with identifying any errors, inconsistencies, or other issues. Last, we conducted extensive data validation checks. The quantitative data were visualized and tested for normality (for pertinent outcomes in *R*), we checked for any patterns (increased variance across outcomes) and also made sure all variables scale sensibly.[9]

[9] Scaling was assessed using Cronbach's Alpha.

5 FINDINGS

The findings from the qualitative and quantitative data analyses can be grouped into five core insights, which we present below. They map onto the goal of the overarching evaluation—namely, the ToC for how R-WASH affects Social Cohesion as well as the comparability of the treatment and control sites ("parallel trends").

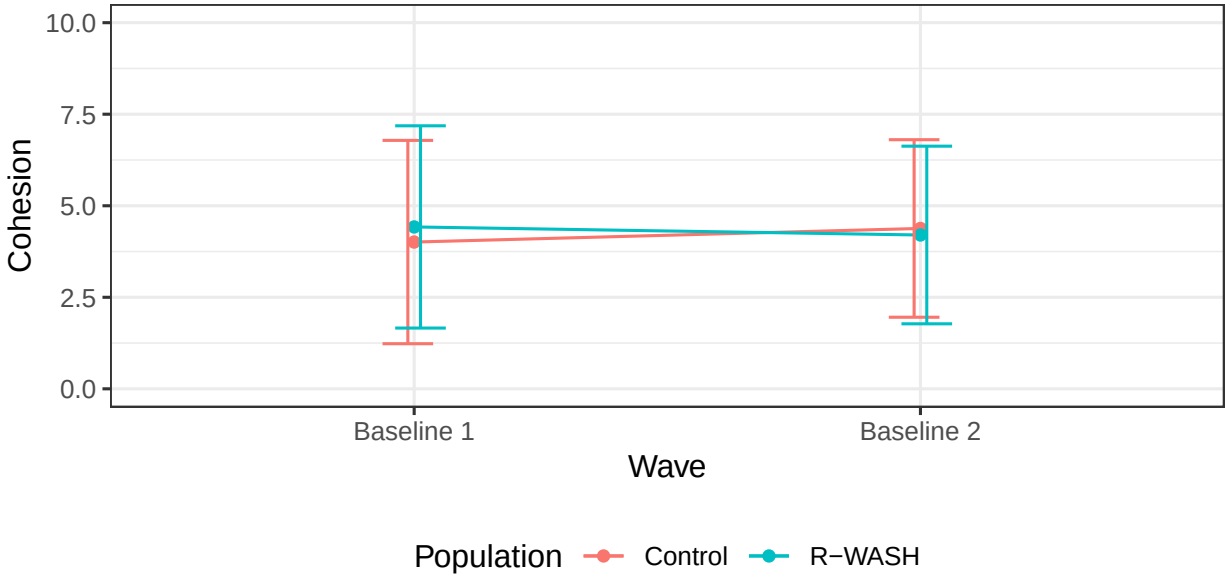
5.1 Parallel tTrends

A first key insight of the data analysis concerns the comparability of the treatment and control sites before the implementation of R-WASH. This so-called parallel trends assumption is a critical identifying assumption of the DID model. Recall that the treatment sites, Kebri Beyah, Doolow and Girba, were chosen by KfW in unison with UNHCR and UNICEF on the basis of comprehensive feasibility studies. The control sites were then chosen by the research team to present suitable counterfactuals or comparison groups. As such, they should be comparable because the field research showed them to be similar based on several relevant characteristics (chiefly the incidence of a refugee/IDP camp, the size of community, the ethnic composition of the camp) and differ less in covariates due to their geographical proximity. Even so, the communities may still be different for reasons that we may have been unable to uncover during the qualitative field work. For this reason, the evaluation relies on a rigorous quantitative assessment whether the levels, but even more important the *trends* of key outcome variables, above all social cohesion, are similar across the treatment and control communities before the R-WASH program is implemented.

To this end, we estimated statistical models that compare the most important outcomes, i.e., social cohesion, across the treatment and control sites. Note that, since the second Baseline could not be implemented in Sudan, the analysis focuses on Ethiopia and Somalia. The key result of this analysis is shown in Figure 10. The Figure convincingly shows that the treatment and control groups are highly comparable both within space but also, most importantly, across time. In particular, the Figure shows the average cohesion-level (here, the amount sent in the dictator game between 0 and 10 USD) across the treatment and comparison group in Baseline 1 and 2. Two results stand out. First, both the control and R-WASH communities have highly similar levels of cohesion, which hovers around 4.5 USD. The similarity across the two groups is consistent for both Baseline 1 and Baseline 2. Second and most relevant, there is no meaningful *trend* (or change) in cohesion from Baseline 1 to 2—neither in the treatment, nor in the control group. As such, the treatment and comparison sites are highly comparable, which is a key assumption of the evaluation.

Equally of interest, we also analyzed whether there are differences across the host and refugee/IDP population. The key findings are shown in Figure 11. The Figure plots the means of pro-sociality

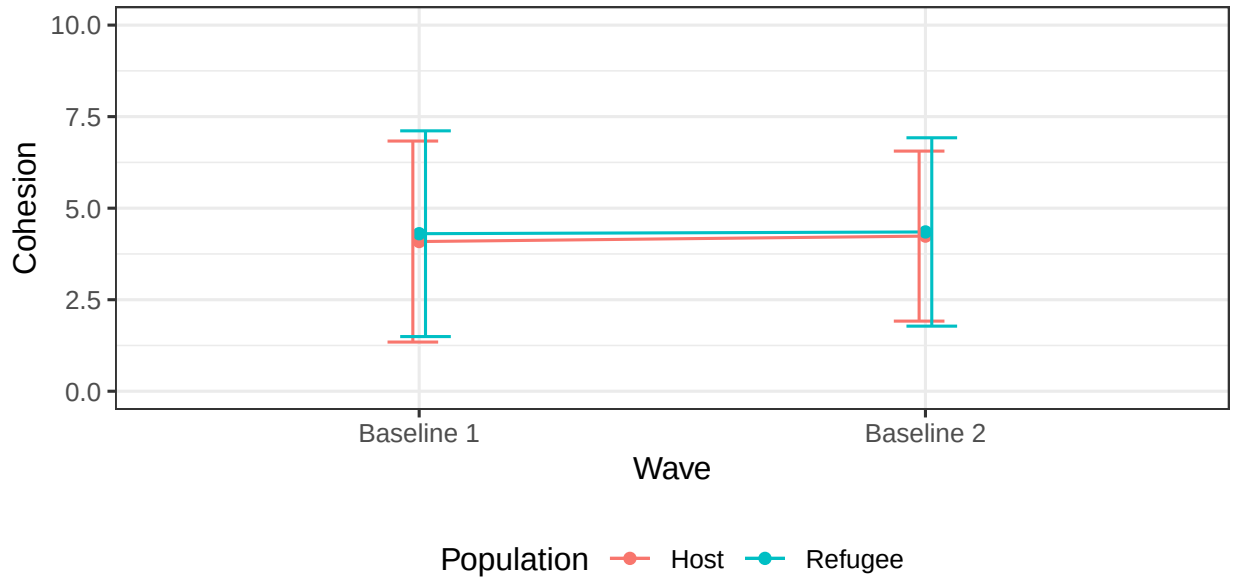
Figure 10: Social cohesion (index) across treatment and control group (Baseline 1 and 2)



Notes: Figure 11 reports mean values of one core pro-sociality outcome across two countries: Somalia and Ethiopia. (Note that the second Baseline could not be implemented in Sudan due to the current conflict, which is why we focus the parallel-trends analysis on the two countries.)

across the first and second Baseline in the host population as well as the refugee/IDP population. As can be readily seen, there are highly similar *trends* across the two waves. Both populations practically stay at 4 USD from Baseline 1 (Summer 2022) and Baseline 2 (February 2023). Besides no change in the *trends*, there is also no noticeable difference in the *levels* of social cohesion between the host and refugee/IDP population. A third noticeable finding is the rather high variance (Error Bars). The variance showcases that trust is highly variable across both groups suggesting that pro-social behavior differs markedly across respondents. Overall, the two Baselines have thus provided solid support for the assumption of parallel trends. The two groups, in other words, are highly comparable.

Figure 11: Social cohesion across waves (Baseline 1 and 2)

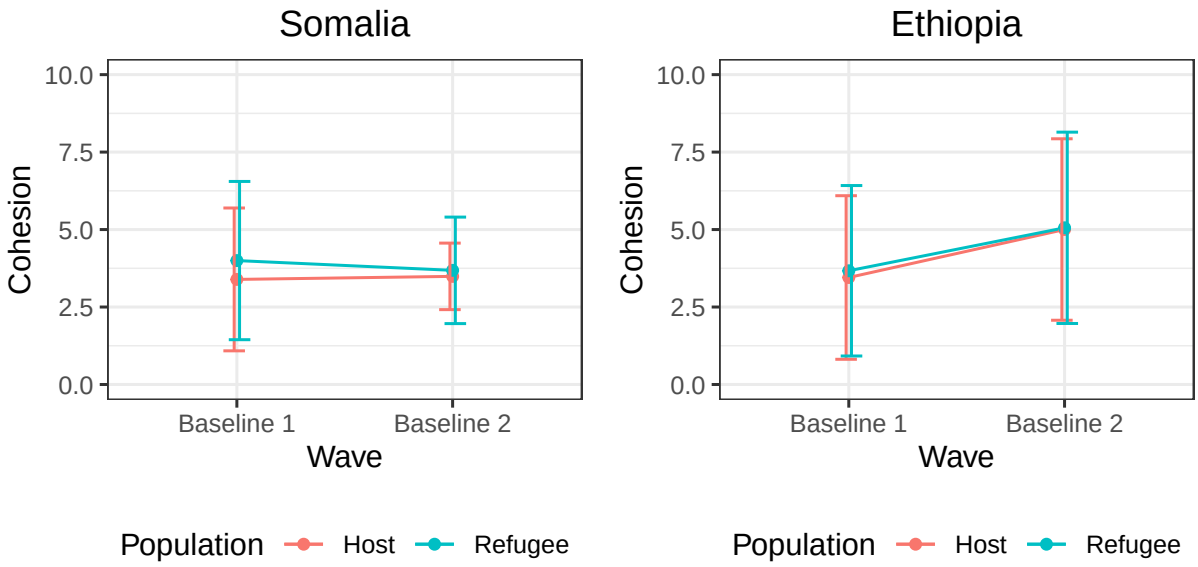


Notes: Figure 11 reports mean values of one core pro-sociality outcome across two countries: Somalia and Ethiopia. (Note that the second Baseline could not be implemented in Sudan due to the current conflict, which is why we focus the analysis on the two countries.)

A similar picture emerges when plotting the parallel trends within the two countries, Ethiopia and Somalia. Figure 12 shows the parallel trends across Ethiopia (right-hand side) and Somalia (left-hand side). In Somalia, the picture is similar to the aggregate sample: There is no change in levels, nor in trends, of social cohesion from Baseline 1 to Baseline 2. In Ethiopia, the situation is a bit more nuanced. Here, we do observe a difference in *levels* from Baseline 1 to Baseline 2. In particular, social cohesion—both in the refugee/IDP and in the host community—improved from Baseline 1 to Baseline 2. This change might, for instance, be a product of moderate improvements in the ongoing conflict in the country since Baseline 1 was fielded, though we cannot be certain about

this attribution.[10] Importantly, however, the *trends* are highly isomorphic across the two groups in Ethiopia, too. In particular, *both* the host population as well as the refugee/IDP population develop highly similarly regarding their levels of pro-sociality. As such, the Figure neatly illustrates that are we to detect any changes across the two groups *after* R-WASH is implemented, it is likely due to the program—not unobserved factors, which, seemingly, affect the two groups in a similar manner.

Figure 12: Social cohesion across waves in two countries (Baseline 1 and 2)



Notes: Figure 12 reports mean values of one core pro-sociality outcome. (Note that the second Baseline could not be implemented in Sudan due to the current conflict, which is why we focus the analysis on the two countries.)

The fact that we showcased stable parallel trends presents two methodological insights to the evaluation. First, the uncovered parallel trends arguably imply that the qualitative matching across the treatment and control sites as well as the communities worked well. In other words, the field work did identify the right comparison sites, which have similar *levels* and *trends* of the key outcome R-WASH is intended to positively impact: social cohesion. This is reassuring in as much as it allows us to estimate the causal effect of R-WASH on key outcomes with relatively high confidence once the Endline data have been collected—whether qualitative or quantitative. Second, the comparability also underlines that the qualitative field work did allow the research team to gain a deep insight into the communities and thus determine whether sites are comparable or not. This builds credence in the qualitative findings that support the updated ToC.

[10] We have no qualitative data to back this up, given that qualitative interviews were only conducted during Baseline 1.

5.2 Levels of Social Cohesion

A second key insight of the analysis concerns the level of social cohesion. The level of social cohesion is of high relevance not only for policy reasons, but also because it is a core goal and hypothesis of the R-WASH evaluation to test whether the program "improve[s] social cohesion within and across groups." Recall that the evaluation has chosen to emphasize this primary hypothesis for two specific reasons. First, we have singled out only one hypothesis because it serves as the central objective of the overarching program's strategy: To determine the impact of R-WASH on social cohesion. Any additional hypotheses formulated were intended to explore the mechanisms through which R-WASH may influence social cohesion, if at all. Second, by concentrating on a single hypothesis, we aim to avoid the pitfalls of multiple comparisons. This is especially crucial considering the relatively limited statistical power of the overall evaluation.

To determine the level of social cohesion, we rely on both quantitative and qualitative data. Quantitatively, Figure 12 implies two key findings. The first, similar to what was outlined in the Baseline Policy Brief, is that social cohesion is at *moderate levels* in the communities of study. We arrive at this conclusion by benchmarking the sent amount in the game to a recent meta analysis by Engel (2011). The authors show that, on average, the amount invested is similar, if a bit lower, compared to the setting we study. In particular, the authors write "dictators [subjects who can distribute money] on average give 28.35% of the pie". In our setting, this Figure is 32% (6.6 USD out of 20). The comparability, of course, is not ideal as we study marginalized (refugee/IDP) communities in the Global South, while most meta analyses study Western samples. Still, the benchmarking allows us to rule out that social cohesion is particularly low and the mean figure is surprisingly close to the global mean.

The moderate level of social cohesion *within* and *across* groups was also a recurring theme in the qualitative research. Contrary to our expectations, conflicts between refugees/IDPs and host communities were not conveyed to us to be very acute.[11] The notable absence of conflict seemed, partly, due to the fact that in all the cases studied the population composition of the refugee/IDP camps and the host communities was largely homogeneous in terms of ethnicity and religion, i.e. there were hardly any cultural fractures. Celebrating religious events together and occasional intermarriages were further drivers of rather reliable social cohesion. That said, resource conflicts and envy do exist between refugees/IDPs and host communities, especially over scarce firewood (in Ethiopia) and, relevantly, water (in Somalia). But, in the context of Ethiopia and Dolow, for example, conflicts over water are mitigated by the fact that the water taps are separate, so that the two groups do not meet while queuing.[12] We did encounter examples of "envy" (see ToC),

[11] However, we should caution that the extent to which people report latent or actual conflicts may be related to their expectations of participating in interviews like ours.

[12] This contextual difference underlines the relevance to conduct the analysis both for all countries combined as well as for all countries individually.

especially in cases when one of the groups (mostly the refugees/IDPs) is supplied with more water, in better quality, and at better prices.

The fact that social cohesion is, perhaps, more pronounced than anticipated has two important implications. First, it raises the bar for R-WASH to have positive effects on this outcome. While we are far from a ceiling effect (where no improvement is possible), the situation may be more peaceful than anticipated, which makes it more difficult to trace positive effects. That said, the amount remains moderate (average of 4.5 USD in the game, for instance), so movement should be possible. Second, we also analyzed whether social cohesion is driven by pertinent social characteristics. The analysis, which relies on an ordinary least squares regression, is given in Table 3. Interestingly, we find that refugees/IDPs, on average, are significantly less likely to act pro-socially. This finding could be the product of limited resources and opportunities. Importantly, however, it underlines the finding that social cohesion, while higher than anticipated, can certainly be improved within and across groups.

Table 3 also provides three additional relevant findings. First, the indicated measure for social cohesion is negatively associated with male, i.e., men are more 'selfish.' This finding underlines the role that women play in local communities in upholding social cohesion (a core causal pathway of the ToC). Second, respondents with greater incomes act less pro-socially. This finding may be a result of the specific cultural context. In particular, living standards are so low that comparatively richer people do *not* necessarily have a meaningful economic surplus, which they could share with society, thus, perhaps, dampening prosocial behavior compared to more affluent contexts. However, the finding also underlines that it is complicated to interpret any association at face-value as there may be an interaction effect between income and gender (women being more pro-social, but also, on average, poorer). Third, more educated people act more pro-socially. This, too, is noteworthy given that the R-WASH program may, by raising overall living standards, aid in the education of underserved communities (though this is a far-fetched causal pathway that did not make it into the ToC). Related, interviews with women also noted that the addition in free time might help young girls attain more education.

5.3 Water Access and Quality

The third overarching purpose of the Baseline phase was to assess the quality and quantity of water supply pre-R-WASH. To this end, all respondents were asked how easy it is to fetch water in their neighborhood (on a scale from 0 to 10) and to what extent respondents felt the water is of high quality (again, on a ten-point scale). The core results from this analysis are presented in Table 4. There are two main findings. First, by and large, the (perceived) water quality and supply is *lower* in the refugee/IDP community. This may not be surprising, but it does underline the

Table 3: Predictors of pro-sociality

	Dependent variable:
	Cohesion_index
Displaced person	-0.202^{***}
	(0.030)
Age	0.001
	(0.002)
Male	-0.048
	(0.031)
Children	-0.012^{***}
	(0.004)
Education	0.075^{***}
	(0.009)
Income	-0.0002^{***}
	(0.00003)
Observations	4,521
R^2	0.034
Adjusted R^2	0.033
Residual Std. Error	0.984 (df = 4514)
F Statistic	26.351^{***} (df = 6; 4514)
Note:	*p$<$0.1; **p$<$0.05; ***p$<$0.01

Notes: Table 3 reports coefficients and standard errors of an ordinary least squares regression of the indicated cohesion index (greater values indicate more cohesive outcomes) on the indicated covariates at BL1.

necessity to pay particular attention to the refugee/IDP community. Second, there is significant variation across sites. In Sudan—as was supported during field work—water quality and access are better compared to Somalia, while Ethiopia ranges in the middle.

Table 4: Water quality and access across sites

Site	Population	Country	Water access (b1)	Water quality (b2)
Control	Host	Ethiopia	4.313	5.857
Control	IDP	Ethiopia	3.716	5.631
Control	Host	Somalia	2.866	3.402
Control	IDP	Somalia	3.867	4.514
Control	Host	Sudan	5.050	5.799
Control	Refugee	Sudan	4.085	5.110
R-WASH	Host	Ethiopia	3.992	4.720
R-WASH	Refugee	Ethiopia	3.357	4.517
R-WASH	Host	Somalia	2.671	2.716
R-WASH	IDP	Somalia	2.644	2.919
R-WASH	Host	Sudan	5.002	5.062
R-WASH	Refugee	Sudan	3.608	3.834

Notes: Table 4 reports mean values on a scale from 0 (very poor) to 10 (very good) of the indicated questions regarding water quality and access.

5.4 Update and Validation of Causal Pathways for the ToC

The fourth overarching finding of the Baseline phase was to explore new, update, and validate causal pathways that link R-WASH to social cohesion. Recall that the preliminary ToC was derived on the basis of a desk review, qualitative expert interviews, an expert survey and a review of the academic literature. The field work—predominantly the qualitative component—was then used to sense-check this theory and propose new causal pathways. While it is beyond the scope of this Baseline report to provide clues for all causal pathways, we want to lay out five particularly important pathways that were uncovered and strengthened during the field work. For all remaining causal pathways, we refer the reader to the IR, which lays out the hypothesized interlinkages in detail.[13]

[13]The authors stress that the ToC will be updated once more after the program was implemented. This may include more targeted programming elements such as the coordination mechanisms to ascertain stakeholder engagement

Before laying open the revised pathways, we briefly recall how the ToC is formulated. As laid out in the IR, we use the language of directed acyclic graphs (DAGs). A DAG includes of two core elements. First, there are random variables. Variables may be observed or unobserved and may take on any distribution. For instance, the treatment variable "R-WASH" is binary (either the program is present or not), while the outcome "social cohesion" is continuous. Second, variables are causally related to one another using arrows, which denote a direct causal effect or **pathways**. Thus, if R-WASH causes social cohesion, we would describe this as R-WASH $\rightarrow$ Social Cohesion. Importantly, because the future cannot predict the past, DAGs cannot by cyclical. Rather, they are acyclic. Third and related, the absence of an arrow means there is no causal relation between two variables.

A first core causal pathway, linking R-WASH to Social Cohesion, operates via improved **safety**, which holds particularly true for women. The theoretical argument is that women who experience fear and anxiety (e.g., while fetching water or because their health situation is dire) have lower physical and mental resources to act pro-socially toward other community members—be they from the refugee/IDP population or the host population (Carleton, Collimore and Asmundson, 2007). This pathway is not new (see IR) but it was corroborated both in the open-ended as well as in the axial coding of the qualitative data. During field work, the connection between water supply and safety for women was made apparent via two interrelated pathways. First, according to affected persons and reference persons, the further away the water source is, the greater the danger for women and girls to be harassed or even raped on the way. The risk is greatest when water must be fetched from rivers due to high costs of water at water supply points (e.g., in Doolow or Luuq). Second, the way to water tanks (in case of failure of the water supply system) is usually longer and therefore more dangerous than the way to a water point (though this does not apply to Somalia). All should be improved upon by R-WASH.

A second core causal pathway, linking R-WASH to Social Cohesion, operates via more **free time**, which, again, holds particularly true for displaced women, but also displaced persons in general (including women and men). This pathway, too, had already been hypothesized. Theoretically speaking, the pathway assumes that free time is a necessary pathway for social cohesion to arise since individuals need sufficient time in order to engage in leisure activities such as sports, community gatherings or festivities that can ultimately boost togetherness (Morata et al., 2023). The open-ended as well as in the structured analysis of the qualitative data corroborated this core pathways. Women of both groups (displaced persons and host community) complained about the long waiting times for water (up to 2-3 hours). Pregnant women or women with disabilities conveyed that they were often unable to carry the water for long distances and/or to wait long times. In such cases, the families often sent younger girls to collect water who then neglect school (a po-

and exchanges between stakeholders as well as the environmental and social screening and management plans.

tential pathway to revisit in the Endline). Easier access to water supply would, thus, free up time for women and girls, which they could use for more relevant educational and economic-related purposes, which could subsequently boost incomes and also, in the long run, social cohesion.

A third core causal pathway, linking R-WASH to Social Cohesion, operates via a risk of **land disputes**. This pathway gained prominence in the field as it had not been hypothesized during the IR phase. The theoretical proposition is that land disputes are known to be particularly detrimental to social cohesion, particularly in the Global South where property rights are often weak (Kalande, 2008). Given the relevance of real estate to earn a living in the sites under study, a potential or real land dispute due to R-WASH may thus present a real hindrance to improving social cohesion—though the Environmental and Social Impact Assessments and Management Plans may offer sufficient safeguard against this risk. That said, the possibility for land disputes is acute across all three countries, but particularly in Ethiopia and Somalia, since R-WASH will need to expand upon the existing water supply system by making use of additional land. In Kebri Beyah, for example, the well is a good 5 kilometers from the city center and thus traverses a several plots of land. This conflict, though currently silent, may resurface when new pipes are built / renovated or wells updated. A second reason for land as a likely negative pathway from R-WASH to social cohesion has to do with the growth of the host community population, which has increased the value of land. In Ethiopia, people of the host community build houses within the area commonly understood as the refugee/IDP settlement area. (The demarcation between camp and host community is not clearly defined.) Our interlocutors expected that the distribution of land will be one of the biggest problems in the implementation of the new refugee law. In Qoloji, too, the conflict over land is already very pronounced because of the size of the camp, which covers a relevant part of the host community's arable and grazing land. By contrast, the risk of land disputes is less acute in Sudan and, as such, showcases that this causal pathway is predominantly derived form the Ethiopian and Somali context.

A fourth core causal pathway linking R-WASH to Social Cohesion is **envy**, which was observed among some host community members. The theoretical argument behind this causal pathway is that envy undermines pro-social behavior by sparking selfishness, anger, and potentially aggression among individuals who harbor envy (Fischer and Torgler, 2006). In some of the settings, like Girba in Sudan, the displaced persons community had a substantively better water supply than the host community (see our context description above). Some host community members expressed frustration and anger over what they perceived as privileged treatment of the refugee community.[14] The quantitative data supports the impression from the field: By analyzing item C4.B ("On a scale from 0 (not at all) to 10 (very much so), were you envious of the public goods other refugees / IDPs have gotten in the last two years?"), we can confirm that envy is a significant obstacle to

[14]Note however, that these qualitative impressions are not mirrored in people's assessments of water quality in Sudan in Table 4.

peace and thus a potential pathways that R-WASH can alleviate.

A fifth core causal pathway linking R-WASH to Social Cohesion is via **incomes**. Income is a long-standing social scientific predictor of social cohesion and related behaviors such as trust (Brandt, Wetherell and Henry, 2015). The relation between income and social capital, however, is not necessarily linear, as Table 3 shows. An increased focus on making ends meet may even undercut social cohesion. That said, in the ToC we formulated, increased income work as a driver of social cohesion as it allows individuals to invest additional economic and social resources into the community. The link from R-WASH to improved income (and, by a similar token, living conditions) is relatively straightforward. In Kebri Beyah, for instance, the qualitative field research demonstrated that a 20 litre jerrycan of water costs 20 ETB, which makes up a huge proportion of the typical total cash transfer to the refugees, which breaks down to 8 ETB (USD 0.25) per person per day (in 2022). As the current water supply by the official system only delivers about 5-6 litres a day (a quarter of the SPHERE minimum standard) people are forced to buy supplementary water. Combined with the lost income opportunities due to the high time required to procure water, the financial burden on households is enormous. This at a time of food insecurity due to climate change and the current supply crisis caused by the Russian war in Ukraine. Again, R-WASH should ideally improve this pathway, which (see next) is a likely driver of social cohesion.

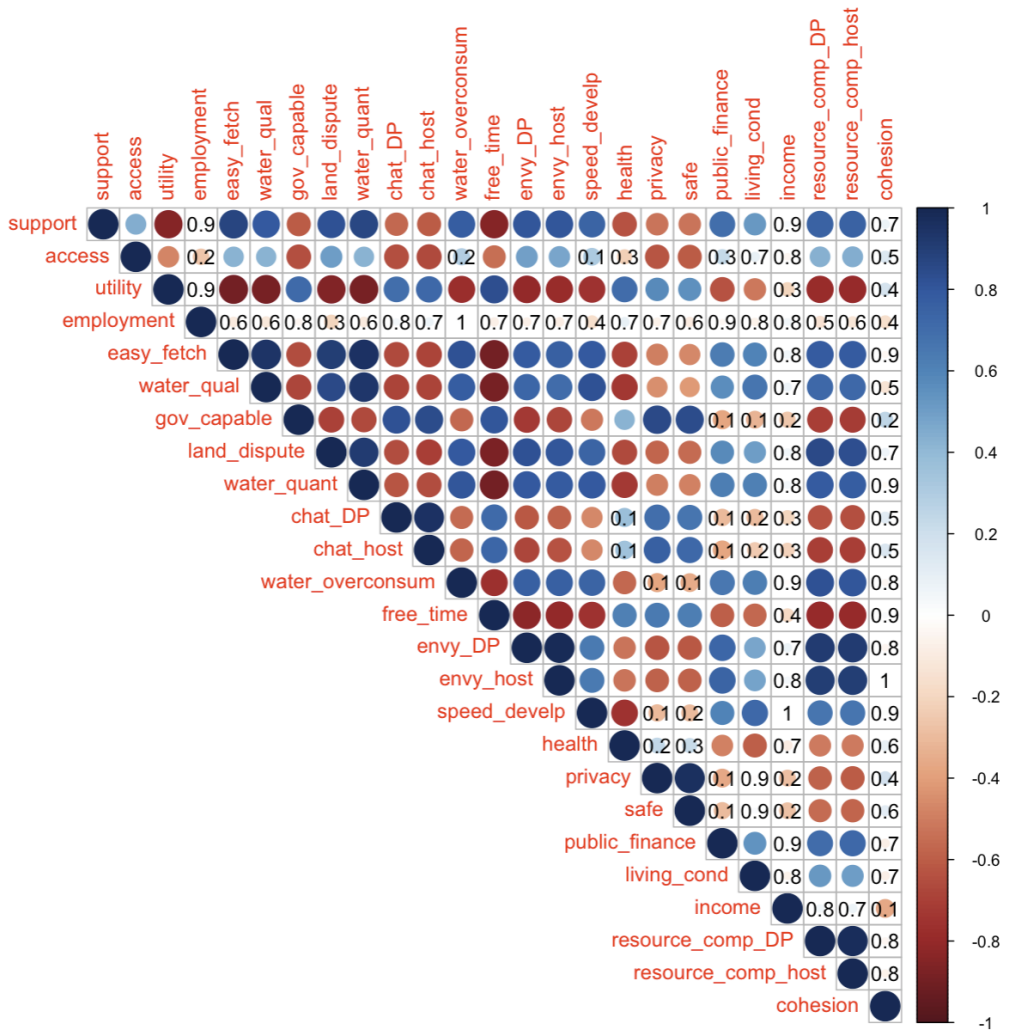
5.5 Correlates of Social Cohesion and ToC Components

A fifth key finding concerns the plausibility of the ToC, which was tested on the basis of the quantitative data. As stated, the causal pathways of the hypothesized ToC were used to inform the quantitative survey. The core goal of the survey regarding the ToC at the Baseline stage was to provide first correlational clues as to whether the hypothesized causal connections are operative. It is important to note that such correlational evidence is *not* conclusive evidence that the ToC for how R-WASH affects social cohesion is sensible. Correlation, after all, does not imply causation. What is more, the ToC is a *causal model*. Simple correlation coefficients are thus only partially informative since they do not adequately adjust for the relevant confounders / mechanisms—i.e., the causal structure (Elwert and Winship, 2014). One should also keep in mind that not all correlations are informative given that not all variables affect one another.

With these caveats in mind, Figure 13 shows the correlation coefficients for all indicated constituent elements of the hypothesized ToC. Each row/column represents a variable—including all variables ranging from a1 (development support) to e1 (social cohesion)—and how a given variable correlates with all other variables in the ToC and, in particular, with social cohesion (e1). As the final row / column shows, all variables are, indeed, positively correlated with Social Cohesion. The strength of the correlation, however, is not overly strong (light blue shading). Most p-values are not significant, underlining that R-WASH needs to be implemented for the ToC to be set in motion. Still, there seem to be positive correlations between the variables that make up the ToC, which is a first reassuring piece of evidence that the Theory is on a good track. The final ToC, however, will only be tested when the endline data will be collected. At this point, the evaluation will be in a position to trace the single causal arrows of the theory (qualitatively and quantitatively) and marry this evidence.

The Correlation Matrix also supports other key pathways, albeit only correlationally. For instance, impressions about development projects overall ("support" or A in the ToC) correlates positively with perceived water access and employment (the latter is not significant) as well as the ability to fetch water and the water's quality. Similarly, water overconsumption (*water overconsum*) predicts resource competition and improved governance capacity (gov capable) positively predicts privacy and safety. These correlations, however, should be interpreted with caution. They merely show how these variables—which are *causal pathways of R-WASH*—currently correlate with one another. Such a correlation neither established causality, nor does it corroborate the ToC of R-WASH.

Figure 13: Correlation Matrix of Hypothesized Pathways in ToC



The Figure shows the correlation coefficients of all constituent parts of the ToC. Dark blue indicates a positive correlation, dark red indicates a negative correlation. Numbers refer to p-values. If there is no number, the p-value is below 0.1 and thus significant.

6 LESSONS LEARNED

The baseline research has yielded four major lessons for the ongoing evaluation. The first and most pertinent lesson is that the evaluation must pay close attention to the political situation across the three countries. While the second baseline was successfully implemented in Ethiopia and Somalia, the situation in Sudan was highly volatile and led to the inability of the implementing partners and research team to obtain the required research approvals to field the surveys. Weeks after, the 2023 Sudan conflict broke out, making it impossible to field the survey. The conflict not only put the evaluation in Sudan on hold but, more importantly, has led to the suspension of the entire R-WASH program in the country. The R-WASH evaluation is, to a degree, shielded against this risk by i) drawing upon evidence from three countries and ii) by having already implemented one baseline in Sudan.

The second insight can be gleaned from the analysis of the status quo, pre-R-WASH. The qualitative and quantitative data have shown a consistent pattern that the water situation is dire. But, perhaps less obviously, social cohesion is at *comparable* levels across the treatment and control group (more below) as well as across the host and refugee/IDP population. This is an important lesson in as much as it underlines that the selected R-WASH communities present a receptive context within which to implement the R-WASH program. While social cohesion can certainly be improved upon (particularly in the refugee/IDP population; see Table 5), there is, at least on the basis of our data, little expectation that the program will exacerbate any looming conflicts. Put differently, we do not see any immediate risk of implementing R-WASH on the basis of our analysis of social cohesion and conflict in the program communities. If anything, the host community seems slightly higher on the social cohesion outcomes, thus underlining the insight that the preparatory work of the program can continue as planned. That said, we must caution that social cohesion is lowest in Somalia and the difference between the host and refugee/IDP populations is also starkest in this country. The low level and the difference echoes the ongoing volatile security situation in the country, which underlines that particular care is necessary as the program gets implemented.

The third insight concerns the comparability of the sites. Parallel trends have already been discussed and shown to be solid. Equally important, the second baseline has shown broadly similar levels of social cohesion across the R-WASH communities and the control sites. As much is shown in the aggregate in Table 6. The similarity in cohesion levels is important as it alleviates concerns about the control sites not being accurate 'counterfactuals.' It also makes the comparative analysis and process tracing a credible way forward. For this reason, we suggest that the plan to survey and qualitatively assess both the control and treatment communities should be continued. There is, in other words, no reason to drop/substitute any control sites. Interestingly, the comparability also, broadly speaking, extends across the sites. While there are differences across the countries, the average social cohesion outcomes are broadly similar. This underlines that the countries—though

Table 5: Social cohesion across outcomes (Baseline 2)

Country	Population	Cohesion_index
Ethiopia	Host	0.062
Ethiopia	Refugee/IDP	-0.011
Somalia	Host	0.082
Somalia	Refugee	-0.197

Notes: Table 5 reports mean values of the aggregate (standardized) social cohesion index across the countries and study populations. Values greater than 0 mean *positive social cohesion* relative to the overall study population. Standardization refers to subtracting each observation by the mean and then dividing by the standard deviation.

culturally not isomorphic—may offer a fertile ground to extract a generalizable ToC.

Table 6: Social cohesion across countries and populations (Baseline 2)

Country	Site	Cohesion_index
Ethiopia	Control	0.004
Ethiopia	R-WASH	0.061
Somalia	Control	-0.030
Somalia	R-WASH	-0.033

Notes: Table 6 reports mean values of the aggregate (standardized) social cohesion index across the countries and treatment sites. Values greater than 0 mean *positive social cohesion* relative to the overall study population.

A fourth insight concerns the issue of gender. While our analysis is pre-program (and can thus not talk about the impact of the program as of now), the qualitative interviews have shown a consistent pattern that the day-to-day challenges regarding water are disproportionally faced by women. At the same time, the decision-making as to the implementation of the utility, including the details of construction, pricing, governance, etc., are predominantly done by men. This is, of course, a product of the cultural context. That said, we do caution that gender mainstreaming is an important UN goal, which may require further reflection on the ground. The direct consequence of the low water supply for women imply that it may be particularly fruitful to include them more

directly in the decision-making process.

References

Andres, Luis Alberto, Christian Borja-Vega, Crystal Fenwick, Jaime de Jesus Filho and Ronald Gomez-Suarez. 2018. "Overview and meta-analysis of global water, sanitation, and hygiene (WASH) impact evaluations." *World Bank Policy Research Working Paper* (8444).

Beard, Victoria A and Diana Mitlin. 2021. "Water access in global South cities: The challenges of intermittency and affordability." *World Development* 147:105625.

Berg, Joyce, John Dickhaut and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10(1):122–142.

Brandt, Mark J, Geoffrey Wetherell and PJ Henry. 2015. "Changes in income predict change in social trust: A longitudinal analysis." *Political Psychology* 36(6):761–768.

Burki, Talha. 2022. "Humanitarian crisis in Tigray amidst civil war." *The Lancet Infectious Diseases* 22(6):774–775.

Carleton, R Nicholas, Kelsey C Collimore and Gordon JG Asmundson. 2007. "Social anxiety and fear of negative evaluation: Construct validity of the BFNE-II." *Journal of Anxiety Disorders* 21(1):131–141.

Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics* 127(4):1755–1812.

Chan, Joseph, Ho-Pong To and Elaine Chan. 2006. "Reconsidering social cohesion: Developing a definition and analytical framework for empirical research." *Social indicators research* 75(2):273–302.

Elwert, Felix and Christopher Winship. 2014. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual review of sociology* 40:31–53.

Engel, Christoph. 2011. "Dictator games: A meta study." *Experimental economics* 14:583–610.

Fischer, Justina AV and Benno Torgler. 2006. "Does envy destroy social fundamentals? The impact of relative income position on social capital." *U of St. Gallen Economics Discussion Paper* (2006-04).

Hashim, Ahmed. 2023. "Attacks on healthcare facilities in Sudan: an unprecedented surge amid a military coup." *Medicine, Conflict and Survival* 39(1):86–92.

Hirwa, Hubert, Qiuying Zhang, Yunfeng Qiao, Yu Peng, Peifang Leng, Chao Tian, Sayidjakhon Khasanov, Fadong Li, Alphonse Kayiranga, Fabien Muhirwa et al. 2021. "Insights on water and

climate change in the greater horn of africa: Connecting virtual water and water-energy-food-biodiversity-health nexus." *Sustainability* 13(11):6483.

Hovil, Lucy and Sasha Jesperson. 2023. "Forced into Slavery: Eritreans Caught Between Refugee and Migration Policies." *Journal of Modern Slavery: A Multidisciplinary Exploration of Human Trafficking Solutions* 8(1).

Husain, Farah, Colleen Hardy, Lemlem Zekele, David Clatworthy, Curtis Blanton and Thomas Handzel. 2015. "A pilot study of a portable hand washing station for recently displaced refugees during an acute emergency in Benishangul-Gumuz Regional State, Ethiopia." *Conflict and Health* 9:1–9.

Hutton, Guy, Laurence Haller, Sanitation Water, World Health Organization et al. 2004. Evaluation of the costs and benefits of water and sanitation improvements at the global level. Technical report.

Kalande, William. 2008. Kenyan land disputes in the context of social conflict theories. In *Fig Commission*. Vol. 7.

Morata, Txus, Paco López, Teresa Marzo and Eva Palasí. 2023. "The influence of leisure-based community activities on neighbourhood support and the social cohesion of communities in Spain." *International Social Work* 66(2):568–584.

Nyadera, Israel Nyaburi, Mohamed Salah Ahmed and Billy Agwanda. 2019. "Transformation of the Somali Civil-War and Reflections for a Social Contract Peacebuilding Process." *Gaziantep University Journal of Social Sciences* 18(4).

Posner, Daniel N. 2017. "When and why do some social cleavages become politically salient rather than others?" *Ethnic and Racial Studies* 40(12):2001–2019.

Schellinger, Sarah M. 2022. *Nubia: lost civilizations*. Reaktion Books.

Srivastava, Prachi and Nick Hopwood. 2009. "A practical iterative framework for qualitative data analysis." *International journal of qualitative methods* 8(1):76–84.

UNHABITAT. 2021. Qoloji Spatial Profile. Technical report.
URL: *https://unhabitat.org/qoloji-spatial-profile*

UNICEF. 2016. "Strategy for water, sanitation and hygiene 2016-2030.".

A Supplementary Information

A.1 List of persons interviewed and sites visited (selection)

A.1.1 Sites visited

In Ethiopia, the following four sites were visited by three consultants, Bernadette Schulz, Prof. Anselm Hager, Innocent Yekeye:

- Kebri Beyah
- Aw-Bare
- Sheder
- Qoloji

In Sudan, the following two sites were visited by two consultants, Innocent Yekeye and Prof. Carlo Koos:

- Girba
- Kassala

In Somalia, the following two sites were visited by the consultant Innocent Yekeye:

- Dolow - Kabasa
- Quansaxley

Across the three sites, the following people were interviewed to obtain a harmonious impression across the three sites:

- Regional senior staff within UNHCR and UNICEF (e.g., Samuel Godrfey, Regional Advisor WASH; Robin Lloyd, Regional WASH officer UNHCR)
- Reagional program management within UNHCR / UNICEF (e.g., Steven Mudhuviwa, program manager of R-WASH)
- Country-level senior WASH staff (e.g., Kitka Goyol, Chief WASH officer Ethiopia; Mohammed Juma, WASH officer Sudan; Janet Simion - WASH Officer, UNHCR; Victor Kinyanjui, Chief WASH officer)

- Regional senior staff within UNHCR / UNICEF (e.g., Utpal Moitra, Chief Field Officer)
- Regional senior WASH staff within UNHCR / UNICEF (e.g., Ayele Munye, Associate WASH officer; Luckson Katsi, WASH coordinator)
- Key implementing partners on the ground (e.g., IMP)
- Key political staff (e.g., Mayor of Kebri Beyah)
- Technical staff (e.g., SachsenWasser in Jijiga)
- Beneficiaries (incl. focus group discussions with women, participant observation at the field sites and water collection points)

A.2 Methodology (additional details)

A.2.1 Quantitative survey instrument

The primary objective of the rigorous impact evaluation is to assess the long-term outcome of social cohesion, which is a complex and intangible concept. Defining and measuring social cohesion poses inherent challenges, particularly due to its varying patterns in different contexts. This presents a significant challenge for the impact evaluation as we aim to measure the construct in a way that a) allows for comparisons across the six sites and b) enables comparisons with other studies, facilitating benchmarking of estimates and the development of a universally applicable ToC.[15]

Our key definition of social cohesion is derived from Chan, To and Chan (2006). The authors' definition is as follows:

> "*Social cohesion is a state of affairs concerning both the vertical and the horizontal interactions among members of society as characterized by a set of attitudes and norms that includes trust, a sense of belonging and the willingness to participate and help, as well as their behavioural manifestations.*
>
> Chan, To and Chan 2006, 290

We condense this definition to its quintessential elements and define social cohesion as '**levels of trust, cooperation and identification in a community**.' We measured the three elements as follows:

First, we measure **trust** using a trust-game (Berg, Dickhaut and McCabe 1995; Row 2 in Table A1). Each survey respondent was given the opportunity to send any amount of an imaginary 10 Dollar sum to the next respondent. Importantly, the game was played with *both* an in-group receiver (refugee/IDP or host population, respectively) as well as an outgroup receiver (refugee/IDP or host population, respectively). The sender was informed that, should the receiver send any money back, the researchers will double the amount. In order to streamline the surveying activities and circumvent providing a source of conflict, enumerators did not implement the second part of the trust experiment because we are only interested in the decision of the respondent whether to trust (the second part refers to sending any amount back to the original sender). The precise wording was as follows:

> "Imagine the following game. You and another person from the local host community who is called Ali get 10 USD. Next, you and Ali have to give any amount of that 10 USD

[15] The cleaned data set is available from KfW upon request.

Table A1: Quasi-experimental measures for social cohesion

Construct	Measure	Scale	Processing
Trust	Trust game (first stage)	continuous	standardization
Cooperation	Trust game (second stage)	continuous	standardization
Identification	Dictator game	continuous	standardization
Social cohesion	trust i + trust ii + dictator	continuous	standardization

Notes: This Table summarizes the three quasi-experimental measures for the constituent parts of our social cohesion definition. To arrive at the social cohesion index, the three measures were each standardized, added together and then standardized again (see main text).

> to the other person. You decide first. Ali decides second. Importantly: each Dollar that you give to Ali, the local host community member, will be tripled by us and then given to Ali. That means, if you give 1 USD of your 10 USD to Ali, you then have 9 USD, while Ali will have 10 USD plus 3 times 1 USD, so 13 USD. Then, Ali can decide to send some money back to you. Let's now play this game. How much of your 10 USD do you give to Ali, which we then triple?"

Second, we measure **cooperation** using the concept of *reciprocity* (Row 2 in Table A1). This measure, too, relies on the well established trust-game by (Berg, Dickhaut and McCabe, 1995). Subjects were told to imagine that another respondent had invested in the trust game and had sent them money. We than informed the respondent that the amount of money would be doubled if they decided to send money back. We then asked the respondent how much money they would like to send back, thus reciprocating / cooperating with the sender. Again, the game was played with *both* an in-group member as well as an outgroup member.

Third, we measured **identification** using a public donation game (Row 3 in Table A1). In essence, we adapted the well-known dictator-game: Subjects were provided with an imaginary 10 USD and asked if they would like to share any of this money with the respective community. We then gave them the opportunity to invest part of the money into a fund tailored toward improving the infrastructure in the community. We then recorded the amount of money the respondent would like to invest. The measure was elicited for *both* the own community as well as the *other* community.

In addition to these behavioral outcomes, we also applied survey-based measures in the quantitative survey.

First, we measured **trust** by adapting the well-established general item (e.g., European Social

Table A2: Survey-based measures for social cohesion

Construct	Measure	Scale	Processing
Trust	Generalized trust	1-5	Standardization
Cooperation	Lost wallet	1-5	Standardization
Identification	Identity group	Binary	Standardization
Social cohesion	trust + wallet + identification	continuous	Standardization

Survey; Row 1 in Table A2). We asked subjects "*Generally speaking, to what extent do you agree with the following sentence: most people [in my ingroup / in the outgroup] can be trusted*". Answer choices were scored on a five-point scale ranging from 1, completely disagree, to 5, completely agree.

Second, we measured **cooperation** by adapting the well-established lost wallet item (e.g., Canadian General Social Survey and Casey, Glennerster and Miguel 2012; Row 2 in Table A2). We asked subjects "*Imagine you lost a wallet that contained 20 dollars and someone from [the in-group / outgroup] found it. How likely is it that the money is returned to you?* Answer choices were scored on a five-point scale ranging from 1 (very unlikely) to 5 (very likely).

Third, we measured **identification** by adapting a well-established item on one's relevant social reference group (e.g., Posner (2017); Row 3 in Table A2). We asked subjects "*We have spoken to many people in this area and they have all described themselves in different ways. Some people describe themselves in terms of their religion, ethnic group, language-group or nationality. Others describe themselves in economic terms, such as farmer. Which specific group do you feel you belong to first and foremost?*" The answer choices included the relevant social reference category (variable name: $cohesion\_survey\_identity$).

Besides measuring social cohesion, the survey also included items for each of the causal arrows in the ToC. These were classifed into outputs, outcomes, medium-term impacts as well as long-term impacts.

Outputs In a second step, we measured the immediate **outputs** of R-WASH. We added four brief items that were asked across all three contexts. First, to measure the implementation of consulting services (A.1), respondents were asked: *'On a scale from 0 (not at all) to 10 (very much so), to what extent are you aware that the local water provider has received support from outside organizations in the last two years?'*. Second, to measure the construction of pipelines/wells (A.2), respondents

were asked: *'In the last two years, how many direct water access points were constructed within close distance (100m) of your home?'*. Third, to measure the community governance of water resources (A.3), respondents were asked: *'Are you aware of an office or board where you can make your water problems heard?'*. Fourth, to direct effects on employment, respondents were asked 'Are you currently employed?'.

Outcomes In a third step, we measured the **outcomes** of R-WASH. To measure water access (B.1), respondents were asked: 'On a scale from 0 (not at all) to 10 (very much so), how easy is it to fetch water in this neighborhood?'. To measure water quality (B.2), respondents were asked: 'On a scale from 0 (not at all) to 10 (very much so), to what extent would you say the drinking water in this area is of high quality?'. To measure governance capacity (B.3), respondents were asked: 'On a scale from 0 (not at all) to 10 (very much so), to what extent are local government officials capable in this area?'. To measure potential land disputes, respondents were asked 'On a scale from 0 (not at all) to 10 (very much so), how severe are conflicts over land in this neighborhood?'.

Medium-term impacts In a fourth step, we measured the **medium-term impacts**. To measure reduced contact (C.1), respondents were asked "In the last month, how often did you have a chat with a stranger from the refugee / IDP community in a given week?". To measure water overconsumption (C.2), respondents were asked "On a scale from 0 (not at all) – 10 (very much so), how much of a problem is it that people consume too much water in this community?". To measure free time (C.3), respondents were asked "How many hours each day do you have free time?". To measure envy (C.4), respondents were asked "On a scale from 0 (not at all) to 10 (very much so), were you envious of the public goods other refugees / IDPs have gotten in the last two years?". To measure frustration about slow implementation (C.5), respondents were asked "On a scale from 0 (not at all) to 10 (very much so), how happy were you with speed at which development projects have been implemented in this area in the last two years?". To measure health (C.6), respondents were asked "How many days were you sick in the last two years?". To measure privacy (C.7), we asked "On a scale from 0 - 10, do you feel like your privacy is respected in this community?". To measure safety (C.8), respondents were asked "On a scale from 0 - 10, how safe do you feel in this community?" (variable name: $mediumterm\_safety$).

Long-term impacts In a fifth step, we measured **long-term impacts**. To measure public spending (D.1), respondents were asked: "On a scale from 0 (not at all) to 10 (very much so), how much have community finances improved over the last two years?". To measure living conditions (D.2), respondents were asked: "On a scale from 0 - 10, how satisfied were you with your living conditions in the last two years?". To measure incomes (D.3), respondents were asked: "What is your average

monthly income (in USD)?". To measure competition (D.4), respondents were asked: "On a scale from 0 (not at all) to 10 (very much so), to what extent do you feel like you had to fight over resources with other refugees / IDPs in the last two years?".

A.2.2 Qualitative survey instrument

The purpose of the baseline interviews was threefold. First, we used the qualitative interviews during and before baseline to *explore* new hypotheses on how R-WASH may affect social cohesion. This step was thus largely exploratory and did not rely on an overly structured survey instrument. Second, we used qualitative interviews to define and corroborate key indicators of interest, above all, social cohesion. Third, we used the interviews to explore current conflict lines and how they moderate the proposed ToC. The following guiding questions were used.

A.2.3 Camp composition

A first set of questions deals with describing the social context of the camps. They were be asked to 1-2 reference persons (pre-selected by UNHCR / UNICEF).

- Which ethnic groups are represented?
- From which countries/regions did they come?
- What is the age structure?
- What is the gender composition?
- What is residents' religious background?
- What other salient ethnic identities play a role (e.g., clans)
- What were the main reasons for fleeing?
- How long have people been living in the camp?
- How much turnover is there?
- Are there any signs that turnover might increase in the near future?

A.2.4 Conflict lines within the camps

A second set of questions dive more deeply into the structure of the camps as well as salient conflict lines.

- Are there groupings or social hierarchies between camp residents?
- If so, on what factors are they based (economic, social, political, religious, ethnic/ tribal)?
- How do the camp residents interact with each other?
- What livelihoods do camp residents have (full provision by UNHCR/ UNICEF or additional sources of income)?
- Do camp residents have the right to work? If yes, where do they work?
- Do camp residents have access to land?
- Who owns the land on which the camp is built?
- Do the available resources meet the needs of the camp residents?
- What is the general pattern of social interactions among refugees/IDPs?
- Are there examples for mutual support in difficult situations when a person needs help?
- How do people perceive others (e.g. refugees from other countries/ ethnic groups/religion or people whose reasons for fleeing differ)?

A.2.5 Host community composition

A third set of questions deals with describing the social context of the host community. They will be asked to 1-2 reference persons (pre-selected by UNHCR / UNICEF).

- Which ethnic groups are represented?
- What livelihoods / professions do people have?
- Do the available resources (e.g. land, food, water, education, accommodation) meet the needs? How are they distributed?
- What socio-economic and political structures, cleavages and dependencies exist?
- What patterns of social cooperation are most salient?

A.2.6 Intergroup cooperation

A fourth set of questions deals with the interaction between camp residents and host communities

- Do camp residents use public services provided mainly in and for the host communities?
- Are there trade relations between camp residents and host community?
- Are there joint celebrations or other forms of cultural exchange?
- Are there cases of intermarriages between the groups?
- Are there opportunities for joint sports (e.g., football pitch)?
- Are sports normally carried out within ethnic or religious groups or do people from various background participate?
- Which other forms of interaction between camp residents and host population exist?
- What is the general pattern of social interactions among refugees/IDPs and between refugees and host population? To what extent is there cooperation and mutual acceptance?
- If work outside the camp is allowed, who are the employers? What are the working conditions? How is the relation between employers and workers seen (by both sides)
- Do workers of host community and camp workers compete for jobs?
- Is there competition on land ownership (if access to land is legally possible for camp residents)?
- Is there competition on other resources (e.g. food, firewood) between camp residents and host population?
- If camp residents use public services provided in and mainly for the host community, are they (or do they feel) in competition with the host population?
- If trade takes place between host community and camp residents, is it seen as fair and mutually beneficial by both sides?
- How do camp residents and people of the host community view their relationship with each other?
- How do camp residents and people of the host community perceive each other?
- Do camp residents feel rejected or excluded?

- Do (poor) people of the host community feel disadvantaged because refugees receive more support?

A.2.7 Intergroup conflict

A fifth set of questions deals with potential conflicts between camp residents and host communities

- How do/ could current and expected conflicts at national or regional level affect the situation on the ground?
- Are there major conflicts i) within the camp, ii) within the host community, iii) between camp residents and host population?
- Which are the main conflict issues?
- To what extent does water play a role?
- Who are the main conflict parties?
- How are specific socio-cultural structures, traditions and conventions supporting conflicts? Are there powerful interest groups behind these conflicts?
- Are there refugee groups and host population from contemporary armed conflicts that directly oppose each other? How are these conflicts carried out in the camp (e.g., rebel recruitment, gangs, etc.)
- In what form do conflictive tensions manifest themselves?
- Are there examples of hate speech, exclusion etc.?
- How many violent incidents have there been in recent years?
- What kind of incidents were they (quarrels, theft, assault, sexual abuse, homicide)? Are there statistical data/ reports on these incidents?
- Who were the direct actors and who were possible drivers of conflict behind?
- To what extent were UNICEF, UNHCR, NGOs or related providers the target of aggression or attacks?
- What conflict resolution structures are in place at local level?

- Do people think there are conflicts in the camp, in the host community and between the two communities? Which are the main conflict issues?
- How are these conflicts expressed?
- Are there people who are excluded or offended or who feel excluded or offended?
- Are there disputes over resources such as land, water, food?
- What specific violent incidents can people remember? What kind of incidents were they? What were they about? Who was involved? What impact did this have on the community? Where these conflicts resolved? If yes: how and by whom?
- Do people know local/ traditional conflict mediation mechanisms? Did they ever use one of these mechanisms or do they know examples how conflicts were resolved? Who would they turn to in the event of a conflict?

A.2.8 WASH

A sixth set of questions deals with the supply of water and sanitation

- What is the current situation regarding water supply in the camp and in the host community (quantity, quality and tariffs)?
- How many liters of water are available on average per person?
- What do people use the water for?
- How many persons does one water tap serve on average?
- Are water taps equally distributed across the camp/host pop, or are some communities disadvantaged regarding access?
- What are the main problems with the current water supply?
- Are there any data on incidence rates for water-borne diseases (diarrhoea, skin diseases) or dehydration?
- How are the current water tariffs, In the camp and in the host community?
- What is the current situation for sanitation (Type and functionality of infrastructure, distribution, distance from households)
- Is there a disposal system for faeces?

- Who is responsible for the family's water supply?
- What is water used for?
- Is the amount of water available sufficient?
- How far do people on average have to walk to get to the next tap?
- How long do people have to queue in the morning to get water?
- Do some communities feel disadvantaged regarding access?
- How does a lack of water access depress living conditions (especially for women and children)?
- Is there competition for water in the camp?
- Is there competition for water in the host community?
- Is there competition for water between the residents of the camp and the host community?
- Do users feel that water tariffs are reasonable?
- Would users accept to pay/ to pay more for improved water quality?
- Does a lack of sanitation lead to insecurity for women (Distance of sanitation facilities from the housing sites, lack of privacy)?
- Does a lack of sanitation lead to health issues and does this, in turn, affect the living conditions?

A.2.9 R-WASH and social cohesion (negative links)

A seventh set of questions deals with a potential positive effect of R-WASH on social cohesion.

- What are the main positive effects expected from the implementation of the R-Wash project, e.g.
  - on health
  - on living conditions
  - on safety and security
  - on social cohesion within the camp and within the host community and between both communities)

  - on the public budget respectively on the budget of the UN organizations
  - others?

- What do the refugees/IDPs, and the host community expect from the implementation of the R-Wash project?
- Do they think that the implementation of R-Wash will improve their living conditions? How?
- Could better water quality improve their health?
- How would better health affect their living conditions (e.g., their ability to work, their income, their children's educational opportunities)?
- Would water points closer to homes reduce the workload of women and girls? How could they use the time saved for themselves and for their families?
- Could improved sanitation infrastructure increase safety, especially for women and children?
- Do people think that increased water supply would reduce competition within the camp/ within the host community and between camp population and host community? What advantages would result from this?

A.2.10 R-WASH and social cohesion (positive links)

An eighth set of questions deals with a potential negative effect of R-WASH on social cohesion.

- Do people think that the project implementation could trigger additional conflicts? To what extent can the provision of wash exacerbate conflicts?
  - does the location matter?
  - does the distribution of construction contracts and job opportunities matter?
  - to what extent might water overuse become a problem?
- Could different tariffs between the refugees and the host communities lead to conflicts?
- Could higher tariffs (due to improved water quality) lead to conflicts?
- Which other possible risks do you see?
- If negative effects and risks are seen, how could they be addressed? Who needs to be involved and how?

- Could political developments and conflicts at national or regional level influence negatively the implementation of the project?
- Do people have any concerns that the implementation of the R-Wash Project could trigger conflicts within the camp, within the host community or between both communities? Which kind of conflicts?
- Would women/ girls miss communication with others when standing in line, if waterpoints are nearer to home and if water availability is increased?
- How important dos it seem that local companies are involved in contracts during implementation?
- How should jobs be distributed during construction?
- How important is it for users to be included in the planning and implementation of the R-Wash? How could this be ensured?
- Are the current water tariffs considered fair? Would people be willing to pay more for better water quality? Would refugees be willing to pay for water?
- Which local institutions/ actors would be best suited to resolve conflicts during the implementation of the project?

A.3 Evaluators' biodata

Baseline research was implemented by the following four researchers.

Prof. Dr. Anselm Hager

Prof. Anselm Hager is a professor based at Humboldt-University in Berlin. His research focuses on rigorous (quasi-)experimental evaluations with a focus on sub-Saharan Africa.

- Email: anselm.hager@gmail.com
- Role: Principal investigator. Advanced statistics specialist.
- Responsibilities:
  - Design of overall evaluation
  - Design of quantitative survey
  - Analysis of qualitative and quantitative data
  - Field work in Ethiopia

Prof. Dr. Carlo Koos

Prof. Carlo Koos is a professor based at the University of Bergen. He is interested in the causes and consequences of war and conflict, war and gender and the impact of development cooperation.

- Email: carlo.koos@uib.no
- Role: Survey design specialist
- Responsibilities:
  - Design of quantitative survey
  - Analysis of data
  - Field work in Sudan

Bernadette Schulz

Bernadette Schulz is an independent development consultant based in Germany. She has over 30 years of experience as an employee of GIZ in leading roles in sub-Saharan Africa, and specializes in qualitative methods and gender.

- Email: bernadettesch@posteo.de
- Role: Qualitative evaluation and peace and fragility specialist
- Responsibilities:
  - Qualitative interviews and focus group discussions
  - Peace and Conflict Analysis.
  - Field work in Ethiopia

Innocent Yekeye

Innocent Yekeye is an independent development consultant working with NEDICO in Zimbabwe. Innocent has over 15 years of experience implementing and managing large-scale surveys in volatile contexts. His research projects have focused on water, education and peace-building.

- Email: innocentyekeye@gmail.com
- Survey implementation specialist
- Responsibilities:
  - Implement and manage survey implementation
  - Qualitative interviews
  - Focus group discussions
  - Field work in Somalia

A.4 Ethics protocol

Figure A.1: Ethics approval from Humboldt University

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Humboldt-Universität zu Berlin | Unter den Linden 6 | 10099 Berlin

Prof. Dr. Anselm Hager
Institut für Sozialwissenschaften

Kultur-, Sozial- und Bildungswissenschaftliche Fakultät

Ethikkommission

Ihr Antrag auf ein Ethikvotum (HU-KSBF-EK_2022_0012)

Datum:
11.02.2022

Bearbeiter:
PSR

Sehr geehrter Herr Prof. Hager,

in Bezug auf Ihren Antrag auf ein Ethikvotum zu Ihrem Forschungsprojekt „Can developmental infrastructure projects improve social cohesion?“ ist die Ethikkommission zu folgendem Votum gekommen:

Aus Sicht der Kommission bestehen keine ethischen Bedenken gegen die Durchführung des Forschungsvorhabens.

Im Namen der Ethikkommission wünsche ich Ihnen bei der Durchführung Ihres Forschungsvorhabens viel Erfolg.

Mit freundlichen Grüßen

i.V. Dr. Patrick Ressler

Postanschrift:
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
Telefon: +49 30 2093-66149
Telefax: +49 30 2093-66151

patrick.ressler@hu-berlin.de
http://fakultaeten.hu-berlin.de/de/ksb

Sitz:
Georgenstraße 47
Raum 1.34
10117 Berlin

A.5 Stakeholder matrix

Stakeholders	Involvement in the project	Interest in the evaluation
Donors		
KfW	Main stakeholder from the donor's perspective. Oversight of R-WASH implementation and the evaluation. Deep knowledge in technical and financial (humanitarian) interventions. Hiring agency of the evaluators.	Blueprint ToC for how WASH projects affect social cohesion. Exemplary implementation of rigorous, mixed-methods intervention across several sites
BMZ	Main donor of the R-WASH project. Largely passive role.	Blueprint ToC for how WASH affect social cohesion. Cross-case evidence on social cohesion measures
Implementing agencies		
UNICEF	The leading implementing agency of R-WASH. Deep contextual knowledge on humanitarians situation on the ground, esp. with regard to children.	Organizational learning for how to implement rigorous modern impact evaluations. Rigorous evidence on how WASH projects affect social cohesion. A blueprint ToC for how WASH projects may generally affect social cohesion in other context.
UNHCR	The co-lead implementing agency. Deep contextual knowledge of camps and the provision of services.	Organizational learning for how to implement rigorous modern impact evaluations. Rigorous evidence on how WASH projects affect social cohesion. A blueprint ToC for how WASH projects may generally affect social cohesion in other context.
Local contractors	Mainly responsible for key outputs of R-WASH, incl. infrastructure improvements and capacity building	Likely limited.